

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5756577号

(P5756577)

(45) 発行日 平成27年7月29日 (2015. 7. 29)

(24) 登録日 平成27年6月5日 (2015. 6. 5)

(51) Int.Cl.

F I

G 0 6 F 9/50 (2006.01)

G 0 6 F 9/46 4 6 2 Z

請求項の数 10 (全 16 頁)

(21) 出願番号	特願2014-557954 (P2014-557954)	(73) 特許権者	390009531
(86) (22) 出願日	平成25年3月21日 (2013. 3. 21)		インターナショナル・ビジネス・マシーンズ・コーポレーション
(65) 公表番号	特表2015-508207 (P2015-508207A)		I N T E R N A T I O N A L B U S I N E S S M A C H I N E S C O R P O R A T I O N
(43) 公表日	平成27年3月16日 (2015. 3. 16)		アメリカ合衆国 1 0 5 0 4 ニューヨーク州 アーモンク ニュー オーチャードロード
(86) 国際出願番号	PCT/CA2013/050229		
(87) 国際公開番号	W02013/142991		
(87) 国際公開日	平成25年10月3日 (2013. 10. 3)	(74) 代理人	100108501
審査請求日	平成26年8月26日 (2014. 8. 26)		弁理士 上野 剛史
(31) 優先権主張番号	13/431, 689	(74) 代理人	100112690
(32) 優先日	平成24年3月27日 (2012. 3. 27)		弁理士 太佐 種一
(33) 優先権主張国	米国 (US)		
早期審査対象出願			
			最終頁に続く

(54) 【発明の名称】 サーバ・クラウドにおけるキャパシティ・オン・デマンドを管理するための装置、コンピュータ実装方法、および製造品 (サーバ・クラウドにおけるキャパシティ・オン・デマンドの管理)

(57) 【特許請求の範囲】

【請求項 1】

少なくとも1つのプロセッサと、

前記少なくとも1つのプロセッサに結合されたメモリと、

前記メモリ中にあり前記少なくとも1つのプロセッサによって実行されるクラウド・キャパシティ・オン・デマンド・マネージャとを備える装置であって、

前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記装置を含むサーバ・クラウド中の複数のサーバのために複数のリソースのキャパシティを管理し、前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記サーバ・クラウド中の前記複数のサーバのうちの少なくとも1つのサーバから、前記複数のリソースのうちの少なくとも1つのリソースのキャパシティを借り、前記サーバ・クラウドが完全な状態である限り、前記借りたキャパシティを前記サーバ・クラウド中の前記複数のサーバのうちの異なるサーバに貸し、前記サーバ・クラウドがもはや完全な状態でないとき、前記クラウド・キャパシティ・オン・デマンド・マネージャが、借りたキャパシティをディセーブルにし、貸したキャパシティを回収し、

前記複数のサーバのうちの1つのサーバがもはや前記サーバ・クラウド中にないせいで前記サーバ・クラウドがもはや完全な状態でないとき、前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記複数のサーバ全てからの全ての借りたキャパシティをディセーブルにし、前記複数のサーバ全てからの全ての貸したキャパシティを回収する、装置。

【請求項 2】

前記複数のサーバのうちの少なくとも 1 つのサーバが、前記サーバ・クラウド中のいずれか他のサーバと共有できるクラウド永続キャパシティを備え、前記クラウド永続キャパシティが、前記サーバ・クラウド中の前記複数のサーバのうちの前記 1 つのサーバ上で永続的にイネーブルにされるキャパシティであって前記サーバ・クラウド中の他のサーバと共有できるキャパシティである、請求項 1 に記載の装置。

【請求項 3】

前記複数のリソースがプロセッサを含む、請求項 1 に記載の装置。

【請求項 4】

前記複数のリソースがメモリを含む、請求項 1 に記載の装置。

10

【請求項 5】

前記複数のサーバのうちの 1 つのサーバがもはや前記サーバ・クラウド中にないせいで前記サーバ・クラウドがもはや完全な状態でないとき、前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記 1 つのサーバから借りた、借りたキャパシティをディセーブルにし、前記 1 つのサーバ上の貸したキャパシティを回収し、まだ前記サーバ・クラウド中にある複数のサーバについての他の借りたキャパシティおよび貸したキャパシティを保持する、請求項 1 に記載の装置。

【請求項 6】

コンピュータに記憶されて実行されるソフトウェアを含むプログラムであって、前記ソフトウェアが、

20

サーバ・クラウド中の複数のサーバのために複数のリソースのキャパシティを管理するクラウド・キャパシティ・オン・デマンド・マネージャを備え、前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記サーバ・クラウド中の前記複数のサーバのうちの少なくとも 1 つのサーバから、前記複数のリソースのうちの少なくとも 1 つのリソースのキャパシティを借り、前記サーバ・クラウドが完全な状態である限り、前記借りたキャパシティを前記サーバ・クラウド中の前記複数のサーバのうちの異なるサーバに貸し、前記サーバ・クラウドがもはや完全な状態でないとき、前記クラウド・キャパシティ・オン・デマンド・マネージャが、借りたキャパシティをディセーブルにし、貸したキャパシティを回収し、前記複数のサーバのうちの 1 つのサーバがもはや前記サーバ・クラウド中にないせいで前記サーバ・クラウドがもはや完全な状態でないとき、前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記複数のサーバ全てからの全ての借りたキャパシティをディセーブルにし、前記複数のサーバ全てからの全ての貸したキャパシティを回収する、プログラム。

30

【請求項 7】

前記複数のサーバのうちの少なくとも 1 つのサーバが、前記サーバ・クラウド中のいずれか他のサーバと共有できるクラウド永続キャパシティを備え、前記クラウド永続キャパシティが、前記サーバ・クラウド中の前記複数のサーバのうちの前記 1 つのサーバ上で永続的にイネーブルにされるキャパシティであって前記サーバ・クラウド中の他のサーバと共有できるキャパシティである、請求項 6 に記載のプログラム。

