(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6277069号
(P6277069)

(45)発行日　平成30年2月7日(2018.2.7)　(24)登録日　平成30年1月19日(2018.1.19)

(51)Int.Cl.　　F I
G06F 9/46 (2006.01)　　G06F 9/46 350
G06F 11/20 (2006.01)　　G06F 11/20

請求項の数 8　(全 23 頁)

(21)出願番号　特願2014-124600(P2014-124600)
(22)出願日　平成26年6月17日(2014.6.17)
(65)公開番号　特開2016-4433(P2016-4433A)
(43)公開日　平成28年1月12日(2016.1.12)
審査請求日　平成26年12月4日(2014.12.4)
審判番号　不服2016-19308(P2016-19308/J1)
審判請求日　平成28年12月26日(2016.12.26)

(73)特許権者　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(74)代理人　100127535
弁理士　豊田　義元
(74)代理人　100159190
弁理士　渡部　比呂志
(72)発明者　山登　庸次
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　西澤　幸久
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

最終頁に続く

(54)【発明の名称】仮想機器管理装置、仮想機器管理方法及び仮想機器管理プログラム

(57)【特許請求の範囲】
【請求項１】
障害の生じた物理機器を検出する検出部と、
前記障害の生じた物理機器以外の稼働中の物理機器のうち物理資源の空き容量のある物理機器を複数特定し、前記稼働中の物理機器の空き容量が無い場合は、予備の物理機器を特定し、特定した複数の前記稼働中又は予備の物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する選択部と、
選択された前記物理機器それぞれに前記仮想機器のそれぞれの再配置を、並行して行うように、物理機器への仮想機器の作成を制御するクラウドコントローラに依頼する依頼部と
を備えたことを特徴とする仮想機器管理装置。
【請求項２】
前記物理資源には、物理機器に障害の生じた場合にのみ使用される確保領域を含み、
前記選択部は、前記障害の生じた物理機器以外の物理機器のうち確保領域を含んだ前記物理資源の空き容量のある物理機器を複数特定し、特定した複数の前記物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する
ことを特徴とする請求項１に記載の仮想機器管理装置。
【請求項３】
前記選択部は、特定した複数の物理機器に順序付けを行い、前記障害の生じた物理機器に配置された仮想機器それぞれの再配置先として、前記順序に基づいて選択した物理機器

の物理資源を選択する処理を繰り返すことを特徴とする請求項１又は２に記載の仮想機器管理装置。
【請求項４】
前記選択部は、特定した複数の物理機器の物理資源の空き容量が多い順に、前記特定した複数の物理機器に順序付けを行うことを特徴とする請求項３に記載の仮想機器管理装置。
【請求項５】
仮想機器を新規に作成する依頼を受付ける受付部と、
新規に作成する依頼を受付けた前記仮想機器を配置する物理機器を、物理機器に障害の生じた場合にのみ使用される確保領域を含まない前記物理資源の空き容量に基づいて選択する作成部とを更に備え、
前記依頼部は、前記作成部によって選択された物理機器に前記仮想機器の作成を依頼することを特徴とする請求項２～４のいずれか一つに記載の仮想機器管理装置。
【請求項６】
前記作成部は、物理機器のうち前記確保領域を含まない前記物理資源の空き容量の最も多い物理機器を特定し、特定した前記物理機器の前記確保領域を含まない物理資源の一部を前記新規に作成する依頼を受付けた仮想機器のうち何れかの仮想機器の配置先として選択する処理を、前記新規に作成する依頼を受付けた全ての仮想機器の配置先を選択するまで繰り返すことを特徴とする請求項５に記載の仮想機器管理装置。
【請求項７】
仮想機器管理装置で実行する仮想機器管理方法であって、
前記仮想機器管理装置が、
障害の生じた物理機器を検出する検出工程と、
前記障害の生じた物理機器以外の稼働中の物理機器のうち物理資源の空き容量のある物理機器を複数特定し、前記稼働中の物理機器の空き容量が無い場合は、予備の物理機器を特定し、特定した複数の前記稼働中又は予備の物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する選択工程と、
選択された前記物理機器それぞれに前記仮想機器のそれぞれの再配置を、並行して行うように、物理機器への仮想機器の作成を制御するクラウドコントローラに依頼する依頼工程と
を含んだことを特徴とする仮想機器管理方法。
【請求項８】
障害の生じた物理機器を検出する検出手順と、
前記障害の生じた物理機器以外の稼働中の物理機器のうち物理資源の空き容量のある物理機器を複数特定し、前記稼働中の物理機器の空き容量が無い場合は、予備の物理機器を特定し、特定した複数の前記稼働中又は予備の物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する選択手順と、
選択された前記物理機器それぞれに前記仮想機器のそれぞれの再配置を、並行して行うように、物理機器への仮想機器の作成を制御するクラウドコントローラに依頼する依頼手順と
をコンピュータに実行させることを特徴とする仮想機器管理プログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、仮想機器管理装置、仮想機器管理方法及び仮想機器管理プログラムに関する。
【背景技術】
【０００２】
ＩａａＳ（Infrastructure as a Service）型クラウドサービスの実施例として、A

ｍａｚｏｎ Ｅｌａｓｔｉｃ Ｃｏｍｐｕｔｅ Ｃｌｏｕｄ（web site, http://aws.amazon.com/ec2）、Ｒａｃｋｓｐａｃｅ Ｃｌｏｕｄ Ｓｅｒｖｅｒ（web site, http://www.rackspacecloud.com/cloud-hosting-products/servers/）がある。

【０００３】

ＩａａＳ型クラウドサービスの基盤として、Ａｍａｚｏｎはプロプライエタリなプラットフォームを用いているが、ＲａｃｋＳｐａｃｅはＯｐｅｎ ＳｏｕｒｃｅのＯｐｅｎＳｔａｃｋ（http://www.openstack.org/）を用いている。

【０００４】

しかし、ＯｐｅｎＳｔａｃｋ等のＩａａＳ基盤は、仮想リソースの管理を行うプリミティブなＡＰＩ（Application Programming Interface）提供がターゲットの中心であり、物理機器の管理はスコープ外であるため、事業者がクラウドサービスを提供する際は考慮が必要である。

【０００５】

具体的には、仮想リソースが動作する物理機器が故障した際の復旧は、ＯｐｅｎＳｔａｃｋは特にサポートしていなく、サービス事業者にて対策が必要である。市中で採用されている方法として、Ｈｉｇｈ ＡｖａｉｌａｂｉｌｉｔｙクラスタソフトウェアのＰａｃｅｍａｋｅｒ等を用いてＨＡ構成を構築し、物理機器故障時はフェールオーバーする方法がある。

【先行技術文献】

【非特許文献】

【０００６】

【非特許文献１】Pacemaker web site、［平成26年5月30日検索］、インターネット（ＵＲＬ：http://www.linux-ha.org/wiki/Pacemaker/）

【発明の概要】

【発明が解決しようとする課題】

【０００７】

しかしながら、上記の従来技術では、仮想機器を復旧するまでの時間が長くなる場合があるという問題がある。

【０００８】

仮想機器が動作する物理機器が故障した際に、復旧する手段として、Ｐａｃｅｍａｋｅｒ等のＨＡ（High Availability）クラスタソフトウェアを用いる方法がある。この方法では、Ｎ－Ａｃｔ(Active)、Ｍ－Ｓｂｙ(standby)の冗長化構成を取り、Ａｃｔｉｖｅの物理機器が故障した際に、Ｓｔａｎｄｂｙの物理機器に自動切り替え（フェールオーバー）を行う。なお、N及びMは任意の整数である。例えば、ＯｐｅｎＳｔａｃｋ上で仮想ルータが動作している物理機器であれば、新しくＡｃｔ機になった物理機器のＮｅｕｔｒｏｎエージェントは、ＯｐｅｎＳｔａｃｋ ＤＢ上の構成情報を用いて、当該論理ホスト上に存在すべき仮想ルータを再構築する。再構築時間は、収容していた仮想ルータ数に比例した時間を要する。これは、仮想マシン等の場合も同様で、Ｎｏｖａの機能を用いて再構築される。このため、収容している仮想機器が多い場合は、全仮想機器が復旧するまでの時間が長くなり、サービス断時間が１時間超に及ぶなど、サービス断時間が大きい問題がある。

【０００９】

また、ＨＡクラスタソフトウェアのＰａｃｅｍａｋｅｒでは、パケットサイズの制限からＨＡクラスタの最大構成が８台程度までしか設定できない。ＨＡクラスタはＮ－Ａｃｔ、Ｍ－Ｓｂｙ構成であるため、クラスタあたり最低１台のＳｔａｎｄｂｙ機が必要である。そのため、予備機比率が１／８以上となり、物理機器の利用効率が悪い。

【００１０】

このように、ＨＡクラスタソフトウェアを用いた方法では、仮想リソースをＳｔａｎｄｂｙ機に復旧するまでの時間が長時間かかる問題や、Ｓｔａｎｄｂｙ機を準備する必要があり物理機器数が増える問題がある。

【００１１】
開示の技術は、上述に鑑みてなされたものであって、仮想機器を復旧するまでの時間を短縮することを目的とする。
【課題を解決するための手段】
【００１２】
本願の開示する仮想機器管理装置は、検出部と、選択部と、依頼部とを有する。検出部は、障害の生じた物理機器を検出する。選択部は、前記障害の生じた物理機器以外の物理機器のうち物理資源の空き容量のある物理機器を複数特定し、特定した複数の前記物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する。依頼部は、選択された前記物理機器それぞれに前記仮想機器のそれぞれの再配置を依頼する。
【００１３】
また、本願の開示する仮想機器管理方法は、検出工程と、選択工程と、依頼工程とを含む。検出工程は、障害の生じた物理機器を検出する。選択工程は、前記障害の生じた物理機器以外の物理機器のうち物理資源の空き容量のある物理機器を複数特定し、特定した複数の前記物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する。依頼工程は、選択された前記物理機器それぞれに前記仮想機器のそれぞれの再配置を依頼する。
【００１４】
また、本願の開示する仮想機器管理プログラムは、検出手順と、選択手順と、依頼手順とを有する。検出手順は、障害の生じた物理機器を検出する。選択手順は、前記障害の生じた物理機器以外の物理機器のうち物理資源の空き容量のある物理機器を複数特定し、特定した複数の前記物理機器の物理資源を、前記障害の生じた物理機器に配置された仮想機器の再配置先として選択する。依頼手順は、選択された前記物理機器それぞれに前記仮想機器のそれぞれの再配置を依頼する。
【発明の効果】
【００１５】
開示する仮想機器管理装置の一つの態様によれば、仮想機器を復旧するまでの時間を短縮することができるという効果を奏する。
【図面の簡単な説明】
【００１６】
【図１】図１は、第１の実施形態に係る仮想機器管理システムの構成の一例を示す図である。
【図２】図２は、仮想機器配置スケジューラ機能部による仮想機器の作成処理を説明するための図である。
【図３】図３は、仮想機器配置スケジューラ機能部による仮想機器の再配置処理を説明するための図である。
【図４】図４は、仮想機器管理システムにおける仮想機器を作成する処理動作を説明するための図である。
【図５】図５は、仮想機器管理システムにおける仮想機器を再配置する処理動作を説明するための図である。
【図６】図６は、仮想機器管理装置が実現する仮想機器配置スケジューラＤＢ及び仮想機器配置スケジューラ機能部を説明するための図である。
【図７】図７は、仮想機器配置情報テーブルのデータ構造の一例を示す図である。
【図８】図８は、物理資源情報テーブルのデータ構造の一例を示す図である。
【図９】図９は、配置先選択部による処理動作を説明するための図である。
【図１０】図１０は、障害検出部による処理動作を説明するための図である。
【図１１】図１１は、再配置先選択部による処理動作を説明するための図である。
【図１２】図１２は、仮想機器の作成を要求された場合の仮想機器配置スケジューラ機能部による処理手順を示すフローチャートである。

