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

特許第6144346号
(P6144346)

(45)発行日　平成29年6月7日(2017.6.7)　(24)登録日　平成29年5月19日(2017.5.19)

(51)Int.Cl.			F I		
G06F	9/50	(2006.01)	G06F	9/46	462Z
G06F	9/46	(2006.01)	G06F	9/46	350

請求項の数 13　(全 24 頁)

(21)出願番号	特願2015-528707 (P2015-528707)
(86)(22)出願日	平成25年8月23日 (2013.8.23)
(65)公表番号	特表2015-529918 (P2015-529918A)
(43)公表日	平成27年10月8日 (2015.10.8)
(86)国際出願番号	PCT/US2013/056524
(87)国際公開番号	WO2014/032031
(87)国際公開日	平成26年2月27日 (2014.2.27)
審査請求日	平成27年4月23日 (2015.4.23)
(31)優先権主張番号	13/593,226
(32)優先日	平成24年8月23日 (2012.8.23)
(33)優先権主張国	米国(US)

(73)特許権者　506223509
アマゾン・テクノロジーズ、インコーポレイテッド
アメリカ合衆国、９８１０８－１２２６
ワシントン州　シアトル　ピーオー　ボックス　８１２２６
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100103034
弁理士　野河　信久
(74)代理人　100075672
弁理士　峰　隆司
(74)代理人　100153051
弁理士　河野　直樹

最終頁に続く

(54)【発明の名称】仮想計算機インスタンスのスケーリング

(57)【特許請求の範囲】

【請求項１】

コンピュータ実装方法であって、

実行可能な命令を有して構成される、１つ以上の計算システムの制御下で、

配置サービスによって、ユーザから、仮想計算機がスケーラブルであるように要求するアプリケーションプログラミングインターフェース（ＡＰＩ）要求を受信することと、ここで、前記ＡＰＩ要求は、要求された前記仮想計算機を設定するために使用される仮想計算機イメージを規定し、

前記配置サービスによって、要求された前記仮想計算機がスケーラブルであること、および、サービスプロバイダ環境で動作しているホスト計算装置が、規定された前記仮想計算機イメージのためのリソース容量の増加に対応することができるように、要求された前記仮想計算機に追加の計算リソースを追加する容量を含むこと、を決定することと、

前記配置サービスによって、要求された前記仮想計算機がスケーラブルである場合、規定された前記仮想計算機イメージにおける増加した作業負荷に対処するのに十分な容量を有する、または、要求された前記仮想計算機がスケーラブルではない場合、規定された前記仮想計算機イメージの過剰な容量がほとんどまたは全くない、要求された前記仮想計算機を前記ホスト計算装置上に設定することと、

前記サービスプロバイダ環境中で動作しているスケーリングサービスから、仮想計算機に割り当てられた計算リソースを調節するという第１の要求を受信することと、

前記第１の要求を受信したことに応答して、前記仮想計算機への１つ以上の計算リソー

スの割り当てを調節することと、
を含む、コンピュータ実装方法。
【請求項２】
前記スケーリングサービスは、
前記仮想計算機を動作させることと関連した１つ以上のメトリクスを監視することと、
前記１つ以上のメトリクスが少なくとも１つの指定閾値を超えていることを検出することと、
前記仮想計算機に割り当てられた計算リソースを調節するという第２の要求を、前記ホスト計算装置に伝送することと、
をさらに実施する、請求項１のコンピュータ実装方法。
【請求項３】
前記ホスト計算装置上で実行している監視サービスは、
前記仮想計算機と関連した１つ以上のメトリクスを監視する、
前記１つ以上のメトリクスが閾値を超えていることを検出する、および
前記１つ以上の計算リソースの割り当てを調節する、
ように構成される、請求項１のコンピュータ実装方法。
【請求項４】
前記スケーリングサービスは、
顧客から、前記仮想計算機に割り当てられた前記計算リソースを調節するという第２の要求を受信することと、
前記仮想計算機に割り当てられた前記計算リソースを調節するという前記第２の要求を伝送することと、
をさらに実施する、請求項１のコンピュータ実装方法。
【請求項５】
前記計算リソースの前記割り当ての前記調節をすることは、
利用可能な計算リソースが所定の限界に達するまで、前記仮想計算機に前記１つ以上の計算リソースを割り当て続けることと、
１つ以上の追加の仮想計算機に前記仮想計算機の作業負荷を分配するように、前記１つ以上の追加の仮想計算機を設定することと、
を含む、請求項１のコンピュータ実装方法。
【請求項６】
前記仮想計算機は、作業負荷を実行することと関連した１つ以上のメトリクスを報告する、ゲストエージェントをさらに含む、請求項１のコンピュータ実装方法。
【請求項７】
前記スケーリングサービスからの前記第１の要求に応答して、前記仮想計算機に割り当てられた前記ホスト計算装置の前記１つ以上の計算リソースに少なくとも部分的に基づき、前記仮想計算機のユーザに料金を請求すること
をさらに含む、請求項１のコンピュータ実装方法。
【請求項８】
顧客が、料金に対して、前記仮想計算機に割り当てられる前記ホスト計算装置の計算リソースを取得することを可能にする、電子市場を提供することをさらに含み、前記料金は、取得された計算リソースの需要と供給に少なくとも部分的に基づく、
請求項１のコンピュータ実装方法。
【請求項９】
計算システムであって、
少なくとも１つのプロセッサと、
前記プロセッサによって実行されるとき、前記計算システムに、
配置サービスによって、ユーザから、仮想計算機がスケーラブルであるように要求するアプリケーションプログラミングインターフェース（ＡＰＩ）要求を受信させ、ここで

、前記ＡＰＩ要求は、要求された前記仮想計算機を設定するために使用される仮想計算機イメージを規定し、

前記配置サービスによって、要求された前記仮想計算機がスケーラブルであること、および、サービスプロバイダ環境で動作しているホスト計算装置が、規定された前記仮想計算機イメージのためのリソース容量の増加に対応することができるように、要求された前記仮想計算機に追加の計算リソースを追加する容量を含むこと、を決定することと、

前記配置サービスによって、サービスプロバイダ環境で動作している前記ホスト計算装置上に、要求された前記仮想計算機がスケーラブルである場合、規定された前記仮想計算機イメージにおける増加した作業負荷に対処するのに十分な容量を有する、または、要求された前記仮想計算機がスケーラブルではない場合、規定された前記仮想計算機イメージの過剰な容量がほとんどまたは全くない、要求された前記仮想計算機を設定させ、

動作している仮想計算機に割り当てられた計算リソースを調節することに関したウェブサービス要求を受信させ、かつ

前記ウェブサービス要求を受信したことに応答して、１つ以上の計算リソースを前記仮想計算機に割り当てさせる、

命令を含む、メモリと、

を備える、計算システム。

【請求項１０】

前記ウェブサービス要求は、前記仮想計算機に割り当てられた前記計算リソースが調節される条件の１セットを指定する、１つ以上のメトリクスを含む、請求項９に記載の計算システム。

【請求項１１】

前記仮想計算機に前記１つ以上の計算リソースを割り当てるための前記命令は、更に、前記計算システムに、利用可能な計算リソースが所定の限界に達するまで、１つ以上の追加の計算リソースを前記仮想計算機に割り当て続けさせ、かつ

前記メモリは、さらに、

前記プロセッサによって実行されるとき、前記計算システムに、前記仮想計算機と１つ以上の追加の仮想計算機に前記仮想計算機の作業負荷を分配するように、前記１つ以上の追加の仮想計算機を設定させる、

命令を含む、

請求項９に記載の計算システム。

【請求項１２】

前記メモリは、

前記プロセッサによって実行されるとき、前記計算システムに、

ユーザが、前記ウェブサービス要求に応答して、前記計算リソースを割り当てることによってスケーリングされることが可能である前記仮想計算機の種類を選択したことを決定させ、かつ

前記ユーザによって選択された前記仮想計算機の前記種類に少なくとも部分的に基づき、前記ユーザに料金を請求させる、

命令をさらに含む、請求項９に記載の計算システム。

【請求項１３】

前記メモリは、

前記プロセッサによって実行されるとき、前記計算システムに、

顧客が、前記計算システム上の前記１つ以上の計算リソースの需要と供給に少なくとも部分的に基づき、前記仮想計算機のために前記計算システムの１つ以上の追加の計算リソースを購入することを可能にする、電子市場を提供させる

命令をさらに含む、請求項９に記載の計算システム。

【発明の詳細な説明】

【背景技術】

【０００１】

ますます多くのアプリケーションおよびサービスが、インターネットなどのネットワーク上で利用可能となっているため、ますます多くのコンテンツ、アプリケーション、および／またはサービスの提供者が、クラウドコンピューティングなどの技術に取り掛かっている。クラウドコンピューティングは概して、ウェブサービスなどのサービスを通して電子リソースへのアクセスを提供するためのアプローチであり、この場合、それらのサービスをサポートするために使用されるハードウェアおよび／またはソフトウェアは、任意の所与の時間におけるサービスの要求を満たすために、動的にスケーラブルである。ユーザまたは顧客は典型的に、クラウドを通したリソースへのアクセスをレンタルする、リースする、または別様にその代金を支払い、したがって、必要とされるハードウェアおよび／またはソフトウェアを購入および維持する必要はない。

【０００２】

この背景では、多くのクラウド計算プロバイダは、複数のユーザが基礎的ハードウェアおよび／またはソフトウェアリソースを共有することを可能にするために、仮想化を利用する。仮想化は、計算サーバ、記憶装置、または他のリソースが、特定のユーザと関連した（例えば、特定のユーザによって所有される）複数の分離したインスタンスに分割されることを可能にすることができる。クラウド計算プロバイダは通常、その顧客のそれぞれに１つ以上の仮想計算機を割り当て、仮想計算機は、それらの顧客に対するアプリケーションおよび／または他の作業負荷を実行するために使用される。しかしながら、要求の増加または他の理由により、顧客の処理負荷が仮想計算機の容量を超え始めるとき、多くの問題および不便さが生じる可能性がある。