【請求項 8】

前記複数のリソースがプロセッサを含む、請求項 6 に記載のプログラム。

40

【請求項 9】

前記複数のリソースがメモリを含む、請求項 6 に記載のプログラム。

【請求項 10】

前記複数のサーバのうちの 1 つのサーバがもはや前記サーバ・クラウド中にないせいで前記サーバ・クラウドがもはや完全な状態でないとき、前記クラウド・キャパシティ・オン・デマンド・マネージャが、前記 1 つのサーバから借りた、借りたキャパシティをディセーブルにし、前記 1 つのサーバ上の貸したキャパシティを回収し、まだ前記サーバ・クラウド中にある複数のサーバについての他の借りたキャパシティおよび貸したキャパシティを保持する、請求項 6 に記載のプログラム。

50

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、一般にコンピュータ・システムに関し、より詳細には、サーバ・クラウド中の複数のサーバとのキャパシティ・オン・デマンドを管理することに関する。

【背景技術】

【0002】

コンピュータ・システムに関する今日の問題の1つは、コンピュータ・ハードウェアのコストと、コンピュータ・リソースに対する需要の変動とのバランスをとることである。ほとんどのネットワーク化されたコンピュータ・システムでは、コンピューティング需要が比較的低いときもあれば、コンピューティング需要が非常に高いときもある。会社が、ピーク需要を満たすことのできるコンピュータ・システムを購入した場合、このコンピュータ・システムのキャパシティの多くは、ピーク時以外の間は使用されないことになる。加えて、ピーク需要を満たすためのキャパシティを購入するにはコストがかさむ。会社が、平均需要を満たすことのできるコンピュータ・システムを購入した場合、コストはより低い、ピーク時の間、コンピュータ・システムの性能は損なわれる。

【0003】

よりフレキシブルな解決法をもたらすための方法の1つは、いくつかのリソースがインストールされているが最初はこれらがディセーブルにされているコンピュータ・システムを、コンピュータ・ユーザが購入できるようにするものである。より大きなキャパシティが必要であると顧客が決定したとき、顧客は、コンピュータ・システムのプロバイダとの手続きに入り、いくつかのリソースを固定期間にわたってイネーブルにすることができる。これは、季節的なピークを有する会社の場合には特にうまくいく。会社は、ピーク・シーズンの間は強化されたコンピューティング・パワーを提供する能力のあるコンピュータ・システムを、妥当なコストで購入することができる。必要時に追加のキャパシティを購入できることは、キャパシティ・オン・デマンドとして知られる。

【0004】

単純な例で説明する。カタログ販売を介して商品を販売する会社が、ホリデー・ショッピングのため毎年11月および12月にピーク需要を経験すると仮定する。この会社は、インストールされているが最初はディセーブルにされている1つまたは複数の追加プロセッサを有するコンピュータ・システムを、購入することができる。次いで会社は、コンピュータ・システムのプロバイダと契約を交わし、設定された期間にわたって追加プロセッサ（複数可）をイネーブルにすることができる。コンピュータ・システムが2つの追加プロセッサを有すると仮定し、また、ピーク購買期間が11月15日～12月14日の30日間に及ぶと仮定する。顧客は、11月15日から始まる60プロセッサ日の追加キャパシティを購入することができる。次いで、この2つの追加プロセッサは、30日間にわたってイネーブルにされることになる（60プロセッサ日の追加キャパシティを提供する）。60プロセッサ日が経過すると、2つの追加プロセッサはディセーブルにされる。

【0005】

サーバ・クラウドは、異なるコンピュータ・システムが共に働くことを可能にする。各サーバは、インストールされているがイネーブルにされていない追加リソース（例えばプロセッサ、メモリなど）を有する場合があるが、これらの追加リソースは、キャパシティ・オン・デマンドに基づいて、必要に応じてイネーブルにすることができる。しかし、クラウド中のサーバがジョブの処理においては通信し協働することができるにもかかわらず、各サーバのキャパシティ・オン・デマンド能力は、他の全てのサーバとは別々であり異なる。したがって、あるサーバが、処理すべき計算集約型のジョブを有し追加のプロセッサ・キャパシティを必要とする場合、このサーバは、サーバ・クラウド中の他のサーバを考慮することなしに、それ自体の利用可能なプロセッサ・キャパシティに基づくやり方でそのようにする。

【発明の概要】

【発明が解決しようとする課題】

【0006】

本発明の目的は、従来技術における課題を解決することにある。

【課題を解決するための手段】

【0007】

クラウド・キャパシティ・オン・デマンド・マネージャが、サーバ・クラウド中のサーバのためのキャパシティ・オン・デマンドを管理する。クラウド・キャパシティ・オン・デマンド・マネージャは、1つまたは複数のサーバからキャパシティを借りることができ、あるサーバから借りたキャパシティをサーバ・クラウド中の異なるサーバに貸す。サーバ・クラウドがもはや完全な状態でないとき、サーバ・クラウド中にもはやないサーバから借りたキャパシティはディセーブルにされ、サーバ・クラウド中にもはやないサーバは、サーバ・クラウドに貸したキャパシティを回収する。