【図１３】図１３は、物理機器に障害が発生した場合の仮想機器配置スケジューラ機能部による処理手順を示すフローチャートである。
【図１４】図１４は、仮想機器管理プログラムを実行するコンピュータを示す図である。
【発明を実施するための形態】
【００１７】
以下に、開示する仮想機器管理装置、仮想機器管理方法及び仮想機器管理プログラムの実施形態について、図面に基づいて詳細に説明する。なお、本実施形態により開示する発明が限定されるものではない。
【００１８】
（第１の実施形態）
図１は、第１の実施形態に係る仮想機器管理システムの構成の一例を示す図である。図１に示すように、仮想機器管理システムは、ユーザ端末１０１、物理機器１０３ａ、物理機器１０３ｂ、物理機器１０３ｃ、クラウドコントローラ１０８、及び仮想機器管理装置１０９を有する。ここで言う「物理機器」とは、仮想機器を生成可能な物理サーバ、ストレージ装置、及びネットワーク機器等である。なお、物理機器１０３ａ、物理機器１０３ｂ及び物理機器１０３ｃを区別しない場合には、物理機器１０３と記載する。また、仮想機器管理システムが有する物理機器１０３の数は図１に示す数に限定されるものではなく、任意に変更可能である。
【００１９】
ユーザ端末１０１は、ユーザが利用する端末であり、ユーザの指示に応じて仮想機器の作成を仮想機器管理装置１０９に要求する。物理機器１０３は、クラウドコントローラ１０８から仮想機器の作成や削除依頼を受け、実際の仮想機器を作成したり削除したりする。例えば、物理機器１０３は、仮想機器を作成する指示をクラウドコントローラ１０８から受付け、仮想機器を作成する。
【００２０】
例えば、物理機器１０３ａは、図示しない仮想ボリューム制御部を有し、仮想ボリューム１０４ａと、仮想ボリューム１０５ａとを作成する。また、物理機器１０３ｂは、図示しない仮想ネットワーク制御部を有し、仮想Ｌ２ネットワーク１０４ｂと、仮想ルータ１０５ｂと、仮想ロードバランサ１０６ｂとを作成する。なお、仮想ネットワーク制御部は、例えば「Ｎｅｕｔｒｏｎ」によって実現される。また、物理機器１０３ｃは、図示しない仮想マシン制御部を有し、仮想マシン１０４ｃと、仮想マシン１０５ｃとを作成する。なお、仮想マシン制御部は、例えば「Ｎｏｖａ」によって実現される。
【００２１】
また、物理機器１０３の稼働状態には、「稼働中」、「予備」及び「故障中/メンテ中」３つの状態がある。「稼働中」は、物理機器が稼働中であることを示す。「予備」は、物理機器が予備系として設けられ稼働中ではないことを示す。「故障中/メンテ中」は、物理機器が故障中やメンテナンス中であることを示す。なお、仮想機器管理システムにおいて、「予備」の物理機器が設けられなくてもよい。
【００２２】
また、物理機器１０３には、物理資源の容量に応じて、仮想機器を配置するために利用可能な物理資源の容量が定義される。ここで、物理資源には、例えば、物理メモリ、ＣＰＵ（Central Processing Unit）、ネットワークポートなどが含まれる。なお、仮想マシンは、フレーバー（仮想マシンのスペック指定）に応じてメモリサイズが異なるため、作成する仮想マシンに応じて利用される物理資源の容量は異なる。しかしながら説明の便宜上、以下では、全ての仮想機器１つにつき、使用される物理資源の容量が同じであるものと仮定する。そして、１つの仮想機器を配置するために使用される物理資源の容量を１単位とし、「１スペース」と呼ぶ。言い換えると、１スペースには、１つの仮想機器を配置可能であり、１つの仮想機器を作成する場合には、いずれかの物理機器のスペースが１つ消費される。
【００２３】

また、物理機器内のスペースの状態は、「空き」、「使用中」、及び「障害用バッファ」の３種類で管理されるものとする。ここで、「空き」は、仮想機器が配置されていないスペースであることを示す。「使用中」は、仮想機器が配置されているスペースであることを示す。「障害用バッファ」は、障害復旧用に確保されたスペースであることを示す。

【００２４】

また、物理機器１０３ａは、高可用ソフトウェア１０７ａを備えている。同様に、物理機器１０３ｂは、高可用ソフトウェア１０７ｂを備えており、物理機器１０３ｃは、高可用ソフトウェア１０７ｃを備えている。なお、高可用ソフトウェア１０７ａ～１０７ｃを区別しない場合には高可用ソフトウェア１０７と記載する。この高可用ソフトウェア１０７には、例えば「Ｐａｃｅｍａｋｅｒ」等が利用できる。高可用ソフトウェア１０７は、物理機器１０３の障害を検知し、仮想機器管理装置１０９に物理機器の障害を通知する。かかる場合、物理機器１０３は、仮想機器を再配置させる指示をクラウドコントローラ１０８から受付け、障害の生じた物理機器１０３に配置された仮想機器を再配置する。

【００２５】

なお、「Ｐａｃｅｍａｋｅｒ」は、信頼性の高い故障検出メカニズムを備えており、スプリットブレイン対策が確立している。「Ｐａｃｅｍａｋｅｒ」は、スプリットブレイン状態（孤立状態）を、Ｑｕｏｒｕｍモジュール等による多数決原理で検出する。

【００２６】

クラウドコントローラ１０８は、物理機器１０３と仮想機器管理装置１０９とに接続されている。このクラウドコントローラ１０８は、ＣＰＵ（Central Processing Unit）、メモリ、データ保持領域、及びネットワーク通信機能を有する装置である。クラウドコントローラ１０８は、仮想機器管理装置１０９からＡＰＩ（Application Programming Interface）経由で仮想機器の作成依頼を受付け、受付けた作成依頼に基づいて、仮想機器の作成を物理機器１０３に指示する。例えば、クラウドコントローラ１０８は、ＯｐｅｎＳｔａｃｋ等である。

【００２７】

仮想機器管理装置１０９は、ユーザ端末１０１と物理機器１０３とクラウドコントローラ１０８とに接続されている。仮想機器管理装置１０９は、ＣＰＵ、メモリ、データ保持領域、及びネットワーク通信機能を有する装置であり、例えば、図１に示すように、仮想機器管理装置１０９は、仮想機器配置スケジューラＤＢ（Data Base）１１０及び仮想機器配置スケジューラ機能部１１１を有する。

【００２８】

仮想機器配置スケジューラＤＢ１１０は、例えば、ＲＡＭ（Random Access Memory）、フラッシュメモリ（Flash Memory）等の半導体メモリ素子、又は、ハードディスク、光ディスク等の記憶装置などである。仮想機器配置スケジューラＤＢ１１０は、仮想機器配置情報及び物理資源情報を記憶する。仮想機器配置情報は、各仮想機器がどの物理機器上に配置されているかを示す情報である。物理資源情報は、各物理機器の稼働状態と物理機器が有する物理資源の空き容量とを示す情報である。なお、仮想機器配置情報の詳細については、図７を用いて後述し、物理資源情報の詳細については、図８を用いて後述する。

【００２９】

仮想機器配置スケジューラ機能部１１１は、物理機器の稼働状態と物理機器が有する物理資源の使用状態とを参照して、ビジネス要件に応じた仮想機器を配置する。例えば、仮想機器配置スケジューラ機能部１１１は、ユーザ端末１０１から、仮想機器の作成を要求された場合、仮想機器の作成要求と仮想機器配置スケジューラＤＢ１１０の情報とを用いて、仮想機器の作成を仲介する。ここで、仮想機器配置スケジューラ機能部１１１は、仮想マシンや仮想ルータ等の仮想機器を新規に作成する通常のオペレーション時に、仮想機器を配置する物理機器を決め、クラウドコントローラ１０８に物理機器を指定して仮想機器の作成を依頼する。

【００３０】

図２は、仮想機器配置スケジューラ機能部１１１による仮想機器の作成処理を説明するための図である。図２では、３台の物理機器１０３ａ～１０３ｃに、仮想ルータであるＬＲ１１～ＬＲ１６、ＬＲ２１～ＬＲ２６、及びＬＲ３１～ＬＲ３６を仮想機器として作成する場合を示す。なお、３台の物理機器１０３ａ～１０３ｃはいずれも稼働中であるものとする。また、図２の例では、仮想機器が仮想ルータである場合を示すが、仮想機器は、仮想マシン等のその他の仮想機器であってもよい。

【００３１】

図２に示すように、仮想機器配置スケジューラ機能部１１１は、仮想機器ＬＲ１１～ＬＲ１６の配置先として物理機器１０３ａを選択し、仮想機器ＬＲ２１～ＬＲ２６の配置先として物理機器１０３ｂを選択し、仮想機器ＬＲ３１～ＬＲ３６の配置先として物理機器１０３ｃを選択する。そして、仮想機器配置スケジューラ機能部１１１は、クラウドコントローラ１０８に配置を依頼する。すなわち、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ａに仮想機器ＬＲ１１～ＬＲ１６を作成するようにクラウドコントローラ１０８に依頼する。また、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｂに仮想機器ＬＲ２１～ＬＲ２６を作成するようにクラウドコントローラ１０８に依頼し、物理機器１０３ｃに仮想機器ＬＲ３１～ＬＲ３６を作成するようにクラウドコントローラ１０８に依頼する。

【００３２】

また、仮想機器配置スケジューラ機能部１１１は、例えば、いずれかの物理機器１０３に障害が生じた場合に、仮想機器配置スケジューラＤＢ１１０の情報を用いて、仮想機器の再配置を仲介する。ここで、仮想機器配置スケジューラ機能部１１１は、高可用ソフトウェア１０７及びクラウドコントローラ１０８と連携することで障害復旧時に仮想機器を再配置する。図３は、仮想機器配置スケジューラ機能部１１１による仮想機器の再配置処理を説明するための図である。図３では、３台の物理機器１０３ａ～１０３ｃのうち、仮想機器としてＬＲ２１～ＬＲ２６を配置する物理機器１０３ｂに障害が生じた場合を示す。なお、３台の物理機器１０３ａ～１０３ｃはいずれも稼働中であるものとする。