【図面の簡単な説明】

【０００３】

本開示に従った種々の実施形態が、図面を参照して記載される。

【図１】種々の実施形態に従って、追加のＣＰＵを割り当てることによって、仮想計算機インスタンスをスケールアップする一実施例を示す。

【図２】種々の実施形態に従って、サービスプロバイダによって展開される自動スケーリングサービスの一実施例を示す。

【図３Ａ】種々の実施形態に従って、ホスト計算機上の仮想計算機インスタンスを自動的にスケーリングするための例示的な過程を示す。

【図３Ｂ】種々の実施形態に従って、ユーザからの要求を受信したことに応答して、仮想計算機インスタンスをスケーリングする例示的な過程を示す。

【図４】種々の実施形態に従って、仮想計算機インスタンスを自動的にスケーリングし、追加の仮想計算機インスタンスを割り当てるための例示的な過程を示す。

【図５】種々の実施形態に従って利用され得る、例示的な計算装置の一般的なコンポーネントのセットの論理配列を示す。

【図６】種々の実施形態に従った態様を実装するための環境の一実施例を示す。

【発明を実施するための形態】

【０００４】

以下の記載において、種々の実施形態は、添付の図面の図の制限のためではなく、一例として示される。本開示における種々の実施形態の参照は、必ずしも同一の実施形態に対してではなく、本開示における種々の実施形態の参照は、少なくとも１つを意味する。特定の実装および他の詳細が考察されるが、これは例示目的でのみ行われることを理解されたい。当業者は、特許請求の範囲に記載された主題の範囲および精神から逸脱することなく、他のコンポーネントおよび構成が使用され得ることを認識するであろう。

【０００５】

本開示の種々の実施形態に従ったシステムおよび方法は、スケーリング計算リソースのための従来のアプローチにおいて経験される上記または他の欠陥のうちの１つ以上を克服し得る。具体的に、種々の実施形態は、自動的に、仮想計算機インスタンスに追加の計算リソース（例えば、プロセッサ、メモリ、ホームネットワーキングデバイスなど）を割り当てる、および／または種々のユーザ指定閾値またはユーザ要求に従って、仮想計算機イ

ンスタンスからの計算リソースの割り当てを解除するためのアプローチを提供する。効果的に、これは、仮想計算機インスタンスが、仮想計算機が提供するリソースに対してオンデマンドでまたは実際の要求に従って、サイズおよび容量を「拡大」または「縮小」することを可能にする。

【０００６】

種々の実施形態に従って、１つのアプローチは、その顧客の代わりにアプリケーションおよび仮想計算機インスタンスをホストする、共有された計算リソース環境（例えば、「クラウド」計算プロバイダ）のサービスプロバイダによって実装されてもよい。アプリケーションおよび仮想計算機インスタンスは、ビスプロバイダによって所有および運営される物理的なリソース（例えば、ホストサーバおよび他のネットワークリソース）上でホストされる。一実施形態に従って、サービスプロバイダは、顧客から仮想計算機イメージを受信し、仮想計算機イメージに少なくとも部分的に基づき、顧客に対する１つ以上の仮想計算機インスタンスを設定する。これらの仮想計算機インスタンスは次いで、サービスプロバイダの物理的な計算リソースを使用して、顧客の種々のアプリケーションおよび／または他のサービスを実行することができる。

【０００７】

一実施形態に従って、各仮想計算機インスタンスは、ホスト計算機（例えば、計算装置）上に設定される。各ホスト計算機は、１つ以上の仮想計算機インスタンスをホストすることができる。少なくとも１つの実施形態において、ホスト計算機はさらに、ホスト計算機のハードウェアデバイスドライバへのアクセスを提供し、かつ１つ以上の仮想計算機インスタンスが直接または仮想化抽象化を通してデバイスにアクセスすることを可能にする、ハイパーバイザおよびサービスホスティング層を含む。

【０００８】

一実施形態に従って、いったん仮想計算機インスタンスがホスト計算機上に設定されると、サービスプロバイダは、顧客から（例えば、アプリケーションプログラミングインターフェース（ＡＰＩ）を介して）、追加のリソースを仮想計算機インスタンスに割り当てるという、またはインスタンスからのリソースの割り当てを解除するという要求を受信してもよい。さらに、ＡＰＩは、顧客が、ＣＰＵ利用などの基礎的リソースの種々の操作メトリクスに関連する、仮想計算機インスタンスに対する１つ以上の顧客によって定義された閾値を指定することを可能にし得る。加えて、顧客は、仮想計算機インスタンスをスケーリングするための決定に関連し得る、顧客のサービスまたはアプリケーションと関連した、種々のランタイム操作メトリクスを指定することが可能になる。これらの操作メトリクスおよび閾値は、顧客が、仮想計算機インスタンスに割り当てられたリソースがスケールアップまたはダウンされるべきである条件を指示することを可能にする。

【０００９】

一実施形態に従って、サービスプロバイダのシステム上のサービスは、仮想計算機インスタンスの実行中に操作メトリクスを監視する。同一または代替の実施形態において、サービスは、仮想計算機インスタンス内で実行しているゲストエージェントから、操作メトリクスを受信してもよい。結果として、操作メトリクスは、サーバから、および／または仮想計算機インスタンス内から生成されてもよい。サービスが、メトリクスのうちの１つ以上が、所定の期間にわたって顧客によって定義された閾値などの閾値を超えていることを検出する場合、それは、リソース（例えば、中央処理装置（ＣＰＵ）、メモリ、他のハードウェアデバイス）を追加または除去することによって、仮想計算機インスタンスのスケーリングアップまたはダウンを開始してもよい。例えば、サービスが、この１時間にわたって少なくとも１０秒間、９０％を超えるＣＰＵ容量で動作していたことを検出する場合、それは、追加のＣＰＵ容量を仮想計算機インスタンスに割り当ててもよい（例えば、追加のＣＰＵまたはＣＰＵコアを割り当てる、より強力なＣＰＵに切り替えるなど）。別の例として、サービスが、仮想計算機インスタンスが特定の期間にわたって１０％未満のＣＰＵ容量で動作していたことを検出する場合、それは、仮想計算機インスタンスから、ある量のＣＰＵ容量の割り当てを解除（例えば、低減）してもよい。いくつかの実施形態

では、仮想計算機インスタンスのスケーリングは、顧客側の任意の手動の関与を必要とせず、自動的に実施され得る。他の実施形態では、仮想計算機インスタンスのスケーリングは、ユーザ（例えば、仮想計算機の所有者）からのインスタンスをスケーリングするという要求を受信したことに応答して、実施され得る。

【００１０】

一実施形態に従って、仮想計算機インスタンスは、単一の仮想計算機インスタンスがもはや顧客の作業負荷を適切にサポートすることができなくなるまで、自動的にスケールアップされ得る。いったんこの限界が達せられると、サービスは、作業負荷に対応するために、追加の仮想計算機インスタンスを自動的に割り当て始めてもよい。加えて、サービスは、すでに記載された方法で、変動する要求を満たすように、追加のＶＭインスタンスのそれぞれを自動的にスケールアップまたはダウンし続けてもよい。いくつかの実施形態では、操作メトリクスおよびユーザによって定義された閾値は、冗長性、アベイラビリティ、耐久性などの要件を部分的に含む場合があるため、ある特定量のＣＰＵまたはＲＡＭなど、他のリソース要件を満たすための単一サーバ上のリソース能力が十分である場合でさえ、異なる物理的なサーバ上でホストされる複数のＶＭへの「スケールアウト」が生じ得る。

【００１１】

種々の実施形態に従って、ホスト内のＶＭインスタンスのスケーリングアップ（またはダウン）の管理は、ウェブベースのグラフィカルユーザインターフェース（ＧＵＩ）、顧客によって定義された閾値、または閉じた測定／アクションループにおける完全に自動的な推論を使用して実施され得る。この自動スケーリングは、いくつかの請求および／または支払いモデルを可能にし得る。例えば、顧客は、ＲＡＭの１ＧＨｚ時間（もしくは他の所定の期間）あたりおよび／またはＧＢ／時間の料金を請求し、個々の装置リソース（ＣＰＵ、ＲＡＭ、ネットワーク）を分離し、別個に請求するなどの汎用スケーラブルＶＭインスタンスに対する料金を請求されてもよい。

【００１２】

種々の実施形態では、ウェブサービスは、ユーザ（例えば、顧客）が仮想計算機インスタンスのスケーリングを要求すること、またはこれらのＶＭインスタンスがリソース容量を拡大もしくは収縮するときに制御する種々の閾値を指定することを可能にするために使用され得る。ウェブサービスは、クエリおよびシンプルオブジェクトアクセスプロトコル（ＳＯＡＰ）ＡＰＩの両方を含むことができる。しかしながら、ウェブサービスがＳＯＡＰベースのＡＰＩコールに限定されず、インターネットなどのネットワークを使用して実施される任意の遠隔の手順／機能／方法の実行を含むことができることに留意されたい。

【００１３】

種々の実施形態では、ウェブサービスは、サイズ変更可能な計算容量（例えば、リソースセンター中の追加のサーバインスタンス）を提供するサービスプロバイダによって展開され得る。この計算容量は、顧客のソフトウェアシステムを構築およびホストするために使用され得る。サービスプロバイダは、ＡＰＩまたはウェブツールおよびユーティリティを使用して、これらのリソースへのアクセスを提供することができる。したがって、ユーザは、リソースを追加または除去し、メトリクス、冗長性、アベイラビリティに基づきスケーリングするなどのために、サービスプロバイダによって露出されるＡＰＩ機能性にアクセスすることができる。

【００１４】

図１は、種々の実施形態に従った、追加のＣＰＵ、仮想ＣＰＵ（ＶＣＰＵ）、物理ＣＰＵ（ＰＣＰＵ）、物理ＣＰＵのコア、またはその部分であって、本明細書において概して「ＣＰＵ」と称される部分を割り当てることによって、仮想計算機インスタンスをスケールアップすることの一実施例１００を示す。例示された実施形態では、ホスト計算装置１０１は、仮想計算機インスタンス１０３および１０４を管理するハイパーバイザ１０２を含む。ハイパーバイザ１０２は、１つ以上のゲストオペレーティングシステムの実行を管理し、異なるオペレーティングシステムの複数のインスタンスが基礎的ハードウェアリソ