10

【0008】

前述および他の特徴および利点は、添付の図面に示す後続のより詳細な記述から明らかになるであろう。

【0009】

本開示は、添付の図面との関連で述べる。図面では、同様の名称は同様の要素を示す。

【図面の簡単な説明】

【0010】

【図1】クラウド・キャパシティ・オン・デマンド・マネージャを備える装置のブロック図である。

20

【図2】キャパシティ・オン・デマンド・マネージャをそれぞれが有する4つのサーバを含む従来技術のサーバ・クラウド・システムのブロック図である。

【図3】クラウド・キャパシティ・オン・デマンド・マネージャを備えるサーバ・クラウド・システムのブロック図である。

【図4】サーバ・クラウド中のサーバ間で共有されるキャパシティを示す、図3のサーバ・クラウド・システムのブロック図である。

【図5】サーバ・クラウド中のサーバ間でキャパシティを共有する方法の流れ図である。

【図6】クラウド永続プロセッサを含むサーバ・クラウド・システムのブロック図である。

30

【図7】サーバ・クラウド中のサーバ間で共有されるキャパシティを示す、図6に示すサーバ・クラウド・システムのブロック図である。

【図8】クラウド・キャパシティ・オン・デマンド・マネージャをハードウェア管理コンソール中に備えるサーバ・クラウド・システムのブロック図である。

【図9】クラウド・キャパシティ・オン・デマンド・マネージャの一実装形態のブロック図である。

【図10】図9に示す借りたキャパシティおよび貸したキャパシティについての特定の一実装形態となり得る実例的なクラウド・リソース・テーブルを示すブロック図である。

【発明を実施するための形態】

【0011】

40

本明細書における特許請求の範囲および開示は、サーバ・クラウド中のサーバのためのキャパシティ・オン・デマンドを管理するクラウド・キャパシティ・オン・デマンド・マネージャを提供する。クラウド・キャパシティ・オン・デマンド・マネージャは、1つまたは複数のサーバからキャパシティを借りることができ、あるサーバから借りたキャパシティをサーバ・クラウド中の異なるサーバに貸することができる。サーバ・クラウドがもはや完全な状態でないとき、サーバ・クラウド中にもはやないサーバから借りたキャパシティはディセーブルにされ、サーバ・クラウド中にもはやないサーバは、サーバ・クラウドに貸したキャパシティを回収する。

【0012】

図1を参照すると、コンピュータ・システム100は、クラウド・キャパシティ・オン

50

・デマンド・マネージャを備えるサーバ・コンピュータ・システムの適切な一実装形態である。サーバ・コンピュータ・システム100は、IBM eServer System xコンピュータ・システムである。しかし、コンピュータ・システムが複雑なマルチユーザ・コンピューティング装置であろうとシングルユーザ・ワークステーションであろうと組み制御システムであろうと、本明細書の開示がどんなコンピュータ・システムにも等しく適用されることは、当業者なら理解するであろう。図1に示すように、コンピュータ・システム100は、1つまたは複数のプロセッサ110、メイン・メモリ120、大容量記憶インタフェース130、表示インタフェース140、およびネットワーク・インタフェース150を備える。これらのシステム・コンポーネントは、システム・バス160の使用によって相互接続される。大容量記憶インタフェース130は、ローカル大容量記憶デバイス155などの大容量記憶デバイスをコンピュータ・システム100に接続するのに使用される。ある特定のタイプのローカル大容量記憶デバイス155は、読取可能かつ書込可能なCD-RWドライブであり、これは、CD-RW195に対してデータの記憶および読取りを行うことができる。

【0013】

メイン・メモリ120は、データ121、オペレーティング・システム122、クラウド処理メカニズム123、クラウド・キャパシティ・オン・デマンド・マネージャ124、インストールされたリソース125、永続リソース126、クラウド永続リソース127、借りたリソース128、および貸したリソース129を含むことが好ましい。データ121は、コンピュータ・システム100中の任意のプログラムへの入力または任意のプログラムからの出力としての働きをする、任意のデータを表す。オペレーティング・システム122は、マルチタスキング・オペレーティング・システムである。クラウド処理メカニズム123は、サーバ100とサーバ・クラウド中の他のサーバとの間の協働をサポートするソフトウェアである。クラウド・キャパシティ・オン・デマンド・マネージャ124は、サーバ・クラウド中のサーバのためのキャパシティ・オン・デマンドを管理し、あるサーバからキャパシティを借りてこのキャパシティを異なるサーバに貸すことができる。インストールされたリソース125は、使用されるようにイネーブルにされているか否かにかかわらず、サーバ100にインストールされたリソースを含む。永続リソース126は、サーバ100上で永続的にイネーブルにされるリソースを含む。クラウド永続キャパシティ127は、サーバ100がメンバとして属するサーバ・クラウド中の任意のサーバに対して永続的にイネーブルにされるリソースのキャパシティを含む。借りたキャパシティ128は、サーバ100がサーバ・クラウド中の他のサーバから借りたリソースのキャパシティを含む。貸したキャパシティ129は、サーバ100がサーバ・クラウド中の他のサーバに貸したリソースのキャパシティを含む。サーバ・クラウド中の他のサーバからキャパシティを借りることによって、クラウド・キャパシティ・オン・デマンド・マネージャ124は、サーバ・クラウド中の追加のリソースを用いて動作するためのより大きなフレキシビリティおよびより低いコストを提供する。

【0014】

コンピュータ・システム100は、周知の仮想アドレッシング・メカニズムを利用する。この仮想アドレッシング・メカニズムにより、コンピュータ・システム100のプログラムは、複数のより小さい記憶エンティティ（メイン・メモリ120およびローカル大容量記憶デバイス155など）へのアクセスではなく、大きい連続したアドレス空間へのアクセスのみを有するかのよう挙動することができる。したがって、データ121、オペレーティング・システム122、クラウド処理メカニズム123、クラウド・キャパシティ・オン・デマンド・マネージャ124、インストールされたリソース125、永続リソース126、クラウド永続キャパシティ127、借りたキャパシティ128、および貸したキャパシティ129は、メイン・メモリ120中にあるように示されているが、これらのアイテムは必ずしも全てが完全にメイン・メモリ120中に同時に含まれるとは限らないことは、当業者なら認識するであろう。また、「メモリ」という用語は、本明細書ではコンピュータ・システム100の仮想メモリ全体を指すのに総称的に使用され、コンピュ

10

20

30

40

50

ータ・システム１００に結合された他のコンピュータ・システムの仮想メモリを含む場合があることにも留意されたい。

【００１５】

プロセッサ１１０は、１つまたは複数のマイクロプロセッサまたは集積回路あるいはその両方から構築されたものとすることができる。プロセッサ１１０は、メイン・メモリ１２０に記憶されたプログラム命令を実行する。メイン・メモリ１２０は、プロセッサ１１０がアクセスできるプログラムおよびデータを記憶する。コンピュータ・システム１００が起動したとき、プロセッサ１１０は最初に、オペレーティング・システム１２２を構成するプログラム命令を実行する。プロセッサ１１０はまた、クラウド・キャパシティ・オン・デマンド・マネージャ１２４を実行する。