【００３３】

図３に示すように、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｂに障害が生じたことを検出する。そして、仮想機器配置スケジューラ機能部１１１は、仮想機器配置スケジューラＤＢ１１０の情報を用いて、仮想機器ＬＲ２１～ＬＲ２６の再配置先を決定する。図３に示す例では、仮想機器配置スケジューラ機能部１１１は、ＬＲ２１、ＬＲ２３、及びＬＲ２５の再配置先として物理機器１０３ａを選択し、ＬＲ２２、ＬＲ２４、及びＬＲ２６の再配置先として物理機器１０３ｃを選択する。

【００３４】

そして、仮想機器配置スケジューラ機能部１１１は、クラウドコントローラ１０８に再配置を依頼する。すなわち、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ａに仮想機器ＬＲ２１、ＬＲ２３、及びＬＲ２５を作成するようにクラウドコントローラ１０８に依頼する。また、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｃに仮想機器ＬＲ２２、ＬＲ２４、及びＬＲ２６を作成するようにクラウドコントローラ１０８に依頼する。

【００３５】

この結果、物理機器１０３ａは、仮想機器ＬＲ２１、ＬＲ２３、及びＬＲ２５を再構築し、仮想機器ＬＲ１１～ＬＲ１６に加えて、仮想機器ＬＲ２１、ＬＲ２３、及びＬＲ２５を配置する。また、物理機器１０３ｃは、仮想機器ＬＲ２２、ＬＲ２４、及びＬＲ２６を再構築し、仮想機器ＬＲ３１～ＬＲ３６に加えて、仮想機器ＬＲ２２、ＬＲ２４、及びＬＲ２６を配置する。

【００３６】

このように、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｂに障害が生じた場合、物理機器１０３ｂに配置された仮想機器ＬＲ２１～ＬＲ２６を、物理機器１０３ａと物理機器１０３ｃとに再配置する。すなわち、仮想機器配置スケジューラ機能部１

１１は、複数台の物理機器を仮想機器の復旧先として利用するので、物理機器故障時の仮想機器復旧時間を短縮できる。

【００３７】

続いて、このような仮想機器管理システムにおける処理動作について、図４及び図５を用いて説明する。図４は、仮想機器管理システムにおける仮想機器を作成する処理動作を説明するための図である。

【００３８】

図４に示すように、ユーザ端末１０１は、仮想機器作成依頼を、仮想機器配置スケジューラ機能部１１１に送信する（ステップＳ１）。仮想機器配置スケジューラ機能部１１１は、仮想機器配置スケジューラＤＢ１１０を参照し（ステップＳ２）、物理資源情報を確認する（ステップＳ３）。これにより仮想機器配置スケジューラ機能部１１１は、仮想機器を作成する物理機器１０３を決定し、ＡＰＩパラメータを準備する（ステップＳ４）。

【００３９】

次に、仮想機器配置スケジューラ機能部１１１は、決定した物理機器１０３に仮想機器を作成させるようにクラウドコントローラ１０８に依頼する（ステップＳ５）。続いて、クラウドコントローラ１０８は、物理機器１０３に仮想機器の作成を依頼する（ステップＳ６）。

【００４０】

そして、物理機器１０３は、仮想機器を作成し（ステップＳ７）、仮想機器の作成が完了したことをクラウドコントローラ１０８に通知する（ステップＳ８）。続いて、クラウドコントローラ１０８は、仮想機器の作成が完了したことを仮想機器配置スケジューラ機能部１１１に通知する（ステップＳ９）。そして、仮想機器配置スケジューラ機能部１１１は、仮想機器の作成が完了したことをユーザ端末１０１に通知する（ステップＳ１０）。

【００４１】

図５は、仮想機器管理システムにおける仮想機器を再配置する処理動作を説明するための図である。図５では、いずれかの物理機器１０３に障害が生じた場合に、仮想機器配置スケジューラ機能部１１１が仮想機器の再配置を仲介する動作を説明する。図５に示すように、仮想機器管理システムでは、物理機器１０３ａ、物理機器１０３ｂ及び物理機器１０３ｃが相互に機器状態を監視している（ステップＳ２１、ステップＳ２２）。以下では、物理機器１０３ａに障害が生じた場合について説明する。ここで、物理機器１０３ａで障害が起きた際は、物理機器１０３ａ上の高可用ソフトウェアは物理機器１０３ａ上のプロセスを停止し、障害を仮想機器配置スケジューラ機能部１１１に通知する。物理機器１０３ｂ及び物理機器１０３ｃも同様に物理機器１０３ａの障害を仮想機器配置スケジューラ機能部１１１に通知する。ここで、物理機器１０３ａが完全に故障している場合は、物理機器１０３ａから仮想機器配置スケジューラ機能部１１１に障害の発生を通知はできないが、物理機器１０３ｂ及び物理機器１０３ｃは、物理機器１０３ａの故障を仮想機器配置スケジューラ機能部１１１に通知できる。このため、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ａの故障を知ることができる。なお図５に示す例では、物理機器１０３ａが完全に故障し、物理機器１０３ａから障害の発生を仮想機器配置スケジューラ機能部１１１に通知できない場合を示す。

【００４２】

かかる場合、物理機器１０３ｂは、物理機器１０３ａに障害が生じたことを仮想機器配置スケジューラ機能部１１１に通知する（ステップＳ２３）。そして、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｂにＡＣＫ（ACKnowledgement）を応答する（ステップＳ２４）。同様に、物理機器１０３ｃは、物理機器１０３ａに障害が生じたことを仮想機器配置スケジューラ機能部１１１に通知する（ステップＳ２５）。そして、仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｃにＡＣＫを応答する（ステップＳ２６）。ここで、仮想機器配置スケジューラ機能部１１１は、最初に受信した通知に従って仮想機器の再配置を始めるが、２番目以降に受信した通知に対してもＡＣＫを応答す

る。
【００４３】
仮想機器配置スケジューラ機能部１１１は、仮想機器配置スケジューラＤＢ１１０を参照し（ステップＳ２７）、物理資源情報を確認する（ステップＳ２８）。これにより仮想機器配置スケジューラ機能部１１１は、物理機器１０３ｂ及び物理機器１０３ｃの物理資源の空き容量を取得して、仮想機器を再配置する物理機器１０３を決定し、ＡＰＩパラメータを準備する（ステップＳ２９）。ここで、仮想機器配置スケジューラ機能部１１１は、複数台の物理機器１０３を仮想機器の復旧先として選択することで、高速の復旧を可能とする。
【００４４】
次に、仮想機器配置スケジューラ機能部１１１は、再配置する仮想機器を物理機器１０３ｂに作成させるようにクラウドコントローラ１０８に依頼する（ステップＳ３０）。続いて、クラウドコントローラ１０８は、物理機器１０３ｂに仮想機器の作成を依頼する（ステップＳ３１）。同様に、仮想機器配置スケジューラ機能部１１１は、再配置する仮想機器を物理機器１０３ｃに作成させるようにクラウドコントローラ１０８に依頼する（ステップＳ３２）。続いて、クラウドコントローラ１０８は、物理機器１０３ｃに仮想機器の作成を依頼する（ステップＳ３３）。ここで、仮想機器配置スケジューラ機能部１１１は、選択した配置先を指定してクラウドコントローラ１０８のＡＰＩを呼び出す。これにより、クラウドコントローラ１０８は、指定された物理機器１０３に対して仮想機器作成を依頼する。
【００４５】
そして、物理機器１０３ｂは、仮想機器を作成し（ステップＳ３４）、仮想機器の作成が完了したことをクラウドコントローラ１０８に通知する（ステップＳ３５）。続いて、クラウドコントローラ１０８は、仮想機器の作成が完了したことを仮想機器配置スケジューラ機能部１１１に通知する（ステップＳ３６）。同様に、物理機器１０３ｃは、仮想機器を作成し（ステップＳ３７）、仮想機器の作成が完了したことをクラウドコントローラ１０８に通知する（ステップＳ３８）。続いて、クラウドコントローラ１０８は、仮想機器の作成が完了したことを仮想機器配置スケジューラ機能部１１１に通知する（ステップＳ３９）。
【００４６】
続いて、図６を用いて、仮想機器管理装置１０９が実現する仮想機器配置スケジューラＤＢ１１０及び仮想機器配置スケジューラ機能部１１１について説明する。図６は、仮想機器管理装置１０９が実現する仮想機器配置スケジューラＤＢ１１０及び仮想機器配置スケジューラ機能部１１１を説明するための図である。
【００４７】
図６に示すように、仮想機器配置スケジューラＤＢ１１０は、仮想機器配置情報テーブル１１０ａ及び物理資源情報テーブル１１０ｂを記憶する。仮想機器配置情報テーブル１１０ａは、各仮想機器がどの物理機器上に配置されているかを示す仮想機器配置情報を記憶する。
【００４８】
図７は、仮想機器配置情報テーブル１１０ａのデータ構造の一例を示す図である。図７に示すように、仮想機器配置情報テーブル１１０ａは、「仮想機器ＩＤ」と、「物理機器ＩＤ」とを対応付けた仮想機器配置情報を記憶する。ここで、仮想機器配置情報テーブル１１０ａが記憶する「仮想機器ＩＤ」は、物理機器１０３に作成された仮想機器を一意に識別する識別子を示す。例えば、「仮想機器ＩＤ」には、「仮想ボリューム＃１」、「仮想ボリューム＃２」等のデータ値が格納される。仮想機器配置情報テーブル１１０ａが記憶する「物理機器ＩＤ」は、物理機器１０３を一意に識別する識別子を示す。例えば、「物理機器ＩＤ」には、「物理機器＃１」、「物理機器＃２」等のデータ値が格納される。
【００４９】
一例をあげると、図７に示す仮想機器配置情報テーブル１１０ａは、識別子が「物理機

器＃１」である物理機器１０３には、仮想機器「仮想ボリューム＃１」及び「仮想ボリューム＃２」が配置されていることを示す。また、図７に示す仮想機器配置情報テーブル１１０ａは、識別子が「物理機器＃２」である物理機器１０３には、仮想機器「仮想Ｌ２ネットワーク＃１」、「仮想ルータ＃１」及び「仮想ロードバランサ＃１」が配置されていることを示す。また、図７に示す仮想機器配置情報テーブル１１０ａは、識別子が「物理機器＃３」である物理機器１０３には、仮想機器「仮想マシン＃１」及び「仮想マシン＃２」が配置されていることを示す。