ースを共有することを可能にする。従来、ハイパーバイザは、動作しているゲストオペレーティングシステムの機能を有して、サーバハードウェア上にインストールされ、この場合、ゲストオペレーティングシステム自体は、サーバとしての機能を果たす。種々の実施形態では、少なくとも２タイプのハイパーバイザ１０２：タイプ１（ベアメタル）ハイパーバイザおよびタイプ２（ホスト型）ハイパーバイザがあり得る。タイプ１ハイパーバイザは、ハードウェアリソース上で直接実行し、１つ以上のゲストオペレーティングシステムを管理および制御し、それはハイパーバイザの上部で実行する。タイプ２ハイパーバイザは、オペレーティングシステム内で実行され、ハードウェアリソース上の第３のレベルで、概念的に１つ以上のゲストオペレーティングをホストする。いずれのタイプのハイパーバイザも、本明細書に記載される実施形態に従って実装され得る。ハイパーバイザ１０２は、ホストドメイン（またはサービス層もしくは仮想化層など）および１つ以上のゲストドメインなど、多くのドメイン（例えば、仮想計算機）をホストすることができる。一実施形態では、ホストドメイン（例えば、Ｄｏｍ－０）は、作成される第１のドメインであり、ハードウェアデバイスおよびハイパーバイザ１０２上で動作している他のドメインの全てを管理するのを助ける。例えば、ホストドメインは、１つ以上のゲストドメイン（例えば、Ｄｏｍ－Ｕ）の作成、破壊、移動、保存、または復元を管理することができる。種々の実施形態に従って、ハイパーバイザ１０２は、ＣＰＵ、入力／出力（Ｉ／Ｏ）メモリ、およびハイパーバイザメモリなどのハードウェアリソースへのアクセスを制御する。例示された実施形態では、ハイパーバイザ１０２は、仮想計算機インスタンスにリソースを割り当てるか、またはリソースの割り当てを解除することによって、仮想計算機インスタンスのスケーリングを実施する、自動スケーリングサービス１１４を含む。あるいは、スケーリングサービスは、Ｄｏｍ－０中、またはホスト計算装置に対して外部に存在することができ、ホスト計算装置は、外部スケーリングサービスから受信されるコマンドを実行するために、シンエージェントを含んでもよい。

【００１５】

一実施形態に従って、ホスト計算装置１０１のハードウェアリソースは、物理メモリ１１６、１つ以上の中央処理装置（ＣＰＵ）（１０７、１０８、１０９、１１０）、および任意の他のハードウェアリソースまたはデバイス１１１を含む。物理メモリ１１６は、ソリッドステートドライブ（ＳＳＤ）、磁気ディスク記憶装置（ＨＤＤ）、ランダムアクセスメモリ（ＲＡＭ）などが含まれるがこれらに限定されない、任意のデータ記憶装置を含むことができる。種々の実施形態では、他のハードウェアリソース１１１としては、ネットワークインターフェースコントローラ（ＮＩＣ）、グラフィックスプロセッシングユニット（ＧＰＵ）、周辺入力／出力（Ｉ／Ｏ）デバイスなどが含まれ得るがこれらに限定されない。

【００１６】

一実施形態に従って、各仮想計算機インスタンス（１０３、１０４）は、少なくとも１つのユーザ１１２、１１３（例えば、サービスプロバイダの顧客）と関連付けられ得る。各仮想計算機インスタンスは、ユーザに代わって、少なくとも１つのアプリケーション（１０５、１０６）または他のサービスを実行することができる。例示された実施形態に従って、仮想計算機インスタンス１０３は、ＣＰＵのうちの１つ以上のセット（例えば、１０７および１０８）が割り当てられ、仮想計算機インスタンス１０４は、ＣＰＵの別のセット（例えば、１１０）が割り当てられる。種々の実施形態では、ＣＰＵは、実際の物理ＣＰＵであってもよく、あるいは仮想計算機に割り当てられる仮想ＣＰＵ容量であってもよい。

【００１７】

種々の実施形態では、ユーザ（１１２、１１３）は、ユーザの仮想計算機と関連した種々の操作メトリクスに対する１つ以上の閾値を指定することが可能になる。図面に示されるように、仮想計算機インスタンス１０３上で実行しているアプリケーション１０５上の処理負荷が、ユーザの仮想計算機と関連した種々の操作メトリクスに対する１つ以上の閾値を超えるとき、システムは、増加した要求を満たすために、追加のＣＰＵ１０９を仮想

計算機インスタンス１０３に割り当てることができる。同様に、処理負荷が減少するとき、システムは、仮想計算機インスタンス１０３に割り当てられたＣＰＵ容量を低減してもよい。

【００１８】

代替の実施形態では、仮想計算機インスタンスのスケーリングは、仮想計算機インスタンスに割り当てられるリソースの量を増加または減少させるという顧客からの要求を受信すると実施され得る。例えば、顧客は、サービスプロバイダによって提供されるＡＰＩを起動し、サービスプロバイダに追加のＣＰＵ容量を仮想計算機インスタンスに割り当てるよう要求してもよい。要求を受信したことに応答して、スケーリングサービス１１４は、追加のＣＰＵ容量を仮想計算機に割り当てることができる。

【００１９】

種々の実施形態に従って、システムはまた、メモリ１１６および／または他のハードウェアリソース（例えば、ＮＩＣ、ＧＰＵ容量など）を割り当てるか、またはその割り当てを解除することによって、仮想計算機インスタンス（１０３、１０４）をスケーリングすることができる。例えば、仮想計算機インスタンス１０３がメモリ容量の９０％に近づいている場合、システムは、追加のメモリ（例えば、物理メモリ、仮想メモリ）を仮想計算機インスタンス１０３に割り当ててもよい。

【００２０】

一実施形態に従って、仮想計算機インスタンスのスケーリングは、それを１つの仮想計算機インスタンスタイプから別のインスタンスタイプに変更することを含むことができ、この場合、各インスタンスタイプは、所定のリソースのセットと関連付けられる。例えば、所定の閾値を超えると、サービスは、顧客に割り当てられる仮想計算機を、「小型」インスタンスタイプ（例えば、１．７ＧＢ　ＲＡＭおよび１６０ＧＢの記憶装置）から「中型」インスタンスタイプ（例えば、３．７５ＧＢ　ＲＡＭおよび４１０ＧＢの記憶装置）に変更してもよい。代替の実施形態では、仮想計算機インスタンスのスケーリングは、例えば、仮想計算機中で実行しているユーザアプリケーションまたはサービスによって要求されるように、かつ定義されたメトリクスおよび閾値に従って、例えば、あらゆる任意の増加量で、あらゆる任意の量のＣＰＵ、メモリ、または他のリソース容量を追加することによって、円滑に連続して実施され得る。

【００２１】

図２は、種々の実施形態に従って、サービスプロバイダによって展開される自動スケーリングサービスの一実施例２００を示す。例示された実施形態では、サービスプロバイダ２０１は、サービスプロバイダがその顧客にリースを提供するホストサーバ（２１９、２２０）などの計算リソースのセットを所有および運営する。少なくとも１つの実施形態に従って、サービスプロバイダ２０１は、各ユーザ（例えば、顧客）が１つ以上の仮想計算機インスタンス（２０９、２１０、２１１、２１２）と関連する、共有されたリソース実行環境を作成する。仮想計算機インスタンスは、計算リソース２１４上で動作し、ネットワーク（例えば、インターネット）上で、種々のデバイス上のユーザによってアクセス可能である。本開示全体を通して使用されるように、ネットワークは、インターネットまたは他の広域ネットワーク（ＷＡＮ）、セルラーネットワーク、ローカルエリアネットワーク（ＬＡＮ）、ストレージエリアネットワーク（ＳＡＮ）、イントラネット、エクストラネットなどを含まれるがこれらに限定されない、互いと通信することが可能であるデバイスの任意の有線または無線ネットワークであってもよい。サービスプロバイダのホストサーバ（２１９、２２０）などの計算リソースは、データセンター、サーバファーム、コンテンツデリバリネットワーク（ＣＤＮ）、ポイントオブプレゼンス（ＰＯＰ）などのリソースの任意の物理または論理グループ中に位置し得る。

【００２２】

一実施形態に従って、サービスプロバイダは、ユーザ（例えば、顧客）が仮想計算機インスタンス（２０９、２１０、２１１、２１２）にアクセスし、それらを管理することを可能にするための、１つ以上のアプリケーションプログラミングインターフェース（ＡＰ

Ｉ）を露出する。例えば、ＡＰＩ２０８は、ユーザに対する１つ以上の仮想計算機インスタンスを設定するために使用される仮想計算機イメージを提示するために、ユーザによって採用され得る。同様に、本明細書に記載される種々の実施形態に従って、ＡＰＩ２０８は、閾値が関連する１つ以上のユーザによって定義された閾値（２１５、２１６、２１７、２１８）およびメトリクスを指定するために採用され得る。例えば、１つの閾値は、すでに記載されたように、仮想計算機インスタンスに割り当てられるＣＰＵの動作容量と関連付けられてもよい。別の閾値は、仮想計算機に割り当てられる利用可能なメモリの量と関連付けられてもよい。別の閾値は、特定の期間にわたって仮想計算機インスタンス上で実行しているアプリケーションによって処理されている要求の平均数であってもよい。加えて、ＡＰＩは、追加のリソース容量を仮想計算機インスタンス（複数を含む）に割り当てるか、または仮想計算機インスタンス（複数を含む）からリソース容量の割り当てを解除するという要求を提示するために、顧客によって使用され得る。

【００２３】

一実施形態に従って、いったんユーザが閾値（２１５、２１６、２１７、２１８）を指定すると、自動監視およびスケーリングサービス２１３は、メトリクスが定義された閾値を超えたときを検出するために、ランタイム実行メトリクスを監視することができる。一実施形態では、自動スケーリングサービス２１３は、仮想計算機インスタンス（２０９、２１０、２１１、２１２）のそれぞれからランタイム情報を収集し、各ＶＭインスタンスからリソースを割り当てるか、または割り当てを解除するための決定を行う、集中型サービスである。代替の実施形態では、自動スケーリングサービス２１３は、各ホスト計算機上で動作しているサービスとして実装され得、ホスト計算機上で仮想計算機インスタンスをスケーリングすることに責任があり得る。