10

【００１６】

コンピュータ・システム１００は単一のプロセッサおよび単一のシステム・バスのみを備えるように示されているが、複数のプロセッサまたは複数のバスあるいはその両方を有するコンピュータ・システムを使用してクラウド・キャパシティ・オン・デマンド・マネージャを実践できることは、当業者なら理解するであろう。加えて、使用されるインタフェースはそれぞれ、計算集約型の処理をプロセッサ１１０からオフロードするのに使用される完全にプログラムされた別々のマイクロプロセッサを備えることが好ましい。しかし、Ｉ／Ｏアダプタを使用してこれらの機能を実施することもできることは、当業者なら理解するであろう。

【００１７】

20

表示インタフェース１４０は、１つまたは複数の表示装置１６５をコンピュータ・システム１００に直接に接続するのに使用される。これらの表示装置１６５は、ノンインテリジェント（すなわちダム）端末である場合と、完全にプログラム可能なワークステーションである場合とがあるが、これらの表示装置１６５を使用して、コンピュータ・システム１００と通信する能力がシステム管理者およびユーザにもたらされる。しかし、１つまたは複数の表示装置１６５との通信をサポートするために表示インタフェース１４０が設けられているが、ユーザおよび他のプロセスとの必要な全ての対話はネットワーク・インタフェース１５０を介して行うこともできるので、コンピュータ・システム１００は必ずしも表示装置１６５を必要とするとは限らないことに留意されたい。

【００１８】

30

ネットワーク・インタフェース１５０は、ネットワーク１７０を介してコンピュータ・システム１００を他のコンピュータ・システムまたはワークステーション１７５に接続するのに使用される。ネットワーク・インタフェース１５０は、ネットワーク１７０が今日のアナログ技法もしくはデジタル技法またはその両方を含むか、あるいは将来の何らかのネットワーキング・メカニズムを介するかにかかわらず、電子デバイスを相互接続するための任意の適切な方法を広く表す。ネットワーク・インタフェース１５０は、ネットワーク１７０上の通信を可能にするハードウェアとソフトウェアの組合せを含むことが好ましい。ネットワーク・インタフェース１５０中のソフトウェアは、適切なネットワーク・プロトコルを使用した他のコンピュータ・システム１７５とのネットワーク１７０経由の通信を管理する通信マネージャを含むことが好ましい。多くの異なるネットワーク・プロトコルを使用してネットワークを実現することができる。これらのプロトコルは、コンピュータがネットワークを介して通信するのを可能にする特殊化されたコンピュータ・プログラムである。ＴＣＰ／ＩＰ（伝送制御プロトコル／インターネット・プロトコル）は、ネットワーク・インタフェース１５０内で通信マネージャによって使用できる適切なネットワーク・プロトコルの一例である。

40

【００１９】

当業者には理解されるであろうが、本発明の態様は、システム、方法、またはコンピュータ・プログラム製品として具体化することができる。したがって、本発明の態様は、完全にハードウェアの実施形態、完全にソフトウェアの実施形態（ファームウェア、常駐ソフトウェア、マイクロコードなどを含む）、または、ソフトウェアとハードウェアの態様

50

を組み合わせた実施形態の形をとることができ、本明細書ではこれらを全て「回路」、「モジュール」、「プロセス」、または「システム」と一般に呼ぶ場合がある。さらに、本発明の態様は、コンピュータ可読プログラム・コードが組み入れられた1つまたは複数のコンピュータ可読媒体において具体化されるコンピュータ・プログラム製品の形をとることができる。

【0020】

1つまたは複数のコンピュータ可読媒体の、任意の組合せを利用することができる。コンピュータ可読媒体は、コンピュータ可読信号媒体またはコンピュータ可読記憶媒体とすることができる。コンピュータ可読記憶媒体は、例えば、電子、磁気、光学、電磁、赤外線、もしくは半導体の、システム、装置、またはデバイス、あるいはこれらの適切な組合せとすることができるが、これらに限定されない。コンピュータ可読記憶媒体のより具体的な例（非網羅的なリスト）は、1つもしくは複数のワイヤを有する電気接続、ポータブル・コンピュータ・ディスク、ハード・ディスク、ランダム・アクセス・メモリ（RAM）、読取専用メモリ（ROM）、消去可能プログラム可能な読取専用メモリ（EPROMもしくはフラッシュ・メモリ）、光ファイバ、ポータブル・コンパクト・ディスク読取専用メモリ（CD-ROM）、光学記憶デバイス、磁気記憶デバイス、またはこれらの任意の適切な組合せを含むことになる。この文書のコンテキストでは、コンピュータ可読記憶媒体は、命令実行システム、装置、もしくはデバイスによって使用されるかまたはそれらに関連して使用されるプログラムを、収録あるいは記憶することのできる、任意の有形媒体とすることができる。

【0021】

コンピュータ可読信号媒体は、例えばベースバンド中にまたは搬送波の一部として、コンピュータ可読プログラム・コードが組み入れられた伝搬データ信号を含むことができる。このような伝搬信号は、電磁、光学、またはこれらの任意の適切な組合せを含めた（ただしこれらに限定されない）様々な形のうちのいずれかをとることができる。コンピュータ可読信号媒体は、コンピュータ可読記憶媒体でない任意のコンピュータ可読媒体であって、命令実行システム、装置、もしくはデバイスによって使用されるかまたはそれらに関連して使用されるプログラムを、通信するか伝搬するかあるいは搬送できる、任意のコンピュータ可読媒体とすることができる。

【0022】

コンピュータ可読媒体上に組み入れられたプログラム・コードは、ワイヤレス、有線、光ファイバ・ケーブル、RFなど、またはこれらの任意の適切な組合せを含めた（ただしこれらに限定されない）、任意の適切な媒体を使用して伝送することができる。