【００５０】

図６に戻る。物理資源情報テーブル１１０ｂは、各物理機器の稼働状態と物理機器が有する物理資源の空き容量とを示す物理資源情報を記憶する。図８は、物理資源情報テーブル１１０ｂのデータ構造の一例を示す図である。図８に示すように、物理資源情報テーブル１１０ｂは、「物理機器ＩＤ」と「稼働状態」と「空き」と「使用中」と「障害用」とを対応付けた物理資源情報を記憶する。

【００５１】

ここで、物理資源情報テーブル１１０ｂが記憶する「物理機器ＩＤ」は、物理機器１０３を一意に識別する識別子を示す。例えば、「物理機器ＩＤ」には、「物理機器＃１」、「物理機器＃２」等のデータ値が格納される。

【００５２】

また、物理資源情報テーブル１１０ｂが記憶する「稼働状態」は、物理機器が稼働中であるか否かを示す。例えば、物理機器が稼働中である場合、「稼働状態」には「稼働中」が格納される。なお、図８では図示していないが、物理機器が予備系として設けられ稼働中ではない場合、「稼働状態」には「予備」が格納される。また、物理機器が故障中やメンテナンス中である場合、「稼働状態」には「故障中／メンテ中」が格納される。

【００５３】

また、「空き」は、物理機器が有する物理資源の容量のうち空き容量を示す。例えば、「空き」には、「３」、「５」、「４」等の値が格納される。また、「使用中」は、物理機器が有する物理資源の容量のうち使用中の容量を示す。例えば、「使用中」には、「１」、「３」、「２」等の値が格納される。また、「障害用」は、物理機器が有する物理資源の容量のうち復旧用に確保された容量を示す。例えば、「障害用」には、「２」等の値が格納される。

【００５４】

一例をあげると、図８に示す物理資源情報テーブル１１０ｂは、物理機器＃１は、稼働中であり、物理資源の空き容量が「３」であり、使用中の容量が「１」であり、復旧用に確保された容量が「２」であることを示す。また、図８に示す物理資源情報テーブル１１０ｂは、物理機器＃２は、稼働中であり、物理資源の空き容量が「５」であり、使用中の容量が「３」であり、復旧用に確保された容量が「２」であることを示す。同様に、図８に示す物理資源情報テーブル１１０ｂは、物理機器＃３は、稼働中であり、物理資源の空き容量が「４」であり、使用中の容量が「２」であり、復旧用に確保された容量が「２」であることを示す。

【００５５】

図６に戻る。仮想機器配置スケジューラ機能部１１１は、作成依頼受付部１１１ａと、配置先選択部１１１ｂと、障害検出部１１１ｃと、再配置先選択部１１１ｄと、作成要求部１１１ｅとを有する。

【００５６】

作成依頼受付部１１１ａは、仮想機器の作成要求をユーザ端末１０１から受付ける。作成依頼受付部１１１ａは、受付けた仮想機器の作成要求を配置先選択部１１１ｂに受け渡す。

【００５７】

配置先選択部１１１ｂは、仮想機器を新規に作成する際に、仮想機器を配置する物理機器を選択する。ここで、配置先選択部１１１ｂは、仮想機器を出来るだけ分散して配置す

るように物理機器１０３を選択する。言い換えると、配置先選択部１１１ｂは、「稼働中」の「空き」スペースの数が平準化するように仮想機器を配置する。

【００５８】

図９は、配置先選択部１１１ｂによる処理動作を説明するための図である。図９では、物理機器＃１～物理機器＃６の６台の物理機器を有する仮想機器管理システムにおいて、仮想機器を新規に作成する場合について説明する。ここで、物理機器＃１～物理機器＃５の稼働状態は「稼働中」であり、物理機器＃６の稼働状態は「予備」である。また、物理機器＃１のスペースの状態は、「空き」２、「使用中」１、「障害用バッファ」２であり、物理機器＃２のスペースの状態は、「空き」０、「使用中」３、「障害用バッファ」２であり、物理機器＃３のスペースの状態は、「空き」４、「使用中」０、「障害用バッファ」２である。また、物理機器＃４のスペースの状態は、「空き」０、「使用中」５、「障害用バッファ」２であり、物理機器＃５のスペースの状態は、「空き」２、「使用中」０、「障害用バッファ」２であり、物理機器＃６のスペースの状態は、「空き」３、「使用中」０、「障害用バッファ」２である。

【００５９】

例えば、配置先選択部１１１ｂは、配置先選択時に、稼働状態が「稼働中」である物理機器の空きスペースの量をチェックし、最も空きスペースが多い稼働中の物理機器を特定する。より具体的には、配置先選択部１１１ｂは、作成する仮想機器のうち１つの仮想機器（例えば、仮想機器＃１）を選択する。そして、図９に示すスペースの状態である場合には、「空き」が４である物理機器＃３を、最も空きスペースが多い稼働中の物理機器に特定する。なお、配置先選択部１１１ｂは、「障害用バッファ」を通常オペレーション時には利用しない。そして、配置先選択部１１１ｂは、特定した物理機器＃３を選択した仮想機器＃１の配置先として選択する。

【００６０】

続いて、配置先選択部１１１ｂは、配置先として選択する処理を、作成を依頼された全ての仮想機器の配置先を選択するまで繰り返す。一例をあげると、配置先選択部１１１ｂは、図９に示す数字順に仮想機器を配置するように物理機器を選択する。このように、配置先選択部１１１ｂは、最も空きスペースが多い稼働中の物理機器のスペースの一部を選択することで「空き」スペースの数を平準化する。

【００６１】

また、配置先選択部１１１ｂは、稼働状態が「稼働中」である物理機器のスペースが全て埋まった場合に、稼働状態が「予備」である物理機器に仮想機器を配置する。このため、配置先選択部１１１ｂは、図９に示す８番のスペースまで仮想機器を配置したら、予備の物理機器に仮想機器を配置する。すなわち、配置先選択部１１１ｂは、図９に示す例において、仮想機器を９台以上作成する場合には、稼働状態が「予備」である物理機器に仮想機器を配置する。

【００６２】

図６に戻る。障害検出部１１１ｃは、障害の生じた物理機器１０３を検出する。ここで、障害検出部１１１ｃは、各物理機器１０３が有する高可用ソフトウェア１０７と連携することで、障害の生じた物理機器１０３を検出する。図１０は、障害検出部１１１ｃによる処理動作を説明するための図である。

【００６３】

図１０では、物理機器１０３ａ～１０３ｃを図示しており、物理機器１０３ａに障害が発生した場合について説明する。また、図１０では、物理機器１０３が有する機能のうち、物理機器１０３ａには、自装置の障害発生時に機能する構成部を示し、物理機器１０３ｂ及び物理機器１０３ｃには、他装置の障害を検知した場合に機能する構成部を示す。

【００６４】

図１０に示すように、物理機器１０３の障害の検知には、高可用ソフトウェア１０７ａ～１０７ｃが用いられる。全ての物理機器１０３は、ＣＩＢ（Cluster Information Base）に、クラスタ内の全物理機器１０３の状態を保持する。高可用ソフトウェア１０７は

、ＲＡ（Resource Agent）を用いて自物理機器の状態を確認する。なお、ＲＡとは、例えば、仮想ボリューム制御部や仮想マシン制御部に相当する。

【００６５】

また、高可用ソフトウェア１０７は、Ｈｅａｒｔｂｅａｔにより、クラスタ内のどの物理機器も他の物理機器の状態を知り得る。このため、高可用ソフトウェア１０７は、ｈｅａｒｔｂｅａｔパケットを使ってクラスタ内に状態を通知する。この仕組みにより、各物理機器は他の物理機器の状態を知る。高可用ソフトウェア１０７は、ある物理機器からのｈｅａｒｔｂｅａｔパケットが継続的にロストすると、他の物理機器は当該物理機器がダウンしたとみなす。

【００６６】

物理機器１０３は、仮想機器配置スケジューラ機能部１１１に物理機器に生じた障害を通知するため、通知ＲＡと通知プロセスとを備える。例えば、Ｐａｃｅｍａｋｅｒが自物理機器の故障を検出した場合、通知ＲＡを使用して仮想機器配置スケジューラ機能部１１１に自物理機器の故障を通知する。一方、通知プロセスは、常駐プロセスとして設定され、ＣＩＢの状態を定期的に確認する。そして、通知プロセスは、他物理機器の故障を検出すると、他物理機器に障害が生じたことを仮想機器配置スケジューラ機能部１１１に通知する。通知ＲＡによる通知及び通知プロセスによる通知は、ＡＣＫが仮想機器配置スケジューラ機能部１１１から返るまで一定回数繰り返される。

【００６７】

続いて、仮想機器配置スケジューラ機能部１１１において、障害検出部１１１ｃは、通知を受信したらＡＣＫを応答し、再配置先選択部１１１ｄに仮想機器の再配置処理を実行させる。また、障害検出部１１１ｃは、２通目以降の通知を無視してＡＣＫを応答する。これにより、複数の物理機器から通知を受けることで冗長化対策をとることができるとともに、復旧処理を繰り返さないようにする。

【００６８】

なお、高可用ソフトウェア１０７は、自物理機器の停止に失敗する場合がある。仮想マシンの場合、復旧により、複数の仮想マシンが同時に存在してしまい、データ領域への同時アクセスによりデータ破壊の可能性が出てしまう。そこで、高可用ソフトウェア１０７が「Ｐａｃｅｍａｋｅｒ」である場合、ＳＴＯＮＩＴＨモジュールを用いて、確実に故障物理機器を落とす。ＳＴＯＮＩＴＨは、ＩＰＭＩ（Intelligent Platform Management Interface）経由で、故障物理機器を停止することで、故障物理機器が動作し続けないことを保証する。Ｑｕｏｒｕｍで過半数を形成した多数派の物理機器が、ＳＴＯＮＩＴＨを起動することで、誤発動を防止する。なお、Ｑｕｏｒｕｍは過半数で判断するため、クラスタの物理機器数が少ない場合に、ある物理機器が故障したら、正常な物理機器が過半数を確保できなくなる。このため、クラスタから故障物理機器を切り離す減設作業が必要である。また、図１０では、高可用ソフトウェア１０７が、Ｐａｃｅｍａｋｅｒである場合を示しているが、他の高可用ソフトウェアでも同様のメカニズムで障害の発生を検知したり、障害の発生を通知したりすることが可能である。

【００６９】

図６に戻る。再配置先選択部１１１ｄは、物理機器に生じた障害を通知された場合に、仮想機器を再作成する物理機器を選択する。ここで、再配置先選択部１１１ｄは、障害の生じた物理機器以外の物理機器にできるだけ順番に割り振られるように物理機器を選択する。例えば、再配置先選択部１１１ｄは、障害の生じた物理機器以外の物理機器のうち物理資源の空き容量のある物理機器を複数特定する。そして、再配置先選択部１１１ｄは、特定した複数の物理機器の物理資源を、障害の生じた物理機器に配置された仮想機器の再配置先として選択する。