【００２４】

いくつかの実施形態では、ホスト計算機は、スケーリングエージェント（２２１、２２２）を含む。スケーリングエージェントは、種々のメトリクスを集中型外部スケーリングサービス２１３に報告してもよく、かつ中央スケーリングサービス２１３からのコマンドを受信し、それらを実行してもよい。一実施形態に従って、仮想計算機インスタンスのいくつかは、種々のメトリクスをスケーリングエージェントに報告する、ゲストエージェント２２４を含んでもよく（例えば、メトリクスは、仮想計算機ならびにユーザ指定のメトリクス内から感知されるように、メモリプレッシャー、ＣＰＵプレッシャーなどを示す）、スケーリングエージェントは次いで、メトリクスを自動スケーリングサービス２１３に報告する。スケーリングサービス２２１は次いで、リソース容量中で仮想計算機インスタンスをスケールアップまたはダウンする決定を行ってもよい。

【００２５】

一実施形態に従って、ユーザのサービスに対する作業負荷または要求が、単一仮想計算機インスタンスがもはや作業に適切に対応するには十分ではない、ある特定の限界に達する場合、自動スケーリングサービス２１３は、ユーザに対する新しい仮想計算機インスタンスを設定し始めることができる。加えて、自動スケーリングサービス２１３は、すでに記載されたように、各インスタンスから計算リソースを追加および／または除去することによって、各個々の仮想計算機インスタンスのスケーリングアップおよびダウンを管理し続けることができる。いくつかの実施形態では、例えば冗長性要件を満たすために、作業の全てが単一インスタンスによって対応され得る場合でさえ、作業負荷を支持するために複数のＶＭインスタンスを必要とする閾値が定義されてもよい。この場合、サービス２１３は、ユーザ指定のサイジングポリシーを満たすために、２つ以上のＶＭインスタンスに割り当てられたリソースを同時に調節してもよい。

【００２６】

一実施形態に従って、仮想計算機インスタンスの自動スケーリングは、仮想計算機の利用の料金を顧客に請求するために使用され得る、多くの異なる請求モデルを可能にすることができる。一実施形態では、顧客は、自動的にスケーラブルな仮想計算機インスタンスを利用するための割増料金が請求されてもよい。例えば、一部の顧客は、その日のある特

定の時間中、またはある特定の場合においてのみ、容量の増加を必要とする場合がある。それらの顧客に対して、オンデマンドで必要とされる容量を自動的に追加し、要求が低下した後、容量を低減することができる、自動スケーリングサービスを利用することが、費用の観点から有利であり得る。他の顧客は、顧客のサービスに対する要求を前もって容易に知らない場合があり、自動スケーリングサービスを活用することは、顧客のアプリケーションが、要件がよく理解される前にアプリケーションに過剰なリソース容量を費やすことなく、要求を満たすことを確認する、アプローチを提供することができる。別の実施形態では、顧客は、所与の期間にわたって利用されたリソース毎に請求されてもよい（例えば、利用されたＣＰＵ時間あたり、１時間あたりのメモリのＧＢあたりなど）。

【００２７】

一実施形態に従って、サービスプロバイダ２０１はさらに、ホストサーバ（２１９、２２０）上に種々の仮想計算機インスタンス（２０９、２１０、２１１、２１２）を設定することに責任がある、配置サービス２２３を使用することができる。配置サービスは、仮想計算機インスタンスがスケーラブル仮想計算機であるかどうかを決定することができる。配置サービス２２３が、仮想計算機インスタンスがスケーラブルであると決定する場合、サービスは、ランタイムまたはオンデマンドで必要とされ得るリソース容量の増加に対応することができるように、過剰な容量を有するホストサーバ上に、仮想計算機を設定することができる。例えば、顧客が割増料金で自動的にスケーラブルな仮想計算機を購入する場合、配置サービスは、ＶＭの作業負荷の増加に対応するのに十分な容量を有するホスト計算機上に、ＶＭを配置してもよい。仮想計算機がスケーラブルではない場合、配置サービスは、過剰なまたは確保された容量がほとんどまたは全くないホスト計算機上に仮想計算機を設定してもよい。

【００２８】

一実施形態に従って、サービスプロバイダ２０１はさらに、顧客が顧客の仮想計算機に対してホスト計算装置の追加のリソースを購入する（例えば、割り当てる）ことを可能にする、電子市場を提供することができる。追加のリソースの料金は、ホスト計算装置上の１つ以上のリソースの需要と供給に少なくとも部分的に基づき得る。例えば、ホスト計算機上に利用可能な大量のＣＰＵ容量があり、要求が低いままであると予測される場合、そのホスト計算機上の仮想計算機に追加のＣＰＵを割り当てるための料金は、低い可能性がある。同様に、利用可能なＣＰＵ容量が少量である場合、追加のＣＰＵに対する料金はより高くなる可能性がある。このようにして需要と供給に基づき価格変動を可能にすることによって、サービスプロバイダは、リソース利用を最適化し、そのネットワークにわたって作業負荷のより効率的な分散を提供することが可能である。

【００２９】

図３Ａは、種々の実施形態に従って、ホスト計算機上の仮想計算機インスタンスを自動的にスケーリングするための例示的な過程３００を示す。この図は、特定の配列の機能動作を示し得るが、過程は必ずしも図示される特定の順序または動作に限定されない。当業者は、この図または他の図で表現される種々の動作が種々の方法で変更、再配置、同時に実施、または応用され得ることを理解するであろう。さらに、種々の実施形態の範囲から逸脱することなく、ある特定の動作または動作配列が、過程に追加または過程から削除され得ることを理解されたい。加えて、本明細書に含有される過程の例示は、コード実行の実際の配列を指定するよりもむしろ、当業者に過程の流れのアイデアを示すよう意図され、それは、異なる流れもしくは配列として実装、性能に対する最適化、または別様に種々の方法で修正されてもよい。

【００３０】

動作３０２において、仮想計算機インスタンスが顧客に対して設定される。仮想計算機インスタンスは、顧客を代表して、共有されたリソース計算環境のサービスプロバイダによって設定され得る。一実施形態に従って、顧客に対して設定される仮想計算機インスタンスは、特定のサービスを提供するアプリケーションを実行する。例えば、顧客は、いくつかの仮想計算機を使用して、データベースサーバとしての１つの仮想計算機インスタン

ス、フロントエンド（例えば、プレゼンテーション論理）サーバとして機能する別個の仮想計算機インスタンス、およびミドルウェア計算サーバとして機能する第３の仮想計算機インスタンスを使用して、サービスを展開してもよい。仮想計算機インスタンスを設定するとき、ユーザは、１つ以上の仮想計算機をスケーリングするための顧客によって定義された閾値を指定してもよい。一実施形態では、顧客は、特定の操作メトリクスに対する種々の値および閾値を指定するために、サービスプロバイダによって提供されるＡＰＩを使用することができる。例えば、ユーザは、インスタンスが１分超の間６０％のＣＰＵ容量で動作している場合、仮想計算機インスタンスのサイズを増大させるべきであると指定してもよい。別の実施形態では、ユーザは、仮想計算機インスタンスとは無関係に閾値のセットを提供し、後に、それが開始されたとき、または別様にすでに動作している後の時間に、これらの閾値のセットをインスタンスと関連付けることができてもよい。

【００３１】

動作３０３において、自動スケーリングサービスは、作業負荷の実行中、仮想計算機インスタンスの１つ以上の操作メトリクスを監視する。例えば、ホスト計算機上に存在するエージェント過程は、ＣＰＵ利用、オープン接続数、ＩＰパケット数、要求数などの種々のランタイム情報を連続的に収集してもよい。収集された情報は、顧客によって指定された命令に従って、各仮想計算機インスタンスをスケールアップまたはダウンするための決定を行うことができる、中央サービスに報告され得る。代替の実施形態では、サービスは、ホスト計算機内でホストされ得、収集されたメトリクスは、報告される必要がない。他の実施形態では、仮想計算機インスタンスは、仮想計算機のスケーリングに関連したユーザによって指定されたメトリクスを報告する、エージェントを含んでもよい。

【００３２】

動作３０４において、サービスは、１つ以上のメトリクスが顧客によって定義された閾値を超えていることを検出する。例えば、サービスは、仮想計算機インスタンスのＣＰＵ使用量が、顧客によって指定された最小タイムフレームに対する使用量閾値を超えていることを検出してもよい。

【００３３】

動作３０５において、サービスは、種々のリソースの容量を増加または減少させるために、仮想計算機インスタンスをスケーリングすることができる。一実施形態では、処理負荷が増加した場合、スケーリングサービスは、追加の計算リソースを仮想計算機インスタンスに割り当てる。例えば、スケーリングサービスは、さらなるＣＰＵ（または仮想単位のＣＰＵ容量の）を仮想計算機インスタンスに追加してもよい。別の実施形態では、スケーリングサービスは、仮想計算機インスタンスからリソースの一部分の割り当てを解除してもよく、および／またはリソースのその部分を他の仮想計算機インスタンスに移動させてもよい。いくつかの実施形態では、割り当て解除のために選択された部分またはサブセットは、顧客によって定義された閾値内にメトリクスを戻すために決定される。

【００３４】

図３Ｂは、種々の実施形態に従って、ユーザからの要求を受信したことに応答して、仮想計算機インスタンスをスケーリングする例示的な過程３１０を示す。動作３１１において、仮想計算機インスタンスは、すでに記載されたように、ユーザに対してホスト計算機上に設定される。いったん設定されると、仮想計算機インスタンスは、ユーザの代わりに作業負荷を実行することができる。動作３１２において、サービスプロバイダは、仮想計算機インスタンスに割り当てられた計算リソース容量を増加または減少させるという要求を受信する。例えば、ユーザは、作業負荷の増加により、仮想計算機インスタンスがより多くのＣＰＵ容量を必要とすることを決定してもよい。ユーザは次いで、追加のＣＰＵを仮想計算機インスタンスに割り当てるために、ＡＰＩを起動してもよい。動作３１３において、スケーリングサービスは、要求に応答して、追加の計算リソースを仮想計算機インスタンスに割り当てるか、または仮想計算機から計算リソースの割り当てを解除する。