【0023】

本発明の態様に関する動作を実施するためのコンピュータ・プログラム・コードは、Java（R）、Smalltalk（R）、C++などのオブジェクト指向プログラミング言語、および、「C」プログラミング言語、Streams処理言語、または同様のプログラミング言語などの従来の手続き型プログラミング言語を含めた、1つまたは複数のプログラミング言語の任意の組合せで書かれてよい。プログラム・コードは、スタンドアロンのソフトウェア・パッケージとして、完全にユーザのコンピュータ上で実行されるか部分的にユーザのコンピュータ上で実行されるか、一部はユーザのコンピュータ上で実行され一部はリモート・コンピュータ上で実行されるか、または完全にリモート・コンピュータもしくはサーバ上で実行されてよい。最後のシナリオでは、リモート・コンピュータは、ローカル・エリア・ネットワーク（LAN）もしくはワイド・エリア・ネットワーク（WAN）を含めた任意のタイプのネットワークを介してユーザのコンピュータに接続されてよく、または、接続は、外部コンピュータに対して（例えばインターネット・サービス・プロバイダを使用してインターネットを介して）行われてもよい。

【0024】

本発明の態様を、本発明の実施形態による方法、装置（システム）、およびコンピュータ・プログラム製品のフローチャート説明またはブロック図あるいはその両方に関して本

10

20

30

40

50

明細書に述べる。フローチャート説明またはブロック図あるいはその両方の各ブロック、ならびに、フローチャート説明またはブロック図あるいはその両方の中のブロックの組合せを、コンピュータ・プログラム命令によって実現できることは理解されるであろう。これらのコンピュータ・プログラム命令を、汎用コンピュータ、専用コンピュータ、または他のプログラム可能データ処理装置のプロセッサに提供して、マシンを生み出すことができ、したがって、コンピュータまたは他のプログラム可能データ処理装置のプロセッサを介して実行される命令は、フローチャートもしくはブロック図またはその両方の1つあるいは複数のブロック中で指定される機能／行為を実現する手段をもたらす。

【0025】

これらのコンピュータ・プログラム命令はまた、コンピュータ可読媒体に記憶されてよく、このコンピュータ可読媒体は、コンピュータ、他のプログラム可能データ処理装置、または他のデバイスに、特定の方式で機能するよう指示することができ、したがって、コンピュータ可読媒体に記憶された命令は、フローチャートもしくはブロック図またはその両方の1つあるいは複数のブロック中で指定される機能／行為を実現する命令を含む製造品を生み出す。

【0026】

コンピュータ・プログラム命令はまた、コンピュータ、他のプログラム可能データ処理装置、または他のデバイスにロードされてよく、それにより、コンピュータ、他のプログラム可能装置、または他のデバイス上で一連の動作ステップが実施されて、コンピュータ実装プロセスが生み出され、したがって、コンピュータまたは他のプログラム可能装置上で実行される命令は、フローチャートもしくはブロック図またはその両方の1つあるいは複数のブロック中で指定される機能／行為を実現するためのプロセスを提供する。

【0027】

本明細書に開示する方法は、ウェブベースのサービスを提供することの一部として実施することができる。このようなサービスは、例えば、この方法を支払いと引き換えにオンライン・ユーザに提供することを含んでよい。

【0028】

図2を参照すると、サーバ・クラウド中のサーバの実例的な従来技術構成が、サーバ・クラウド・システム200として示されている。サーバ・クラウド・システム200は、4つのサーバ210A～210Dを含み、これらのサーバは全て、サーバ・クラウド230中で何らかのネットワーキング・メカニズムを介して相互接続される。各サーバは、インストールされたプロセッサおよび永続プロセッサを備える。インストールされたプロセッサは、サーバに物理的にインストールされたプロセッサの数を指定し、永続プロセッサの数は、サーバ上で永続的にイネーブルにされるプロセッサの数を指定する。図2の特定の例では、各サーバは、225A～225Dおよび226A～226Dに示すように、8つのインストールされたプロセッサおよび2つの永続プロセッサを備える。従来技術のサーバ・クラウド・システム200中では、各サーバは、それ自体のキャパシティ・オン・デマンド・マネージャ232A～232Dを備え、これらのキャパシティ・オン・デマンド・マネージャは、キャパシティ・オン・デマンドを提供するための既知の方法によって、各サーバのキャパシティを増大させることができる。しかし、サーバごとのキャパシティ・オン・デマンド・マネージャは、サーバ・クラウド230中の他のサーバ上のキャパシティから独立して扱われる。したがって、各キャパシティ・オン・デマンド・マネージャ232A～232Dの機能は、サーバがサーバ・クラウド230の一部であるときと、サーバがサーバ・クラウド230の一部でないときとで違いがない。

【0029】

図3に、従来技術のサーバ・クラウド・システム200にいくつかの点で類似するサーバ・クラウド・システム300を示す。サーバ・クラウド・システム300は、サーバ・クラウド330中で何らかのネットワーキング・メカニズムを介して相互接続される4つのサーバ310A～310Dを含む。サーバ310A～310Dはそれぞれ、図1に示すサーバ・コンピュータ・システム100とすることができる。各サーバは、8つのインス

10

20

30

40

50

トールされたプロセッサ 3 2 5 A～3 2 5 D、および 2 つの永続プロセッサ 3 2 6 A～3 2 6 D を備える。しかし、サーバ 3 1 0 A は追加で、クラウド・キャパシティ・オン・デマンド・マネージャ 1 2 4 も備え、これは、サーバ・クラウド中のサーバのためのキャパシティ・オン・デマンドを管理し、必要時に、あるサーバから、別のサーバに貸されるキャパシティを借りることができる。各サーバ 3 1 0 A～3 1 0 D は、キャパシティを借りることとキャパシティを貸すことの両方ができるので、各サーバは、借りたプロセッサ 3 2 8 A～3 2 8 D、および貸したプロセッサ 3 2 9 A～3 2 9 D を追跡する。3 2 8 A～3 2 8 D は、図 3 では「借りたプロセッサ」として示されているが、実際には、借りた、プロセッサのキャパシティである。同様に、3 2 9 A～3 2 9 D は、図 3 では「貸したプロセッサ」として示されているが、実際には、貸した、プロセッサのキャパシティである。図 3 に示す構成では、各サーバが、その 2 つの永続プロセッサによってそれ自体の処理負荷を扱うことができると仮定する。