【００７０】

図１１は、再配置先選択部１１１ｄによる処理動作を説明するための図である。図１１では、物理機器＃１～物理機器＃６の６台の物理機器を有する仮想機器管理システムにおいて、物理機器＃４が故障した際の復旧について説明する。ここで、物理機器＃１～物理

機器＃５の稼働状態は「稼働中」であり、物理機器＃６の稼働状態は「予備」である。また、物理機器＃１のスペースの状態は、「空き」２、「使用中」１、「障害用バッファ」２であり、物理機器＃２のスペースの状態は、「空き」０、「使用中」３、「障害用バッファ」２であり、物理機器＃３のスペースの状態は、「空き」４、「使用中」０、「障害用バッファ」２である。また、物理機器＃４のスペースの状態は、「空き」０、「使用中」１０、「障害用バッファ」２であり、物理機器＃５のスペースの状態は、「空き」２、「使用中」０、「障害用バッファ」２であり、物理機器＃６のスペースの状態は、「空き」３、「使用中」０、「障害用バッファ」２である。

【００７１】

例えば、再配置先選択部１１１ｄは、障害の生じた物理機器以外の物理機器のうち物理資源の空き容量のある物理機器を複数特定する。ここで、再配置先選択部１１１ｄは、物理機器の障害発生時には、「空き」のスペースに加えて、「障害用バッファ」のスペースも使用する。これにより、再配置先選択部１１１ｄは、「空き」が０の物理機器も含めて、より多くの物理機器が復旧処理を分担できるようにする。図１１に示す例では、再配置先選択部１１１ｄが、稼働中である物理機器＃１～物理機器＃３及び物理機器＃５を特定した場合を示す。

【００７２】

そして、再配置先選択部１１１ｄは、特定した複数の物理機器の物理資源を、障害の生じた物理機器＃４に配置された仮想機器の再配置先として選択する。例えば、再配置先選択部１１１ｄは、特定した複数の物理機器に順序付けを行う。ここで、再配置先選択部１１１ｄは、特定した複数の物理機器の物理資源の空き容量が多い順に、特定した複数の物理機器に順序付けを行う。例えば、再配置先選択部１１１ｄは、「空き」のスペースと「障害用バッファ」のスペースとの合計スペースを物理資源の空き容量とし、合計スペースが多い順に物理機器に順序付けを行う。図１１の例では、再配置選択部１１１ｄが、合計スペースが６である物理機器＃３、合計スペースが５である物理機器＃１、合計スペースが４である物理機器＃５、そして、合計スペースが２である物理機器＃２の順で順序付けした場合を示す。

【００７３】

続いて、再配置先選択部１１１ｄは、障害の生じた物理機器に配置された仮想機器それぞれの再配置先として、順序に基づいて選択した物理機器の物理資源を選択する処理を繰り返す。一例をあげると、再配置先選択部１１１ｄは、図１１に示す数字順に仮想機器を再配置するように物理機器を選択する。より具体的には、再配置選択部１１１ｄは、物理機器＃３、物理機器＃１、物理機器＃５、そして、物理機器＃２の順で選択した物理機器の物理資源を仮想機器の再配置先として選択する処理を繰り返す。ここで、再配置先選択部１１１ｄは、「空き」のスペースや「障害用バッファ」のスペースが無くなるまでは、各物理機器に仮想機器を順番に配置する。また、再配置先選択部１１１ｄは、スペースが無くなった物理機器は飛ばすようにする。

【００７４】

なお、再配置先選択部１１１ｄは、稼働状態が「稼働中」である全ての物理機器の「空き」のスペース及び「障害用バッファ」のスペースが満たされるまで、稼働状態が「予備」である物理機器を選択しない。このように、仮想機器配置スケジューラ機能部１１１は、「空き」のスペースに加えて、仮想機器の作成時には利用されない「障害用バッファ」のスペースを予め準備しておき、障害時に多くの物理機器に仮想機器を再配置することで、高速の復旧を可能とする。また、再配置選択部１１１ｄは、障害が発生した物理機器に配置された仮想機器の全てを再配置可能ではない場合には、特定した物理機器の「空き」のスペースと「障害用バッファ」のスペースとに再配置可能な範囲で、仮想機器ごとに再配置先を選択する。

【００７５】

また、Ｐａｃｅｍａｋｅｒのクラスタ構成は、最大８台程度で組み、障害の検知を行う。また、仮想機器配置スケジューラ機能部１１１は、クラスタを跨いで別物理機器に仮想

機器を作成してもよいため、再作成が依頼される物理機器はクラスタのサイズ以上でも良い。また、全てが埋まった際に利用される予備機は存在してもしなくてもよい。クラスタ構成上はＮ－Ａｃｔ、０－Ｓｂｙで、Ｓｔａｎｄｂｙ機を準備する必要はないため、物理機器の利用効率を高めることも出来る。

【００７６】

図６に戻る。作成要求部１１１ｅは、配置先選択部１１１ｂにより選択された物理機器１０３に、仮想機器を作成するようにクラウドコントローラ１０８に依頼する。また、作成要求部１１１ｅは、再配置先選択部１１１ｄにより選択された物理機器１０３に、仮想機器を作成するようにクラウドコントローラ１０８に依頼する。

【００７７】

図１２は、仮想機器の作成を要求された場合の仮想機器配置スケジューラ機能部１１１による処理手順を示すフローチャートである。図１２に示すように、作成依頼受付部１１１ａは、仮想機器の作成をユーザ端末１０１から依頼されたか否かを判定する（ステップＳ１０１）。ここで、作成依頼受付部１１１ａは、仮想機器の作成をユーザ端末１０１から依頼されたと判定した場合（ステップＳ１０１、Ｙｅｓ）、作成を依頼された仮想機器を特定する（ステップＳ１０２）。なお、作成依頼受付部１１１ａは、仮想機器の作成をユーザ端末１０１から依頼されたと判定しなかった場合（ステップＳ１０１、Ｎｏ）、繰り返し物理機器の作成をユーザ端末１０１から依頼されたか否かを判定する。

【００７８】

続いて、配置先選択部１１１ｂは、作成を依頼された仮想機器を１つ選択する（ステップＳ１０３）。そして、配置先選択部１１１ｂは、空き容量の最も多い物理機器を特定する（ステップＳ１０４）。また、配置先選択部１１１ｂは、空き容量の最も多い物理機器を特定できたか否かを判定する（ステップＳ１０５）。ここで、配置先選択部１１１ｂは、空き容量の最も多い物理機器を特定できたと判定した場合（ステップＳ１０５、Ｙｅｓ）、ステップＳ１０９に移行する。一方、配置先選択部１１１ｂは、空き容量の最も多い物理機器を特定できたと判定しなかった場合（ステップＳ１０５、Ｎｏ）、予備系の物理機器が存在するか否かを判定する（ステップＳ１０６）。ここで、配置先選択部１１１ｂは、予備系の物理機器が存在すると判定しなかった場合（ステップＳ１０６、Ｎｏ）、処理を終了する。

【００７９】

一方、配置先選択部１１１ｂは、予備系の物理機器が存在すると判定した場合（ステップＳ１０６、Ｙｅｓ）、空き容量の最も多い予備系の物理機器を特定する（ステップＳ１０７）。また、配置先選択部１１１ｂは、空き容量の最も多い予備系の物理機器を特定できたか否かを判定する（ステップＳ１０８）。ここで、配置先選択部１１１ｂは、空き容量の最も多い予備系の物理機器を特定できたと判定しなかった場合（ステップＳ１０８、Ｎｏ）、処理を終了する。一方、配置先選択部１１１ｂは、空き容量の最も多い予備系の物理機器を特定できたと判定した場合（ステップＳ１０８、Ｙｅｓ）、ステップＳ１０９に移行する。

【００８０】

ステップＳ１０９において、配置先選択部１１１ｂは、特定した物理機器を選択した仮想機器の配置先に選択する（ステップＳ１０９）。そして、配置先選択部１１１ｂは、作成を依頼された全ての仮想機器の配置先を選択したか否かを判定する（ステップＳ１１０）。ここで、配置先選択部１１１ｂは、作成を依頼された全ての仮想機器の配置先を選択したと判定しなかった場合（ステップＳ１１０、Ｎｏ）、作成を依頼された全ての仮想機器の配置先を選択するまでステップＳ１０３からステップＳ１１０までの処理を繰り返し実行する。

【００８１】

一方、配置先選択部１１１ｂは、作成を依頼された全ての仮想機器の配置先を選択したと判定した場合（ステップＳ１１０、Ｙｅｓ）、クラウドコントローラ１０８に配置を依頼し（ステップＳ１１１）、処理を終了する。