【００３５】

図４は、種々の実施形態に従って、仮想計算機インスタンスを自動的にスケーリングし

、追加の仮想計算機インスタンスを割り当てるための例示的な過程４００を示す。
【００３６】
動作４０１において、すでに記載されたように、仮想計算機インスタンスがユーザに対して設定される。仮想計算機インスタンスは次いで、１つ以上の予め指定された操作メトリクスに関して監視される。動作４０２において、サービスは、仮想計算機インスタンスに対する１つ以上の操作メトリクスが、顧客によって定義された閾値を横切っている（超えているか、または下回っている）ことを検出してもよい。動作４０３において、スケーリングサービスは、追加の計算リソースを仮想計算機インスタンスに割り当てることによって、仮想計算機インスタンスを自動的にスケーリングする。例えば、スケーリングサービスは、追加のメモリ容量またはＣＰＵ容量を仮想計算機インスタンスに追加してもよい。
【００３７】
動作４０４において、スケーリングサービスは、仮想計算機インスタンスが、サービスの必要とされる作業負荷を適切に満たすようにスケーリングすることができないことを決定してもよい。例えば、それは、仮想計算機インスタンスがサービスプロバイダによって許可される最大サイズまで拡大していることを決定してもよい。動作４０５において、スケーリングサービスは、作業負荷に対応するために、１つ以上の追加の仮想計算機インスタンスを自動的に設定し始めてもよい。追加の仮想計算機インスタンスのそれぞれもまた、すでに記載された方法でスケーリングされ得る（動作４０５）。
【００３８】
図５は、例示的な計算装置５００の一般的なコンポーネントのセットの論理配列を示す。本実施例において、装置は、メモリデバイスまたは素子５０４中に記憶され得る命令を実行するためのプロセッサ５０２を含む。当業者に明らかとなるように、装置は、プロセッサ５０２による実行のためのプログラム命令のための第１のデータ記憶装置、画像またはデータのための別個の記憶装置、他のデバイスと情報を共有するための着脱式メモリなど、多くの種類のメモリ、データ記憶装置、または非一時的コンピュータ可読記憶媒体を含むことができる。装置は典型的に、タッチスクリーンまたは液晶ディスプレー（ＬＣＤ）などのなんらかの種類の表示素子５０６を含むが、ポータブルメディアプレーヤーなどのデバイスが、音声スピーカーを通してなど、他の手段を介して情報を伝達してもよい。考察されたように、多くの実施形態における装置は、ユーザからの従来の入力を受信することが可能な少なくとも１つの入力素子５０８を含む。この従来の入力としては例えば、プッシュボタン、タッチパッド、タッチスクリーン、ホイール、ジョイスティック、キーボード、マウス、キーパッド、またはユーザが装置にコマンドを入力することができる任意の他のそのようデバイスもしくは素子が含まれ得る。しかしながら、いくつかの実施形態では、従来の入力装置は、ボタンを全く含まなくてもよく、ユーザが装置と接触している必要なく、装置を制御することができるように、視覚および音声によるコマンドの組み合わせによってのみ制御されてもよい。いくつかの実施形態では、図５の計算装置５００は、Ｗｉ－Ｆｉ（登録商標）、Ｂｌｕｅｔｏｏｔｈ（登録商標）、ＲＦ、有線、または無線通信システムなど、種々のネットワーク上で通信するための、１つ以上のネットワークインターフェース素子５０８を含むことができる。多くの実施形態における装置は、インターネットなどのネットワークと通信することができ、他の計算装置と通信することができてもよい。
【００３９】
本開示の実施形態が以下の節を考慮して説明され得る。
１．仮想計算機をスケーリングするためのコンピュータ実装方法であって、
実行可能な命令を有して構成される、１つ以上のコンピュータシステムの制御下で、
少なくとも１つの顧客に対する仮想計算機インスタンスを設定することであって、仮想計算機インスタンスは、ホスト計算装置上に設定される、仮想計算機インスタンスを設定することと、
アプリケーションプログラミングインターフェース（ＡＰＩ）を介して顧客から、仮

想計算機インスタンスと関連した顧客によって定義された閾値を受信することと、
仮想計算機インスタンスの実行中、仮想計算機インスタンスと関連した１つ以上のメトリクスを監視することと、
要求を受信したことに応答して、仮想計算機への１つ以上の計算リソースの割り当てを調節することと、
１つ以上のメトリクスおよび顧客によって定義された閾値に少なくとも部分的に基づき、１つ以上の計算リソースの仮想計算機インスタンスへの割り当てを調節することであって、計算リソースは、処理リソース、ネットワークリソース、またはメモリリソースのうちの少なくとも１つを含む、割り当てを調節することと、
を含む、方法。

２．追加の計算リソースが仮想計算機インスタンスに割り当てることができないことを決定することと、
追加の計算リソースが仮想計算機インスタンスに割り当てることができないことを決定したことに応答して、顧客に対する第２の仮想計算機インスタンスを設定することと、
をさらに含む、付記１に記載の方法。

３．顧客と関連したアカウントに、仮想計算機インスタンスを動作するための料金を請求することであって、料金は、仮想計算機インスタンスに割り当てられた調節された計算リソースに少なくとも部分的に基づく、料金を請求することと、
をさらに含む、付記１に記載の方法。

４．コンピュータ実装方法であって、
実行可能な命令を有して構成される、１つ以上のコンピュータシステムの制御下で、
サービスプロバイダ環境中で動作しているホスト計算装置によって、仮想計算機をホスト計算装置上に設定させることと、
サービスプロバイダ環境中で動作しているスケーリングサービスから、仮想計算機に割り当てられたリソースを調節するという要求を受信することと、
要求を受信したことに応答して、仮想計算機への１つ以上の計算リソースの割り当てを調節することと、
を含む、コンピュータ実装方法。

５．スケーリングサービスは、
仮想計算機を動作させることと関連した１つ以上のメトリクスを監視することと、
１つ以上のメトリクスが少なくとも１つの指定閾値を超えていることを検出することと、
１つ以上のメトリクスが指定閾値を超えていることを検出したことに応答して、仮想計算機に割り当てられたリソースを調節するという要求を、ホストコンピュータ装置に伝送することと、
をさらに実施する、付記４に記載のコンピュータ実装方法。

６．ホスト計算装置上で実行している監視サービスは、
仮想計算機と関連した１つ以上のメトリクスを監視する、
１つ以上のメトリクスが閾値を超えていることを検出する、および
１つ以上のメトリクスが閾値を超えていることを検出したことに応答して、１つ以上の計算リソースの割り当てを調節する、
ように構成される、付記４に記載のコンピュータ実装方法。

７．スケーリングサービスは、
アプリケーションプログラミングインターフェース（ＡＰＩ）を介して顧客から、仮想計算機に割り当てられたリソースを調節するという要求を受信することと、
要求を受信したことに応答して、仮想計算機に割り当てられたリソースを調節するという要求を伝送することと、
をさらに実施する、付記４に記載のコンピュータ実装方法。

８．所定の限界が達せられるまで、仮想計算機に１つ以上の追加のリソースを割り当て続けることと、

１つ以上の追加の仮想計算機にわたって仮想計算機の作業負荷を分配するように、１つ以上の追加の仮想計算機を設定することと、
をさらに含む、付記４に記載のコンピュータ実装方法。
９．割り当てを調節することは、
１つ以上の追加の仮想計算機が、仮想計算機によって提供される少なくとも１つのサービスと関連した冗長性、アベイラビリティ、または耐久性のうちの少なくとも１つを満たすために必要とされることを決定することと、
１つ以上の追加の仮想計算機を設定することと、
をさらに含む、付記４に記載のコンピュータ実装方法。
１０．仮想計算機は、作業負荷を実行することに関連した１つ以上のメトリクスを報告する、ゲストエージェントをさらに含む、付記４に記載のコンピュータ実装方法。
１１．スケーリングサービスからの要求に応答して、仮想計算機に割り当てられたホスト計算装置の１つ以上のリソースに少なくとも部分的に基づき、ユーザに請求書を送付すること
をさらに含む、付記４に記載のコンピュータ実装方法。
１２．顧客が、料金に対して、仮想計算機に割り当てられるホスト計算装置のリソースを取得することを可能にする、電子市場を提供することをさらに含み、料金は、１つ以上のリソースの需要と供給に少なくとも部分的に基づく、
付記４に記載のコンピュータ実装方法。
１３．配置サービスによって、仮想計算機がスケーラブルであるように要求するＡＰＩ要求を受信することと、
配置サービスによって、ホスト計算装置が追加のリソースを仮想計算機に追加する容量を含むことを決定することと、
配置サービスによって、仮想計算機をホスト計算装置上に設定することと、
をさらに含む、付記４に記載のコンピュータ実装方法。
１４．計算システムであって、
少なくとも１つのプロセッサと、
プロセッサによって実行されるとき、計算システムに、
ユーザに対する仮想計算機を設定する、サービスプロバイダ環境中で動作している仮想計算機に割り当てられたリソースを調節することに関したウェブサービス要求を受信させ、かつ
要求を受信したことに応答して、仮想計算機をホストするサーバに、１つ以上の計算リソースを仮想計算機に割り当てさせる、
命令を含む、メモリと、
を備える、計算システム。
１５．ウェブサービス要求は、仮想計算機に割り当てるための計算リソースを指定する、１つ以上の入力パラメータを含有する、付記１４に記載の計算システム。
１６．ウェブサービス要求は、どのサーバが仮想計算機に割り当てられた計算リソースを調節するように命令されているかに応答して、条件の１セットを指定する、１つ以上の入力パラメータを含有する、付記１４に記載の計算システム。
１７．メモリは、実行時に、計算システムに、
仮想計算機を動作させることと関連した１つ以上のメトリクスを監視させ、
１つ以上のメトリクスが少なくとも１つの指定閾値を超えていることを検出させ、
１つ以上のメトリクスが指定閾値を超えていることを検出したことに応答して、仮想計算機に割り当てられたリソースを調節するという要求を、ホストコンピュータ装置に伝送させる、
命令をさらに含む、付記１４に記載の計算システム。
１８．スケーリングサービスは、
アプリケーションプログラミングインターフェース（ＡＰＩ）を介して顧客から、仮想計算機をスケーリングするという要求を受信することと、