10

【0030】

次に図 4 に移るが、この例では、サーバ 3 1 0 D が処理負荷の増大のせいで追加のプロセッサ・キャパシティを必要とすると仮定し、さらに、サーバ 3 1 0 D が、既にイネーブルにされている 2 つの永続プロセッサ 3 2 6 D に加えて、3 つの追加プロセッサのキャパシティを必要とすると仮定する。従来技術では、サーバ 3 1 0 D 上のキャパシティ・オン・デマンド・マネージャが、インストールされたプロセッサのうちのさらに 3 つをイネーブルにすることができ、その結果、サーバ 3 1 0 D には合計 5 つの永続プロセッサがあることになる。しかし、クラウド・キャパシティ・オン・デマンド・マネージャ 1 2 4 は、サーバ・クラウド中の 1 つまたは複数のサーバから一時的に借りてサーバ 3 1 0 D に貸すことのできる未使用キャパシティがサーバ・クラウド中にある可能性があることを認識する。図 4 の特定の例では、サーバ 3 1 0 A、3 1 0 B、および 3 1 0 C の各々が、各サーバ上の 2 つの永続プロセッサのうちの一方を使用してそれらの作業負荷を処理できると仮定する。このことは、各サーバが、追加のキャパシティを必要とするサーバ 3 1 0 D に貸せる 1 つのプロセッサのキャパシティを有することを意味する。したがって、クラウド・キャパシティ・オン・デマンド・マネージャ 1 2 4 は、図 4 の 3 2 6 A～3 2 6 C に示すように、永続プロセッサの数を 2 から 1 に減少させ、図 4 の 3 2 9 A～3 2 9 C に示すように、貸したプロセッサの数を 0 から 1 に増加させる。次いで、クラウド・キャパシティ・オン・デマンド・マネージャは、図 4 の 3 つの借りたプロセッサ 3 2 8 D を提供する矢印付き点線によって示すように、これらの貸したプロセッサ 3 2 9 A～3 2 9 C のキャパシティをサーバ 3 1 0 D に貸すことができる。図 4 で貸されるものおよび借りられるものは、プロセッサ・キャパシティであることに留意されたい。したがって、図 3 に示すサーバ・クラウド・システム 3 0 0 中でイネーブルにされるプロセッサの総数は、4 つの各サーバ上の 2 つの永続プロセッサ 3 2 6 A～3 2 6 B であり、合計 8 つのプロセッサがイネーブルにされる。図 3 に示すサーバ・クラウド・システム 3 0 0 中でイネーブルにされるプロセッサの総数は依然として 8 つであり、これらは、3 つのサーバの各々からの 1 つの永続プロセッサ 3 2 6 A～3 2 6 C と、サーバ 3 1 0 D 上の 2 つの永続プロセッサ 3 2 6 D と、サーバ 3 1 0 D 上の 3 つの借りたプロセッサ 3 2 8 D である。クラウド・キャパシティ・オン・デマンド・マネージャが、あるサーバからキャパシティを借りてこのキャパシティをサーバ・クラウド中の異なるサーバに貸すことができることにより、顧客が特定のサーバのために追加のキャパシティを購入するのではなくサーバ上の未使用キャパシティを利用するのを可能にすることになる極めてフレキシブルなシステムがもたらされる。この結果、サーバ・クラウド中の総キャパシティをより効率的に、よりコスト効果のある方式で使用するシステムとなる。

20

30

40

【0031】

図 5 を参照すると、方法 5 0 0 は、サーバ・クラウドが確立されるときに開始する（ステップ 5 1 0）。サーバ・クラウドが完全な状態であり（ステップ 5 2 0＝はい）、サーバ・クラウド中のサーバが追加のキャパシティを必要とし（ステップ 5 3 0＝はい）、サーバ・クラウド中の 1 つまたは複数の他のサーバが貸すためのキャパシティを有するとき

50

(ステップ540＝はい)は、1つまたは複数の他のサーバはキャパシティを貸し(ステップ550)、サーバは借りたキャパシティを使用する(ステップ560)。サーバ・クラウド中のどのサーバも追加キャパシティを必要としないとき(ステップ530＝いいえ)は、方法500はステップ520にループバックして継続する。同様に、他のどのサーバも貸すためのキャパシティを有さないとき(ステップ540＝いいえ)は、方法500はステップ520にループバックして継続する。方法500は、サーバ・クラウドがもはや完全な状態でなくなるまで(ステップ520＝いいえ)継続することができ、サーバ・クラウドが完全な状態でなくなった時点で、借りたキャパシティはディセーブルにされ(ステップ570)、貸したキャパシティは回収される(ステップ580)。ステップ570および580は、2つの異なる方式で機能する場合があることに留意されたい。第1の実装形態では、あるサーバがもはやサーバ・クラウドのメンバでないせいでサーバ・クラウドがもはや完全な状態でないとき(ステップ520＝いいえ)は、サーバ・クラウド中の全てのサーバ上の借りたキャパシティはステップ570でディセーブルにされ、サーバ・クラウド中の全てのサーバ上の貸したキャパシティは回収される。第2の実装形態では、あるサーバ(紛失サーバ)がもはやサーバ・クラウドのメンバでないせいでサーバ・クラウドがもはや完全な状態でないとき(ステップ520＝いいえ)は、紛失サーバからキャパシティを借りてまだサーバ・クラウド中にあるサーバがあればそのサーバ中で、紛失サーバに対する借りたキャパシティはステップ570でディセーブルにされる。紛失サーバは、それ自体がもはやサーバ・クラウドのメンバでないことを検出すると、ステップ580で、貸したキャパシティを回収する。第1の実装形態は、オール・オア・ナッシング手法であり、サーバ・クラウド中のいずれかのサーバが失われると、全ての借りたキャパシティはディセーブルにされ、全ての貸したキャパシティは回収される。第2の実装形態では、選択的に、紛失サーバから借りたキャパシティがディセーブルにされ、紛失サーバ上の貸したキャパシティが回収されるが、サーバ・クラウド中の残りのサーバは、紛失サーバが失われたことによって影響を受けない借りたキャパシティおよび貸したキャパシティを使用して機能することができる。