【００８２】
図１３は、物理機器１０３に障害が発生した場合の仮想機器配置スケジューラ機能部１１１による処理手順を示すフローチャートである。図１３に示すように、障害検出部１１１ｃは、物理機器の障害を通知されたか否かを判定する（ステップＳ２０１）。ここで、障害検出部１１１ｃは、物理機器の障害を通知されたと判定した場合（ステップＳ２０１、Ｙｅｓ）、仮想機器配置情報テーブル１１０ａを参照して、障害が発生した物理機器に配置された仮想機器を特定する（ステップＳ２０２）。なお、障害検出部１１１ｃは、物理機器の障害を通知されたと判定しなかった場合（ステップＳ２０１、Ｎｏ）、繰り返し物理機器の障害を通知されたか否かを判定する。
【００８３】
続いて、再配置先選択部１１１ｄは、空き容量及び障害用容量の少なくともいずれかがある物理機器を複数特定する（ステップＳ２０３）。そして、再配置先選択部１１１ｄは、複数の物理機器を特定できたか否かを判定する（ステップＳ２０４）。ここで、再配置先選択部１１１ｄは、複数の物理機器を特定できたと判定しなかった場合（ステップＳ２０４、Ｎｏ）、予備系の物理機器が存在するか否かを判定する（ステップＳ２０５）。ここで、再配置先選択部１１１ｄは、予備系の物理機器が存在すると判定しなかった場合（ステップＳ２０５、Ｎｏ）、ステップＳ２０７に移行する。
【００８４】
一方、再配置先選択部１１１ｄは、予備系の物理機器が存在すると判定した場合（ステップＳ２０５、Ｙｅｓ）、空き容量及び障害用容量の少なくともいずれかがある予備系の物理機器を特定する（ステップＳ２０６）。そして、再配置先選択部１１１ｄは、稼働中及び予備系を合わせて１以上の物理機器を特定できたか否かを判定する（ステップＳ２０７）。ここで、再配置先選択部１１１ｄは、稼働中及び予備系を合わせて１以上の物理機器を特定できたと判定した場合（ステップＳ２０７、Ｙｅｓ）、ステップＳ２１１に移行する。一方、再配置先選択部１１１ｄは、稼働中及び予備系を合わせて１以上の物理機器を特定できたと判定しなかった場合（ステップＳ２０７、Ｎｏ）、処理を終了する。
【００８５】
再配置先選択部１１１ｄは、ステップＳ２０４において、複数の物理機器を特定できたと判定した場合（ステップＳ２０４、Ｙｅｓ）、特定した複数の物理機器に、障害が発生した物理機器に配置された仮想機器を全て再配置可能であるか否かを判定する（ステップＳ２０８）。ここで、再配置先選択部１１１ｄは、障害が発生した物理機器に配置された仮想機器を全て再配置可能であると判定した場合（ステップＳ２０８、Ｙｅｓ）、ステップＳ２１１に移行する。一方、再配置先選択部１１１ｄは、障害が発生した物理機器に配置された仮想機器を全て再配置可能であると判定しなかった場合（ステップＳ２０８、Ｎｏ）、予備系の物理機器が存在するか否かを判定する（ステップＳ２０９）。ここで、再配置先選択部１１１ｄは、予備系の物理機器が存在すると判定した場合（ステップＳ２０９、Ｙｅｓ）、空き容量及び障害用容量の少なくともいずれかがある予備系の物理機器を特定する（ステップＳ２１０）。
【００８６】
ステップＳ２１１において、再配置先選択部１１１ｄは、特定した物理機器に順序付けを行う（ステップＳ２１１）。例えば、再配置先選択部１１１ｄは、空き容量及び障害用容量が多い順に、特定した物理機器に順序付けを行う。なお、再配置先選択部１１１ｄは、特定した物理機器が１つである場合には、ステップＳ２１１の処理を省略してもよい。
【００８７】
続いて、再配置先選択部１１１ｄは、仮想機器の再配置先を選択する（ステップＳ２１２）。例えば、再配置先選択部１１１ｄは、障害の生じた物理機器に配置された仮想機器それぞれの再配置先として、順序に基づいて選択した物理機器の物理資源を選択する処理を繰り返す。なお、再配置先選択部１１１ｄは、障害が発生した物理機器に配置された仮想機器の全てを再配置可能ではない場合には、特定した物理機器の空き容量及び障害用容量に再配置可能な範囲で、仮想機器ごとに再配置先を選択する。

【００８８】
そして、再配置先選択部１１１ｄは、クラウドコントローラ１０８に再配置を依頼し（ステップＳ２１３）、処理を終了する。なお、再配置先選択部１１１ｄは、仮想機器管理システムにおいて予備系の物理機器がない場合には、ステップＳ２０５、ステップＳ２０６、ステップＳ２０９及びステップＳ２１０の処理を省略してもよい。
【００８９】
上述したように、第１の実施形態に係る仮想機器管理装置１０９は、障害の生じた物理機器以外の物理機器のうち物理資源の空き容量のある物理機器を複数特定する。そして、第１の実施形態に係る仮想機器管理装置１０９は、特定した複数の物理機器の物理資源を、障害の生じた物理機器に配置された仮想機器の再配置先として選択する。すなわち、第１の実施形態に係る仮想機器管理装置１０９は、複数台の物理機器を仮想機器の復旧先として利用する。これにより、第１の実施形態に係る仮想機器管理装置１０９は、仮想機器を復旧するまでの時間を短縮することができる。
【００９０】
より具体的には、従来方式では、Ｎ－Ａｃｔ、Ｍ－Ｓｂｙでクラスタを組み物理機器に障害が起きた際に、Ｐａｃｅｍａｋｅｒ等の高可用ソフトウェアの機能によりＳｔａｎｄｂｙ機にフェールオーバーし、ＯｐｅｎＳｔａｃｋ等のクラウドコントローラのＤＢを元に仮想機器を再構築していた。ここで、従来方式では、ＨＡクラスタソフトウェアを用いたフェールオーバーは、１台のＳｔａｎｄｂｙ機に仮想機器を新たに再構築するため、全仮想機器の復旧に時間がかかるという問題がある。
【００９１】
一方、第１の実施形態に係る仮想機器管理装置１０９では、Ｎ－Ａｃｔ、０－Ｓｂｙでクラスタを組み、物理機器に障害が起きた際は、高可用ソフトウェアの機能により障害を検知するが、フェールオーバーせずに物理機器の障害を仮想機器管理装置１０９に通知する。仮想機器管理装置１０９は、各仮想機器に対して、再配置する複数の物理機器を決定し、配置する物理機器を指定してクラウドコントローラ１０８に再作成依頼を行う。そして、クラウドコントローラ１０８は、指定された物理機器に仮想機器を作成する。
【００９２】
このように、第１の実施形態に係る仮想機器管理装置１０９は、故障した物理機器上で動作していた仮想機器を、複数台の物理機器に再作成することで高速に復旧する。言い換えると、仮想機器管理装置１０９は、複数台の物理機器を復旧先として利用するため、物理機器故障時の仮想機器復旧時間が短縮される。例えば、移行先物理機器が３台の場合は、復旧処理時間が１/３に短縮できる。
【００９３】
また、第１の実施形態に係る仮想機器管理装置１０９は、Ｐａｃｅｍａｋｅｒ等の高可用ソフトウェアでＮ－Ａｃｔ、０－Ｓｂｙでクラスタを組み障害検知を行う。ここで、第１の実施形態に係る仮想機器管理装置１０９は、クラスタの枠を超えて故障復旧を行うことが出来るため、移行先物理機器の台数をクラスタサイズ以上にとることもできる。これにより、復旧時間をより短縮できる。
【００９４】
更に、仮想機器管理システムでは、障害検知のためのクラスタはＮ－Ａｃｔ、０－Ｓｂｙであるため、Ｓｔａｎｄｂｙ用の物理機器を準備する必要がなく、物理機器数の増大を抑えることができる。
【００９５】
第１の実施形態に係る仮想機器管理装置１０９は、物理機器に障害が起きた際の、仮想機器復旧を高速に行う事を狙っているが、実施形態はこれに限定されるものではない。例えば、物理機器が完全に故障した場合以外の、以下のユースケースにも、拡張して対応することが出来る。例えば、物理機器が完全には故障していなくても、複数あるファンの一つが故障した場合等は、サーバを停止してファン交換したい場合がある。かかる場合、仮想機器管理装置１０９は、物理機器上の仮想機器一括移動のＡＰＩ／ＧＵＩをオペレータ

に提供する。そして、仮想機器管理装置１０９は、ＡＰＩ／ＧＵＩ経由でのオペレータの依頼を受けて、故障復旧時と同様に仮想機器を一括で別物理機器に移動する。これにより、例えば管理者は、ファンの一つが故障した物理機器をメンテナンスすることが出来る。

【００９６】

なお、上述した実施形態では、配置先選択部１１１ｂは、仮想機器を新規に作成する通常のオペレーション時に、仮想機器を出来るだけ分散して配置するように物理機器１０３を選択するものとして説明したが実施形態はこれに限定されるものではない。例えば、配置先選択部１１１ｂは、仮想機器を新規に作成する通常のオペレーション時には、仮想機器を分散させることなく配置するように物理機器１０３を選択するようにしてもよい。

【００９７】

また、図１１に示す例では、再配置先選択部１１１ｄが、特定した複数の物理機器の物理資源の空き容量が多い順に、特定した複数の物理機器に順序付けを行う場合について説明したが、実施形態はこれに限定されるものではない。例えば、再配置先選択部１１１ｄは、物理資源の空き容量とは関係なく、特定した複数の物理機器に任意に順序付けを行うようにしてもよい。

【００９８】

（第２の実施形態）

さて、これまで本発明の実施形態について説明したが、本発明は上述した実施形態以外にも、その他の実施形態にて実施されてもよい。そこで、以下では、その他の実施形態を示す。

【００９９】

（システム構成）

また、本実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部又は一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部又は一部を公知の方法で自動的に行うこともできる。この他、上述の文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

【０１００】

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的又は物理的に分散・統合して構成することができる。

【０１０１】

（プログラム）

また、上記第１の実施形態に係る仮想機器管理装置１０９が実行する処理をコンピュータが実行可能な言語で記述した仮想機器管理プログラムを生成することもできる。この場合、コンピュータが仮想機器管理プログラムを実行することにより、上記実施形態と同様の効果を得ることができる。さらに、かかる仮想機器管理プログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録された仮想機器管理プログラムをコンピュータに読み込ませて実行することにより上記実施形態と同様の処理を実現してもよい。以下に、図１等に示した仮想機器管理装置１０９と同様の機能を実現する仮想機器管理プログラムを実行するコンピュータの一例を説明する。

【０１０２】

図１４は、仮想機器管理プログラムを実行するコンピュータ１０００を示す図である。図１４に示すように、コンピュータ１０００は、例えば、メモリ１０１０と、ＣＰＵ１０２０と、ハードディスクドライブインタフェース１０３０と、ディスクドライブインタフェース１０４０と、シリアルポートインタフェース１０５０と、ビデオアダプタ１０６０と、ネットワークインタフェース１０７０とを有する。これらの各部は、バス１０８０によって接続される。

【０１０３】

メモリ１０１０は、ＲＯＭ（Read Only Memory）１０１１およびＲＡＭ（Random Access Memory）１０１２を含む。ＲＯＭ１０１１は、例えば、ＢＩＯＳ（Basic Input Output System）等のブートプログラムを記憶する。ハードディスクドライブインタフェース１０３０は、ハードディスクドライブ１０３１に接続される。ディスクドライブインタフェース１０４０は、ディスクドライブ１０４１に接続される。ディスクドライブ１０４１には、例えば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が挿入される。シリアルポートインタフェース１０５０には、例えば、マウス１０５１およびキーボード１０５２が接続される。ビデオアダプタ１０６０には、例えば、ディスプレイ１０６１が接続される。

【０１０４】

ここで、図１４に示すように、ハードディスクドライブ１０３１は、例えば、ＯＳ１０９１、アプリケーションプログラム１０９２、プログラムモジュール１０９３およびプログラムデータ１０９４を記憶する。上記実施形態で説明した仮想機器管理プログラムは、例えばハードディスクドライブ１０３１やメモリ１０１０に記憶される。

【０１０５】

また、仮想機器管理プログラムは、例えば、コンピュータ１０００によって実行される指令が記述されたプログラムモジュールとして、例えばハードディスクドライブ１０３１に記憶される。具体的には、上記実施形態で説明した障害検出部１１１ｃと同様の情報処理を実行する検出手順と、再配置先選択部１１１ｄと同様の情報処理を実行する選択手順と、作成要求部１１１ｅと同様の情報処理を実行する依頼手順とが記述されたプログラムモジュール１０９３が、ハードディスクドライブ１０３１に記憶される。

【０１０６】

また、仮想機器管理プログラムによる情報処理に用いられるデータは、プログラムデータ１０９４として、例えば、ハードディスクドライブ１０３１に記憶される。そして、ＣＰＵ１０２０が、ハードディスクドライブ１０３１に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてＲＡＭ１０１２に読み出して、上述した各手順を実行する。