要求を受信したことに応答して、仮想計算機をスケーリングするという命令を伝送することと、
をさらに実施する、付記１４に記載の計算装置。
１９．メモリは、プロセッサによって実行されるとき、計算装置に、
所定の限界が達せられるまで、１つ以上の追加のリソースを仮想計算機に割り当て続けさせ、かつ
１つ以上の追加の仮想計算機にわたって仮想計算機の作業負荷を分配するように、１つ以上の追加の仮想計算機を設定させる、
命令をさらに含む、付記１４に記載の計算装置。
２０．メモリは、プロセッサによって実行されるとき、計算装置に、
ユーザが、ウェブサービス要求に応答して、計算リソースを割り当てることによってスケーリングされることが可能である種類の仮想計算機インスタンスを選択したことを決定させ、かつ
ユーザによって選択された仮想計算機の種類に少なくとも部分的に基づき、ユーザに請求書を送付させる、
命令をさらに含む、付記１４に記載の計算装置。
２１．メモリは、プロセッサによって実行されるとき、計算装置に、
顧客が、計算装置上の１つ以上のリソースの需要と供給に少なくとも部分的に基づき、計算装置の１つ以上の追加のリソースを購入することを可能にする、電子市場を提供させる
命令をさらに含む、付記１４に記載の計算装置。
２２．ユーザに対する仮想計算機を設定することは、
配置サービスによって、仮想計算機がスケーラブルであるように要求するアプリケーションプログラミングインターフェース（ＡＰＩ）要求を受信することと、
配置サービスによって、仮想計算機をホスト計算装置上に設定することと、
をさらに含む、付記１４に記載の計算装置。
２３．ホスト計算装置上でユーザに対する仮想計算機を設定することであって、仮想計算機は、作業負荷を実行することが可能である、仮想計算機を設定することと、
仮想計算機をスケーリングするための命令を受信することであって、命令は、スケーリングサービスからホスト計算装置に受信され、スケーリングサービスは、ホスト計算装置に対して外部に存在する、命令を受信することと、
命令を受信したことに応答して、１つ以上の計算リソースの仮想計算機への割り当てを調節することであって、１つ以上の計算リソースは、ホスト計算装置のハイパーバイザによって割り当て可能である、割り当てを調節することと、
を含む、動作のセットを実施するための、１つ以上のプロセッサによって実行可能な命令の１つ以上の配列を記憶する、非一時的コンピュータ可読記憶媒体。
２４．仮想計算機は、少なくとも１つの顧客の代わりに、共有されたリソース計算環境サービスプロバイダによって設定され、スケーリングサービスは、サービスプロバイダによって展開される、付記２３に記載の非一時的コンピュータ可読記憶媒体。
２５．スケーリングサービスは、
仮想計算機によって実行される作業負荷と関連した１つ以上のメトリクスを監視することと、
１つ以上のメトリクスが少なくとも１つの指定閾値を超えていることを検出することと、
１つ以上のメトリクスが指定閾値を超えていることを検出したことに応答して、仮想計算機インスタンスをスケーリングするという命令を伝送することと、
をさらに実施する、付記２３に記載の非一時的コンピュータ可読記憶媒体。
２６．スケーリングサービスは、
アプリケーションプログラミングインターフェース（ＡＰＩ）を介して顧客から、仮想計算機をスケーリングするという要求を受信することと、

要求を受信したことに応答して、仮想計算機をスケーリングするという命令を伝送することと、
をさらに実施する、付記２３に記載の非一時的コンピュータ可読記憶媒体。
【００４０】
考察されたように、記載された実施形態に従って、種々の環境で異なるアプローチが実装され得る。例えば、図６は、種々の実施形態に従った態様を実装するための環境６００の一実施例を示す。理解されるように、ウェブベースの環境が説明目的で使用されるが、種々の実施形態を実装するために、異なる環境が必要に応じて使用されてもよい。システムは、電子クライアント装置６０２を含み、それは、適切なネットワーク６０４上で要求、メッセージ、または情報を送信および受信し、かつ装置のユーザに情報を送り返すように動作可能な任意の適切な装置を含むことができる。電子クライアント装置の例としては、パーソナルコンピュータ、携帯電話、携帯型メッセージ装置、ラップトップコンピュータ、セットトップボックス、携帯情報端末、電子ブックリーダーなどが含まれる。ネットワークとしては、イントラネット、インターネット、セルラーネットワーク、ローカルエリアネットワーク、もしくは任意のそのようなネットワーク、またはそれらの組み合わせが含まれる、任意の適切なネットワークが含まれ得る。電子クライアントシステムにために使用されるコンポーネントは、選択されたネットワークおよび／または環境の種類によって少なくとも部分的に決まってもよい。任意の適切なネットワークを介して通信するためのプロトコルおよびコンポーネントはよく知られており、本明細書では詳述されない。ネットワーク上での通信は、有線または無線接続、およびそれらの組み合わせを介して可能となり得る。本実施例では、ネットワークは、環境が要求を受信し、それに応答してコンテンツを提供するためのウェブサーバ６０６を含むようなインターネットを含むが、当業者に明らかとなるように、他のネットワークに対して、同様の目的を果たす代替の装置が使用され得る。
【００４１】
例示的な環境は、少なくとも１つのアプリケーションサーバ６０８およびデータストア６１０を含む。いくつかのアプリケーションサーバ、層、または他の素子、過程、もしくはコンポーネントがあってもよく、それらは連鎖または別様に構成されてもよく、適切なデータストアからデータを取得するなどのタスクを実施するために相互に作用することができることが理解されるべきである。本明細書で使用される場合、用語「データストア」は、データを記憶、アクセス、および検索することが可能な任意の装置または装置の組み合わせを指し、それは、任意の標準、分散、またはクラスター環境において、データサーバ、データベース、データ記憶装置、およびデータ記憶媒体の任意の組み合わせおよび数を含んでもよい。アプリケーションサーバは、クライアント装置に対する１つ以上のアプリケーションの態様を実行するために必要に応じてデータストアと統合し、アプリケーションに対するデータアクセスおよびビジネスロジックの大部分に対応するための、任意の適切なハードウェアおよびソフトウェアを含むことができる。アプリケーションサーバは、データストアと連携したアクセス制御サービスを提供し、ユーザに転送されるテキスト、グラフィックス、オーディオ、および／またはビデオなどのコンテンツを生成することができ、それらは本実施例において、ＨＴＭＬ、ＸＭＬ、または別の適切な構造化言語の形態で、ウェブサーバによってサーバに提供されてもよい。全ての要求および応答の処理、ならびにクライアント装置６０２とアプリケーションサーバ７０８との間のコンテンツの配信は、ウェブサーバ６０６によって処理され得る。本明細書で考察される構造化コードが、本明細書の他の部分で考察される任意の適切なデバイスまたはホスト計算機上で実行され得るため、ウェブおよびアプリケーションサーバは必要とされず、ただの例示的なコンポーネントにすぎないことを理解されたい。
【００４２】
データストア６１０は、特定の態様に関連するデータを記憶するためのいくつかの別々のデータ表、データベース、または他のデータ記憶装置機構および媒体を含むことができる。例えば、例示されるデータストアは、生成データ６１２およびユーザ情報６１６を記

憶するための機構を含み、それらは、生成側に対するコンテンツを提供するために使用され得る。データストアはまた、ログまたはセッションデータ６１４を記憶するための機構を含むことが示される。ページイメージ情報およびアクセス権情報など、データストア中に記憶される必要があり得る多くの他の態様が存在し得、それらは、必要に応じて上記に列挙された機構のいずれかに、またはデータストア６１０中の追加の機構に記憶され得ることを理解されたい。データストア６１０は、それと関連したロジックを通して、アプリケーションサーバ６０８から命令を受信し、それに応答してデータを取得、更新、または別様に処理するように動作可能である。一実施例では、ユーザは、ある特定の種類の項目に対する検索要求を提示してもよい。この場合、データストアは、ユーザの身元を検証するためにユーザ情報にアクセスしてもよく、その種類の項目に関する情報を取得するために、カタログ詳細情報にアクセスすることができる。情報は次いで、ユーザがユーザデバイス６０２上のブラウザーを介して見ることができるウェブページ上の結果リストなどで、ユーザに戻され得る。対象となる特定の項目に関する情報は、専用ページまたはブラウザーのウィンドウで見ることができる。

【００４３】

各サーバは典型的に、そのサーバの一般的な管理および動作に対する実行可能なプログラム命令を提供する、オペレーティングシステムを含み、典型的に、サーバのプロセッサによって実行されるとき、サーバがその目的とする機能を実施することを可能にする命令を記憶する、コンピュータ可読媒体を含む。オペレーティングシステムおよびサーバの一般的な機能性に対する好適な実装は既知であるか、または市販されており、特に本開示を考慮すると、当業者によって容易に実装される。

【００４４】

一実施形態における環境は、１つ以上のコンピュータネットワークまたは直接接続を使用して、通信リンクを介して相互接続される、いくつかのコンピュータシステムおよびコンポーネントを利用する、分散計算環境である。しかしながら、コンピュータシステムは、図６に示されるよりも少ないまたは多いコンポーネントを有するシステムにおいて、同等に動作し得ることが、当業者によって理解されるであろう。したがって、図６のシステム６００の描写は、本開示の範囲に限定するものではなく、本質的に例示的なものとして見なされるべきである。

【００４５】

本明細書で考察または示唆される種々の実施形態は、幅広い種類の動作環境で実装され得、それらは場合によっては、多くのアプリケーションのいずれかを動作させるために使用され得る１つ以上のユーザコンピュータ、計算装置、または処理装置を含むことができる。ユーザまたはクライアント装置は、標準的な動作システムを起動するデスクトップまたはラップトップコンピュータなどの多くの汎用パーソナルコンピュータ、ならびに携帯電話ソフトウェアを起動し、多くのネットワークおよびメッセージプロトコルをサポートすることが可能な、携帯電話、無線、および手持ち式装置のいずれかを含むことができる。コンピュータシステムはまた、開発およびデータベース管理などの目的で、種々の市販のオペレーティングシステムおよび他の既知のアプリケーションのいずれかを起動する多くのワークステーションを含むことができる。これらのデバイスはまた、ダミー端子、シンクライアント、ゲームシステム、およびネットワークを介して通信することが可能な他のデバイスなど、他の電子デバイスを含むことができる。