【0032】

ステップ520で、サーバ・クラウドが完全な状態であるかどうかの判定は、任意の適切な方式で行うことができる。例えば、特定の一実装形態では、サーバ・クラウド中のサーバ間でトークンが回されて、サーバ・クラウドが維持される。あるサーバがそのトークンを定義済み期間内に送らなかった場合は、このサーバはもはや正しく機能していないと見なされ、このことは、サーバ・クラウドがもはや完全な状態でないことを意味する。代替の一実装形態では、クラウド・キャパシティ・オン・デマンド・マネージャは、サーバ・クラウドのメンバのログをとることができ、サーバ・クラウド中の各サーバに定期的に問い合わせることができる。各サーバが適切な応答によって応答した場合、クラウド・キャパシティ・オン・デマンド・マネージャは、サーバ・クラウドがまだ完全な状態であることを知る。サーバのうちの1つが応答しなかった場合は、クラウド・キャパシティ・オン・デマンド・マネージャは、応答しなかったサーバが正しく機能していないこと、したがってもはやサーバ・クラウド中になことを知る。次いで、クラウド・キャパシティ・オン・デマンド・マネージャは、上に論じた措置を講じて、借りたキャパシティをディセーブルにし、貸したキャパシティを回収することができる。本明細書における開示および特許請求の範囲は、現在知られている方法であろうと将来開発される方法であろうと、サーバ・クラウドが完全な状態であるかどうか判定するための任意の適切な方法に及ぶ。

【0033】

クラウド・キャパシティ・オン・デマンド・マネージャ124は、サーバ・クラウド中のサーバにまたがってキャパシティを管理することができるので、このことは、図6に示す新しい概念を引き起こす。サーバ・クラウド・システム600は、4つのサーバ610A～610Dを含み、これらのサーバは、同じインストールされたプロセッサ325A～325D、永続プロセッサ326A～326D、借りたプロセッサ328A～328D、および貸したプロセッサ329A～329Dを有する。サーバ610A～610Dはそれ

10

20

30

40

50

ぞれ、図1に示すサーバ・コンピュータ・システム100とすることができる。加えて、「クラウド永続プロセッサ」と本明細書で呼ぶ新しい概念が導入され、それにより、永続的にイネーブルにされサーバ・クラウド中のどのサーバによっても使用できる、サーバ上のキャパシティが表される。図6の特定の例では、各サーバ610A～610Dは、2つのクラウド永続プロセッサ627A～627Dを備える。これらのクラウド永続プロセッサについてのこれらのキャパシティは、クラウド永続プロセッサが存在するサーバを含めた、サーバ・クラウド630中のどのサーバによっても使用することができる。

【0034】

図7を参照するが、キャパシティの貸し借りを促した図4の条件と同じ条件、すなわち、サーバ610Dがさらに3つのプロセッサ分の3つのキャパシティを必要とするという条件を仮定する。サーバ610Dは、2つのクラウド永続プロセッサ627Dを使用することができ、図7の矢印付き点線で示すように、第3の必要なプロセッサをサーバ610Cから借りることができる。サーバ610C上の1つの貸したプロセッサ329Cは、クラウド永続プロセッサ627Cの数を2から1に減少させることに留意されたい。クラウド永続プロセッサと呼ばれるこの新しい特徴を定義することによって、クラウド・キャパシティ・オン・デマンド・マネージャは、サーバ・クラウド中のサーバ間でどのリソースが借りられるかまたは貸されるかに関して、より大きなフレキシビリティを有することができる。例えば、クラウド・キャパシティ・オン・デマンド・マネージャ124は、最初に、全てのクラウド永続プロセッサを借りることができる。というのは、これらは、サーバ上のいずれかの永続プロセッサを借りる前にクラウド処理のために確保されるからである。代替方式では、クラウド永続プロセッサは、サーバ・クラウド中のサーバ間で共有することが許可される唯一のキャパシティとすることができる。永続プロセッサとは別に定義されたクラウド永続プロセッサのキャパシティを有することによって、クラウド・キャパシティ・オン・デマンド・マネージャ124は、サーバ・クラウド全体にわたるキャパシティの貸し借りにおいてより大きなフレキシビリティを有する。

【0035】

図3、4、6、および7では、クラウド・キャパシティ・オン・デマンド・マネージャ124は、サーバ・クラウド中の1つのサーバ上にあるように示されている。代替の一実装形態では、クラウド・キャパシティ・オン・デマンド・マネージャ124は、図8に示すように、サーバ・クラウド中の別個のエンティティ上にあってもよい。サーバ・クラウド・システム800では、4つのサーバ810A、610B、610C、および610Dが、サーバ・クラウド830中で相互接続される。サーバ・クラウド830中ではハードウェア管理コンソール820も接続され、ハードウェア管理コンソール820は、クラウド・キャパシティ・オン・デマンド・マネージャ124を備える。ハードウェア管理コンソール820は、サーバ・クラウド中のサーバを構成するのを可能にし、サーバ・クラウド中のサーバ中のリソースおよびキャパシティを管理するためのユーザ・インタフェースを提供する。ハードウェア管理コンソール820はまた、サーバ・クラウド中のサーバを監視することができ、サーバ・クラウドがもはや完全な状態でないときを検出することができる。このように、ハードウェア管理コンソール820は、サーバの関係をサーバ自体に管理させるのではなく、サーバ外の独立した制御ポイントを提供し、これにより、サーバがサーバ・クラウドを去るときがより容易になる。例えば、図7で、サーバ610Aが動作不良を起こして応答しなくなった場合、クラウド・キャパシティ・オン・デマンド・マネージャ124は、もはやその仕事を行うことができない。図8に示すようにクラウド・キャパシティ・オン・デマンド・マネージャ124を別個のハードウェア管理コンソール820中に配置することによって、クラウド・キャパシティ・オン・デマンド・マネージャ124は、どのサーバがサーバ・クラウドを去るにかかわらず機能し続けることができる。