【０１０７】

なお、仮想機器管理プログラムに係るプログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０３１に記憶される場合に限られず、例えば、着脱可能な記憶媒体に記憶されて、ディスクドライブ１０４１等を介してＣＰＵ１０２０によって読み出されてもよい。あるいは、仮想機器管理プログラムに係るプログラムモジュール１０９３やプログラムデータ１０９４は、ＬＡＮ（Local Area Network）やＷＡＮ（Wide Area Network）等のネットワークを介して接続された他のコンピュータに記憶され、ネットワークインタフェース１０７０を介してＣＰＵ１０２０によって読み出されてもよい。

【０１０８】

（その他）

なお、本実施形態で説明した特定プログラムは、インターネットなどのネットワークを介して配布することができる。また、特定プログラムは、ハードディスク、フレキシブルディスク（ＦＤ）、ＣＤ－ＲＯＭ、ＭＯ、ＤＶＤなどのコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行することもできる。

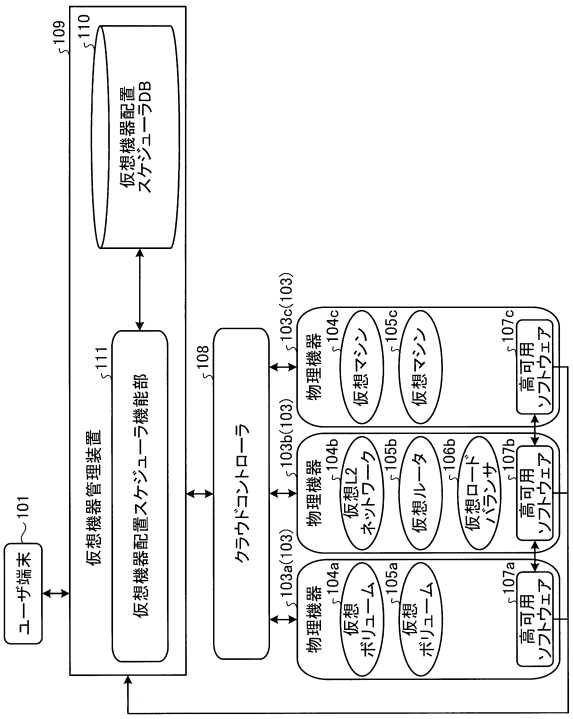
【符号の説明】

【０１０９】

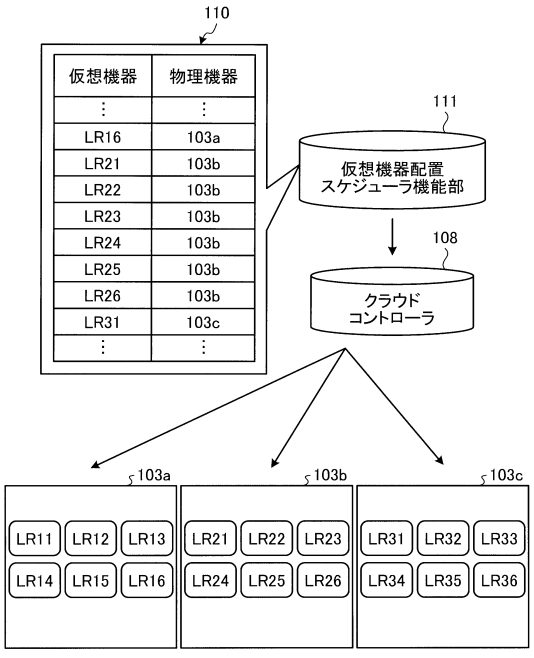
１０９　仮想機器管理装置
１１０　仮想機器配置スケジューラＤＢ
１１０ａ　仮想機器配置情報テーブル
１１０ｂ　物理資源情報テーブル
１１１　仮想機器配置スケジューラ機能部

１１１ａ　作成依頼受付部
１１１ｂ　配置先選択部
１１１ｃ　障害検出部
１１１ｄ　再配置先選択部
１１１ｅ　作成要求部
１０００　コンピュータ
１０１０　メモリ
１０１１　ＲＯＭ
１０１２　ＲＡＭ
１０２０　ＣＰＵ
１０３０　ハードディスクドライブインタフェース
１０３１　ハードディスクドライブ
１０４０　ディスクドライブインタフェース
１０４１　ディスクドライブ
１０５０　シリアルポートインタフェース
１０５１　マウス
１０５２　キーボード
１０６０　ビデオアダプタ
１０６１　ディスプレイ
１０７０　ネットワークインタフェース
１０８０　バス
１０９１　ＯＳ
１０９２　アプリケーションプログラム
１０９３　プログラムモジュール
１０９４　プログラムデータ

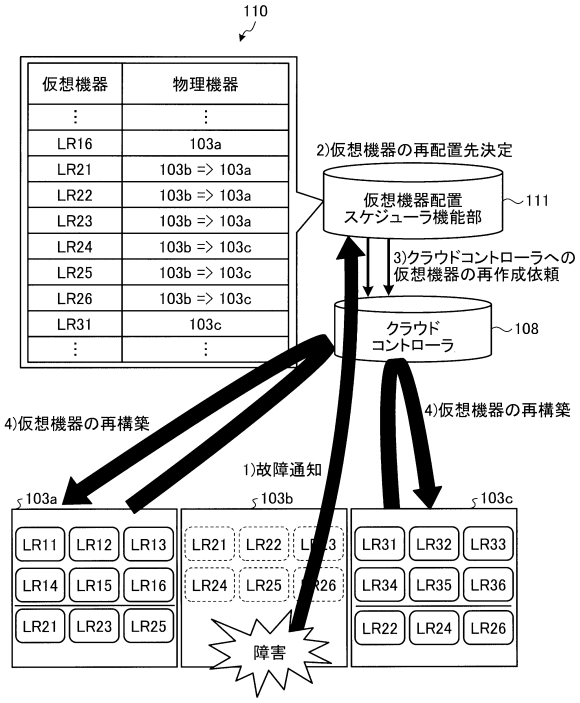
【図１】



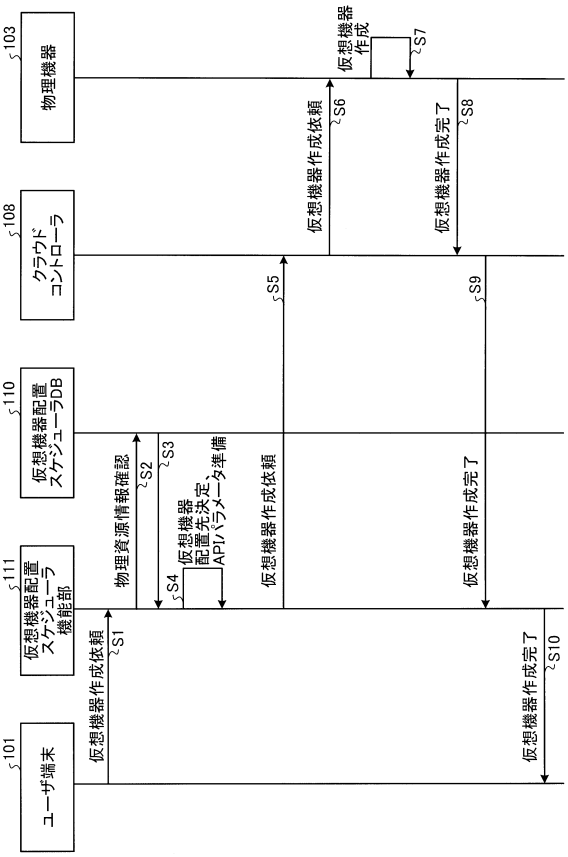
【図２】



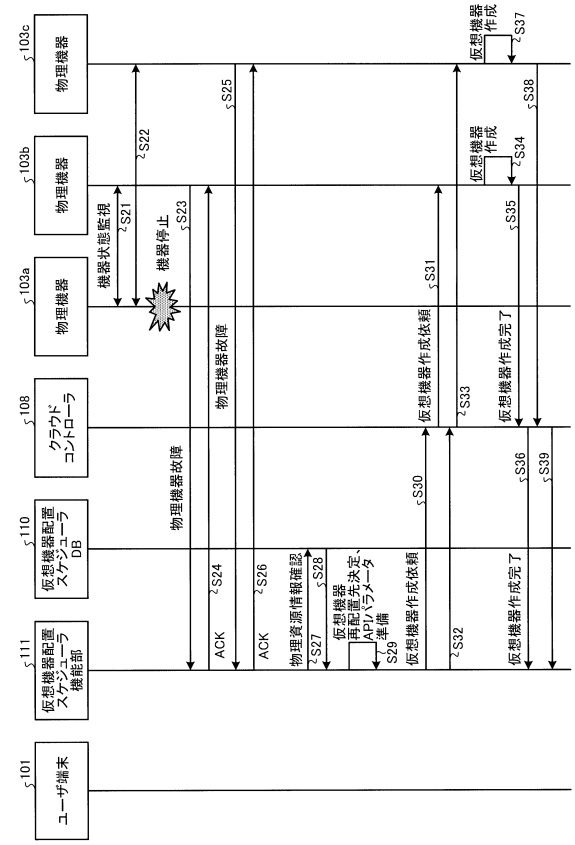
【図３】


110
仮想機器
物理機器
LR16
103a
LR21
103b => 103a
LR22
103b => 103a
LR23
103b => 103a
LR24
103b => 103c
LR25
103b => 103c
LR26
103b => 103c
LR31
103c
2)仮想機器の再配置先決定
仮想機器配置
スケジューラ機能部
111
3)クラウドコントローラへの
仮想機器の再作成依頼
クラウド
コントローラ
108
4)仮想機器の再構築
4)仮想機器の再構築
1)故障通知
103a
103b
103c
LR11
LR12
LR13
LR14
LR15
LR16
LR21
LR23
LR25
LR21
LR22
LR24
LR25
LR31
LR32
LR33
LR34
LR35
LR36
LR22
LR24
LR26
障害