【００４６】

ほとんどの実施形態は、ＴＣＰ／ＩＰ、ＯＳＩ、ＦＴＰ、ＵＰｎＰ、ＮＦＳ、ＣＩＦＳ、およびＡｐｐｌｅＴａｌｋなど、種々の市販のプロトコルのいずれかを使用して、通信をサポートするために、当業者によく知られている少なくとも１つのネットワークを利用する。ネットワークは、例えば、ローカルエリアネットワーク、広域ネットワーク、仮想プライベートネットワーク、インターネット、イントラネット、エクストラネット、公衆交換電話網、赤外線ネットワーク、無線ネットワーク、およびそれらの任意の組み合わせであってもよい。

【００４７】
ウェブサーバを利用する実施形態では、ウェブサーバは、ＨＴＴＰサーバ、ＦＴＰサーバ、ＣＧＩサーバ、データサーバ、Ｊａｖａ（登録商標）サーバ、およびビジネスアプリケーションサーバを含む、種々のサーバまたは中間階層アプリケーションのいずれかを起動することができる。サーバ（複数を含む）はまた、Ｊａｖａ（登録商標）、Ｃ、Ｃ＃、もしくはＣ＋＋などの任意のプログラミング言語、またはＰｅｒｌ、Ｐｙｔｈｏｎ、もしくはＴＣＬなどの任意のスクリプト言語、ならびにそれらの組み合わせで書かれる１つ以上のスクリプトまたはプログラムとして実装されてもよい、１つ以上のウェブアプリケーションを実行することなどによって、ユーザデバイスからの要求に応答して、プログラムまたはスクリプトを実行することが可能であってもよい。サーバ（複数を含む）はまた、Ｏｒａｃｌｅ（登録商標）、Ｍｉｃｒｏｓｏｆｔ（登録商標）、Ｓｙｂａｓｅ（登録商標）、およびＩＢＭ（登録商標）から市販されているものが含まれるがこれらに限定されない、データベースサーバを含んでもよい。

【００４８】
環境は、上記で考察されるような種々のデータストならびに他のメモリおよび記憶装置を含むことができる。これらは、ネットワーク上のコンピュータのうちの１つ以上にローカルな（および／もしくは常駐する）、またはコンピュータのいずれかまたは全部から遠隔の記憶媒体など、種々の位置に存在することができる。実施形態の特定のセットにおいて、情報は、当業者にはよく知られているストレージエリアネットワーク（「ＳＡＮ」）に存在してもよい。同様に、コンピュータ、サーバ、または他のネットワーク装置による機能を実施するための任意の必要なファイルが、必要に応じて、ローカルおよび／またはリモートで記憶されてもよい。システムがコンピュータ化された装置を含む場合、コンピュータ化された各装置は、バスを介して電気的に連結されてもよいハードウェア要素を含むことができ、要素は、例えば、少なくとも１つの中央処理装置（ＣＰＵ）、少なくとも１つの入力装置（例えば、マウス、キーボード、コントローラー、タッチスクリーン、またはキーパッド）、および少なくとも１つの出力装置（例えば、表示装置、プリンター、またはスピーカー）を含む。コンピュータ化された装置を含むシステムはまた、ディスクドライブ、光学式記憶装置、およびランダムアクセスメモリ（「ＲＡＭ」）またはリードオンリーメモリ（「ＲＯＭ」）などのソリッドステート記憶装置などの１つ以上の記憶装置、ならびにリムーバブルメディアデバイス、メモリカード、フラッシュカードなどを含んでもよい。

【００４９】
コンピュータ化された装置を含むシステムはまた、上記に記載されるようなコンピュータ可読記憶媒体リーダー、通信装置（例えば、モデム、ネットワークカード（無線または有線）、赤外線通信装置など）、および作業メモリを含むことができる。コンピュータ可読記憶媒体リーダーは、遠隔、ローカル、固定、および／またはリムーバブル記憶装置、ならびにコンピュータ可読情報を一時的および／またはより永久的に含有、記憶、伝送、および検索するための記憶媒体を表す、コンピュータ可読記憶媒体と接続され得るか、またはコンピュータ可読記憶媒体を受信するように構成され得る。システムおよび種々の装置はまた、典型的に、クライアントアプリケーションまたはウェブブラウザーなど、オペレーティングシステムおよびアプリケーションプログラムを含む、少なくとも１つの作業メモリ装置内に位置する多くのソフトウェアアプリケーション、モジュール、サービス、または他の要素を含む。代替の実施形態が上記に記載される実施形態からの多くの変更を有してもよいことを理解されたい。例えば、カスタマイズされたハードウェアもまた使用されてもよく、および／または特定の要素がハードウェア、ソフトウェア（アプレットなどの高移植性ソフトウェアを含む）、または両方で実装されてもよい。さらに、ネットワーク入力／出力装置などの他の計算装置への接続が採用されてもよい。

【００５０】
コードまたはコードの部分を含有するための記憶媒体およびコンピュータ可読媒体は、所望の情報を記憶するために使用され得、かつシステム装置によってアクセスされ得る、

ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ（登録商標）、フラッシュメモリもしくは他のメモリ技術、ＣＤ－ＲＯＭ、デジタル多用途ディスク（ＤＶＤ）、もしくは他の光学式記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置、もしくは他の磁気記憶装置、または任意の他の媒体を含む、コンピュータ可読命令、データ構造、プログラムモジュール、または他のデータなどの情報の記憶および／または伝送のための任意の方法または技術で実装される、揮発性および不揮発性、リムーバブルおよび非リムーバブル媒体などであるがこれらに限定されない、記憶媒体および通信媒体を含む、当該技術分野で既知であるか、または使用される任意の適切な媒体を含むことができる。本明細書に提供される本開示および教示に基づき、当業者は、種々の実施形態を実装するための他の手段および／または方法を理解するであろう。

【００５１】

結果的に、本明細書および図面は、制限的な意味よりもむしろ例示的な意味でみなされるものとする。しかしながら、特許請求の範囲に記載されるような本発明のより広い精神および範囲から逸脱することなく、種々の修正および変更がそれに行われてもよいことが明らかとなるであろう。

以下に、本願出願当初の特許請求の範囲に記載された発明を付記する。

［Ｃ１］

コンピュータ実装方法であって、

実行可能な命令を有して構成される、１つ以上のコンピュータシステムの制御下で、

サービスプロバイダ環境中で動作しているホスト計算装置によって、仮想計算機を前記ホスト計算装置上に設定させることと、

前記サービスプロバイダ環境中で動作しているスケーリングサービスから、前記仮想計算機に割り当てられたリソースを調節するという要求を受信することと、

前記要求を受信したことに応答して、前記仮想計算機への１つ以上の計算リソースの割り当てを調節することと、

を含む、コンピュータ実装方法。

［Ｃ２］

前記スケーリングサービスは、

前記仮想計算機を動作させることと関連した１つ以上のメトリクスを監視することと、

前記１つ以上のメトリクスが少なくとも１つの指定閾値を超えていることを検出することと、

前記１つ以上のメトリクスが前記指定閾値を超えていることを検出したことに応答して、前記仮想計算機に割り当てられたリソースを調節するという前記要求を、前記ホストコンピュータ装置に伝送することと、

をさらに実施する、［Ｃ１のコンピュータ実装方法。

［Ｃ３］

前記ホスト計算装置上で実行している監視サービスは、

前記仮想計算機と関連した１つ以上のメトリクスを監視する、

前記１つ以上のメトリクスが閾値を超えていることを検出する、および

前記１つ以上のメトリクスが前記閾値を超えていることを検出したことに応答して、前記１つ以上の計算リソースの前記割り当てを調節する、

ように構成される、［Ｃ１］のコンピュータ実装方法。

［Ｃ４］

前記スケーリングサービスは、

アプリケーションプログラミングインターフェース（ＡＰＩ）を介して顧客から、前記仮想計算機に割り当てられたリソースを調節するという要求を受信することと、

前記要求を受信したことに応答して、前記仮想計算機に割り当てられたリソースを調節するという前記要求を伝送することと、

をさらに実施する、［Ｃ１］のコンピュータ実装方法。

［Ｃ５］
所定の限界が達せられるまで、前記仮想計算機に前記１つ以上の追加のリソースを割り当て続けることと、
１つ以上の追加の仮想計算機にわたって前記仮想計算機の作業負荷を分配するように、１つ以上の追加の仮想計算機を設定することと、
をさらに含む、［Ｃ１］のコンピュータ実装方法。
［Ｃ６］
前記仮想計算機は、前記作業負荷を実行することと関連した１つ以上のメトリクスを報告する、ゲストエージェントをさらに含む、［Ｃ１］のコンピュータ実装方法。
［Ｃ７］
前記スケーリングサービスからの要求に応答して、前記仮想計算機に割り当てられた前記ホスト計算装置の前記１つ以上のリソースに少なくとも部分的に基づき、前記ユーザに請求書を送付すること
をさらに含む、［Ｃ１］のコンピュータ実装方法。
［Ｃ８］
顧客が、料金に対して、前記仮想計算機に割り当てられる前記ホスト計算装置のリソースを取得することを可能にする、電子市場を提供することをさらに含み、前記料金は、前記１つ以上のリソースの需要と供給に少なくとも部分的に基づく、
［Ｃ１］のコンピュータ実装方法。
［Ｃ９］
配置サービスによって、前記仮想計算機がスケーラブルであるように要求するＡＰＩ要求を受信することと、
前記配置サービスによって、前記ホスト計算装置が追加のリソースを前記仮想計算機に追加する容量を含むことを決定することと、
前記配置サービスによって、前記仮想計算機を前記ホスト計算装置上に設定することと、
をさらに含む、［Ｃ１］のコンピュータ実装方法。
［Ｃ１０］
計算システムであって、
少なくとも１つのプロセッサと、
前記プロセッサによって実行されるとき、前記計算システムに、
ユーザに対する仮想計算機を設定する、サービスプロバイダ環境中で動作している仮想計算機に割り当てられたリソースを調節することに関したウェブサービス要求を受信させ、かつ
前記要求を受信したことに応答して、前記仮想計算機をホストするサーバに、１つ以上の計算リソースを前記仮想計算機に割り当てさせる、
命令を含む、メモリと、
を備える、計算システム。
［Ｃ１１］
前記ウェブサービス要求は、どの前記サーバが前記仮想計算機に割り当てられた前記計算リソースを調節するように命令されているかに応答して、条件の１セットを指定する、１つ以上の入力パラメータを含有する、［Ｃ１０］に記載の計算システム。
［Ｃ１２］
前記メモリは、前記プロセッサによって実行されるとき、前記計算装置に、
所定の限界が達せられるまで、前記１つ以上の追加のリソースを前記仮想計算機に割り当て続けさせ、かつ
１つ以上の追加の仮想計算機にわたって前記仮想計算機の作業負荷を分配するように、１つ以上の追加の仮想計算機を設定させる、
命令をさらに含む、［Ｃ１０］に記載の計算装置。
［Ｃ１３］