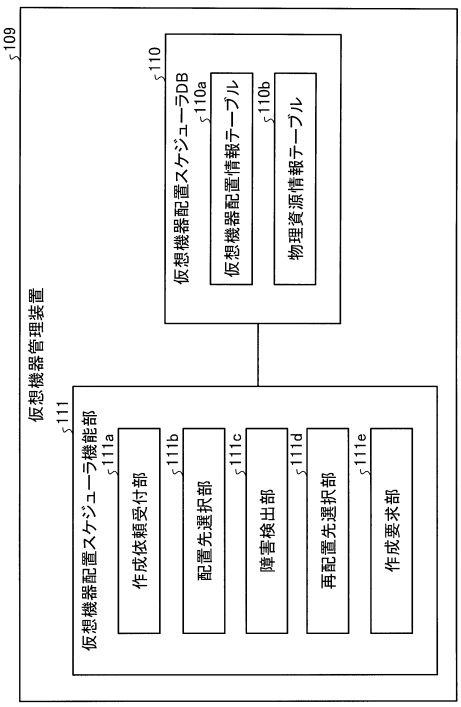
【図４】


103
物理機器
仮想機器
作成
S7
108
クラウド
コントローラ
仮想機器作成依頼
S6
仮想機器作成完了
S8
S5
S9
110
仮想機器配置
スケジューラDB
物理資源情報確認
S2
S3
S4
仮想機器
配置先決定、
APIパラメータ準備
仮想機器作成依頼
仮想機器作成完了
111
仮想機器配置
スケジューラ
機能部
S1
仮想機器作成依頼
仮想機器作成完了
S10
101
ユーザ端末

【図５】


103c
物理機器
仮想機器
作成
S37
S25
S22
103b
物理機器
仮想機器
作成
S34
S38
機器状態監視
S21
機器停止
S23
S35
103a
物理機器
S31
物理機器故障
108
クラウド
コントローラ
仮想機器作成依頼
S33
仮想機器作成完了
物理機器故障
S36
S39
110
仮想機器配置
スケジューラ
DB
S30
S24
ACK
S26
ACK
物理資源情報確認
S28
S27
仮想機器
再配置先決定、
APIパラメータ
準備
S29
仮想機器作成依頼
S32
仮想機器作成完了
111
仮想機器配置
スケジューラ
機能部
101
ユーザ端末

【図６】


109
仮想機器管理装置
110
仮想機器配置スケジューラDB
110a
仮想機器配置情報テーブル
110b
物理資源情報テーブル
111
仮想機器配置スケジューラ機能部
111a
作成依頼受付部
111b
配置先選択部
111c
障害検出部
111d
再配置先選択部
111e
作成要求部

【図７】

110a

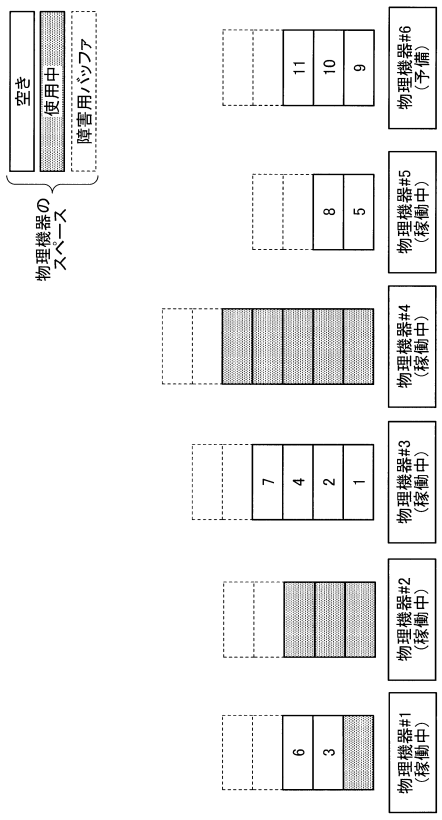
仮想機器ID	物理機器ID
仮想ボリューム#1	物理機器#1
仮想ボリューム#2	物理機器#1
仮想L2ネットワーク#1	物理機器#2
仮想ルータ#1	物理機器#2
仮想ロードバランサ#1	物理機器#2
仮想マシン#1	物理機器#3
仮想マシン#2	物理機器#3

【図８】

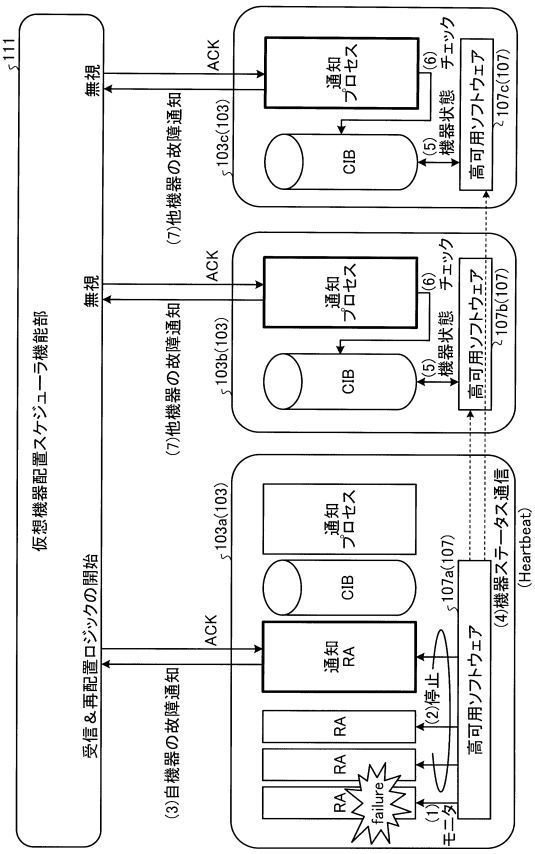
110b

物理機器ID	稼働状態	空き	使用中	障害用
物理機器#1	稼働中	3	1	2
物理機器#2	稼働中	5	3	2
物理機器#3	稼働中	4	2	2

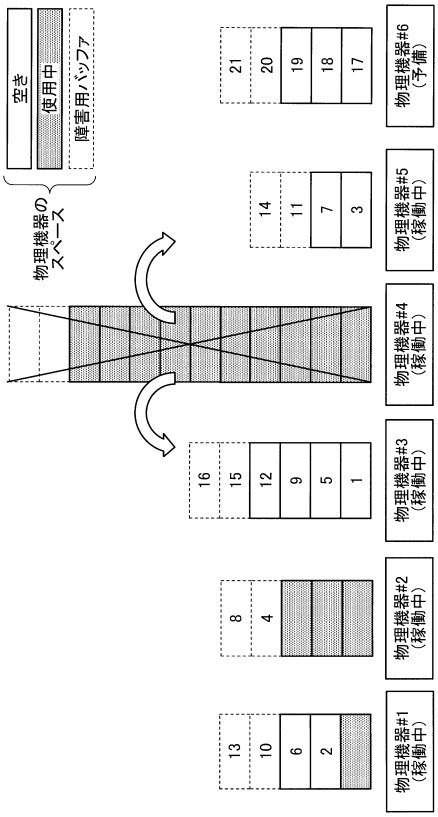
【図９】



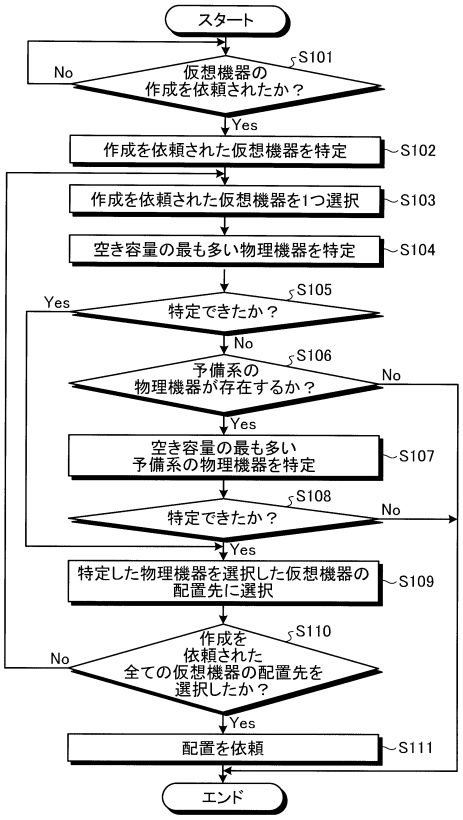
【図１０】



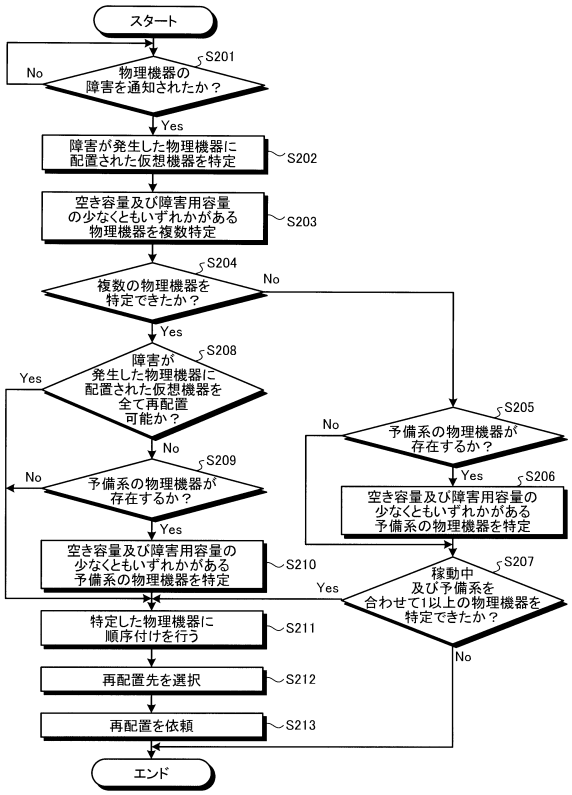
【図１１】



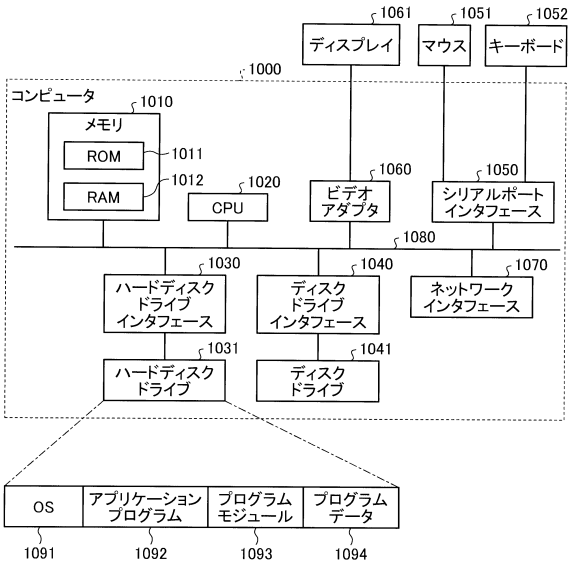
【図１２】



【図１３】



【図１４】



フロントページの続き

(72)発明者　長尾　伸二
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　渡邊　拓磨
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

合議体
審判長　仲間　晃
審判官　辻本　泰隆
審判官　佐久　聖子

(56)参考文献　特開２００９－２８２７１４（ＪＰ，Ａ）
米国特許出願公開第２０１３／０２３２５０４（ＵＳ，Ａ１）
特開２０１１－２３２９１６（ＪＰ，Ａ）
特開２０１１－２３３１４６（ＪＰ，Ａ）
特開２０１２－２０８５９８（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 9/46- 9/54
G06F11/20