前記メモリは、前記プロセッサによって実行されるとき、前記計算装置に、
前記ユーザが、前記ウェブサービス要求に応答して、前記計算リソースを割り当てることによってスケーリングされることが可能である種類の前記仮想計算機インスタンスを選択したことを決定させ、かつ
前記ユーザによって選択された前記仮想計算機の前記種類に少なくとも部分的に基づき、前記ユーザに請求書を送付させる、
命令をさらに含む、［Ｃ１０］に記載の計算装置。
［Ｃ１４］
前記メモリは、前記プロセッサによって実行されるとき、前記計算装置に、
顧客が、前記計算装置上の前記１つ以上のリソースの需要と供給に少なくとも部分的に基づき、前記計算装置の１つ以上の追加のリソースを購入することを可能にする、電子市場を提供させる
命令をさらに含む、［Ｃ１０］に記載の計算装置。
［Ｃ１５］
前記ユーザに対する前記仮想計算機を設定することは、
配置サービスによって、前記仮想計算機がスケーラブルであるように要求するアプリケーションプログラミングインターフェース（ＡＰＩ）要求を受信することと、
前記配置サービスによって、前記仮想計算機を前記ホスト計算装置上に設定することと、
をさらに含む、［Ｃ１０］に記載の計算装置。

【図１】

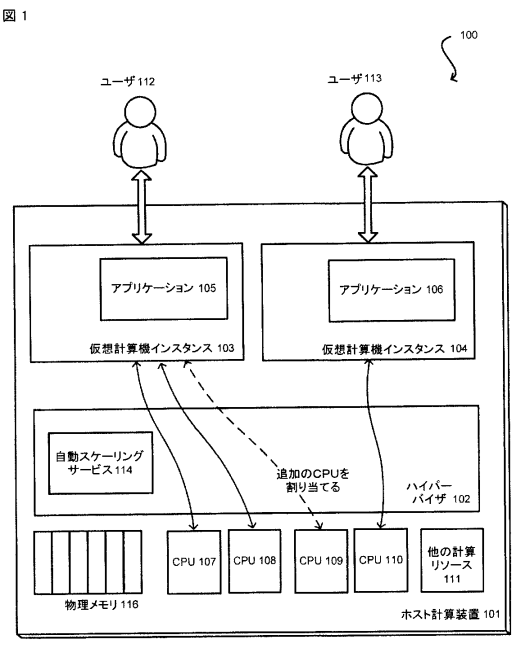

FIGURE 1

【図２】

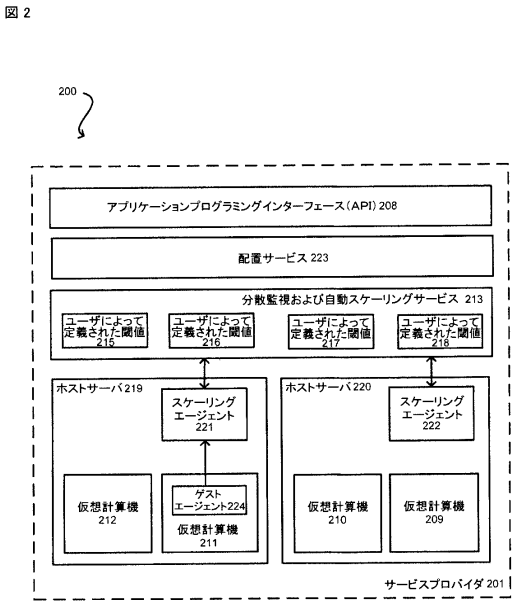

FIGURE 2

【図３Ａ】

図 3A

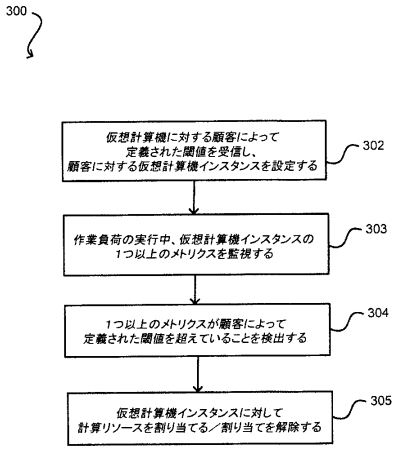

FIGURE 3A

【図３Ｂ】

図 3B

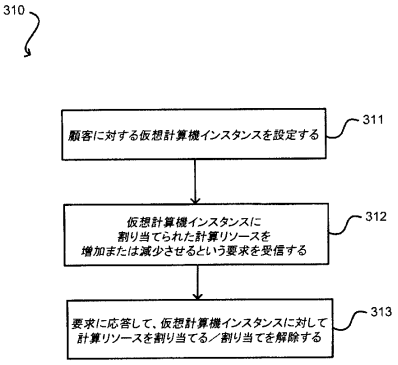

FIGURE 3B

【図４】

図 4

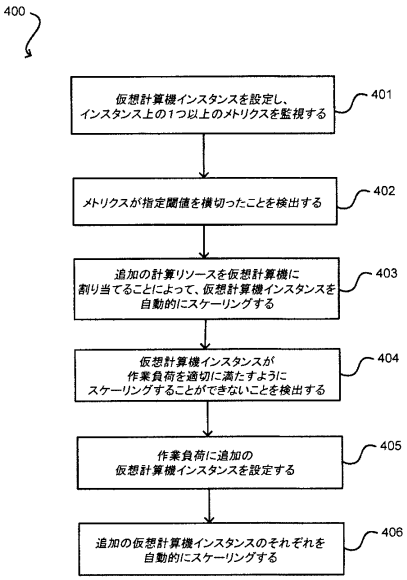

FIGURE 4

【図５】

図 5

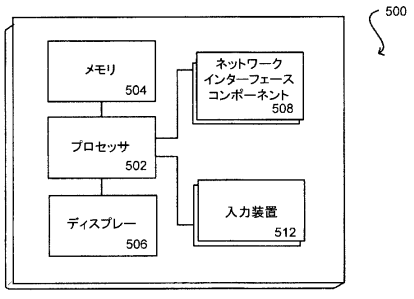

FIGURE 5

【図６】

図 6

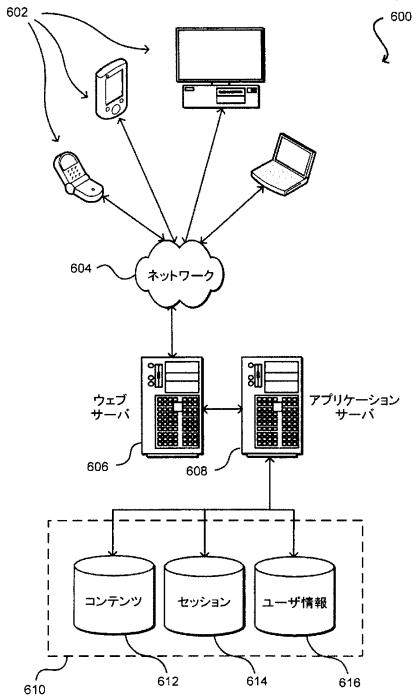

FIGURE 6

フロントページの続き

(74)代理人 100140176
弁理士 砂川 克
(74)代理人 100124394
弁理士 佐藤 立志
(74)代理人 100112807
弁理士 岡田 貴志
(74)代理人 100111073
弁理士 堀内 美保子
(72)発明者 マー、マイケル・デービッド
アメリカ合衆国、ワシントン州 ４８１０９、シアトル、テリー・アベ・ノース ４１０
(72)発明者 コワルスキー、マルチン・ピー.
アメリカ合衆国、ワシントン州 ４８１０９、シアトル、テリー・アベ・ノース ４１０

審査官 田中 幸雄

(56)参考文献 特開２０１１－２０３９１０（ＪＰ，Ａ）
特開２０１２－９９０６２（ＪＰ，Ａ）
特開２０１１－１７０６７９（ＪＰ，Ａ）
特開２００７－２００３４７（ＪＰ，Ａ）
寺田亜紀，ネット世界の「雲」をつかめ！ クラウドの成分分析と導入メリット，Ｇ－ＣＬＯＵＤ Ｍａｇａｚｉｎｅ，日本，（株）技術評論社，２０１１年 3月１５日，第2号，１２８－１３１ページ
荒井康宏ほか，徹底解説！ Ａｍａｚｏｎ Ｗｅｂ Ｓｅｒｖｉｃｅｓ WebエンジニアのためのAWSガイド，Ｇ－ＣＬＯＵＤ Ｍａｇａｚｉｎｅ，日本，（株）技術評論社，２０１０年 9月１０日，１０－１９ページ
波戸邦夫ほか，インタークラウドに向けたシステムアーキテクチャの提案，電子情報通信学会技術研究報告，日本，社団法人電子情報通信学会，２０１１年 2月２４日，Ｖｏｌ．１１０ Ｎｏ．４４９，１５１－１５６ページ

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ ９／５０
Ｇ０６Ｆ ９／４６