【0036】

図9を参照すると、クラウド・キャパシティ・オン・デマンド・マネージャ124は、クラウド・メンバシップ910、すなわちどのサーバがサーバ・クラウドのメンバである

かを追跡し、借りたキャパシティ 920 および貸したキャパシティ 930 を追跡する。クラウド・キャパシティ・オン・デマンド・マネージャ 124 はまた、キャパシティ照会メカニズム 940 を備え、このキャパシティ照会メカニズム 940 は、サーバ・クラウド中の各サーバに照会して、サーバが貸すためのキャパシティを有するかどうか判定することができる（図 5 のステップ 540 参照）。借りたキャパシティ 920 および貸したキャパシティ 930 は、任意の適切な方式で追跡することができる。例えば、図 10 に示すクラウド・リソース・テーブル 1010 は、借りたキャパシティ 920 および貸したキャパシティ 930 の状況を常に把握するための適切な方法の 1 つを表す。エントリ 1020 は、P3 のキャパシティ ID、プロセッサのリソース・タイプ、および、サーバ A がこのキャパシティの所有元でありこのキャパシティがまだ貸し出されていないことを示す。エントリ 1030 は、P14 のキャパシティ ID、プロセッサのリソース・タイプ、および、サーバ C がこのキャパシティの所有元でありこのキャパシティがサーバ D に貸されたことを示す。借りたキャパシティおよび貸したキャパシティの状況を常に把握することによって、クラウド・キャパシティ・オン・デマンド・マネージャ 124 は、サーバ・クラウドがもはや完全な状態でないときに、借りたキャパシティをディセーブルにすることができる（図 7 のステップ 570 参照）。各サーバは、サーバがもはやサーバ・クラウド中にないときを検出できるメカニズムを備えることが好ましく、それに応答して、クラウド中の他のサーバに既に貸し出したキャパシティがあればそれを回収することになることに留意されたい（図 8 のステップ 580 参照）。

10

【0037】

20

上記の例ではプロセッサについて論じているが、プロセッサは、サーバ・クラウド内で貸し借りできるキャパシティを有するリソースの適切な一例を表す。本明細書における開示および特許請求の範囲は、サーバ中の任意の適切なリソース、およびサーバ・クラウド中の任意の適切なリソースに明白に及び、限定ではないがこれらのリソースは、プロセッサ、メモリ、入出力（I/O）スロット、ネットワーク・アダプタなどを含む。また、クラウド・キャパシティ・オン・デマンド・マネージャによって貸し借りされているものは、リソース自体ではなく、リソースのキャパシティであることにも留意されたい。したがって、サーバ 310D が 2 つの永続プロセッサ 326D および 3 つの借りたプロセッサ 328D を有するとき、このことは、サーバ 310D 中の 8 つのインストールされたプロセッサ 325D のうちの 5 つを使用できることを意味する。「借りたプロセッサ」 328D は、他のサーバから借りた、プロセッサのキャパシティを表す。サーバ上の永続プロセッサと借りたプロセッサの合計は、インストールされたプロセッサの総数を超えることはできないことに留意されたい。

30

【0038】

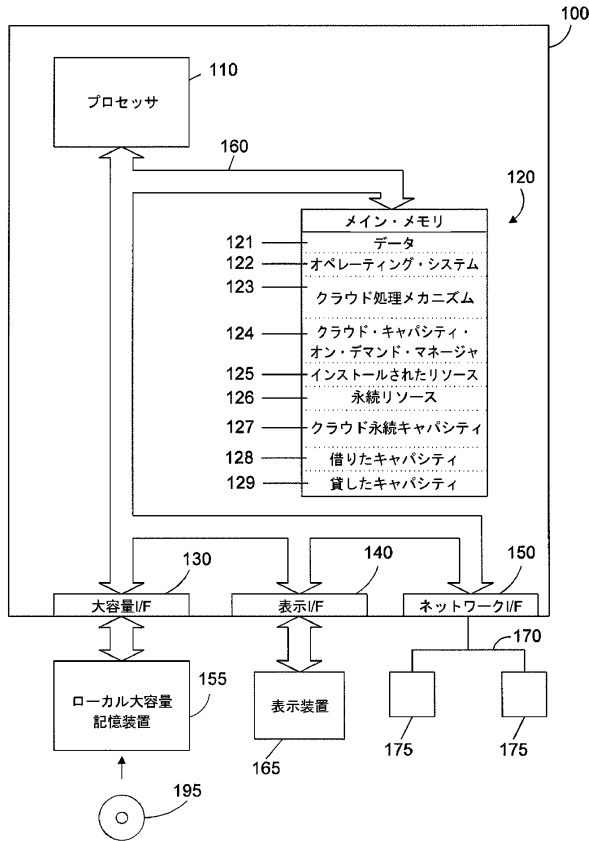
本開示および特許請求の範囲は、サーバ・クラウド中のサーバのためのキャパシティ・オン・デマンドを管理するクラウド・キャパシティ・オン・デマンド・マネージャに関する。クラウド・キャパシティ・オン・デマンド・マネージャは、1 つまたは複数のサーバからキャパシティを借りることができ、あるサーバから借りたキャパシティをサーバ・クラウド中の異なるサーバに貸することができる。サーバ・クラウドがもはや完全な状態でないとき、サーバ・クラウド中にもはやないサーバから借りたキャパシティはディセーブルにされ、サーバ・クラウド中にもはやないサーバは、サーバ・クラウドに貸したキャパシティを回収する。

40

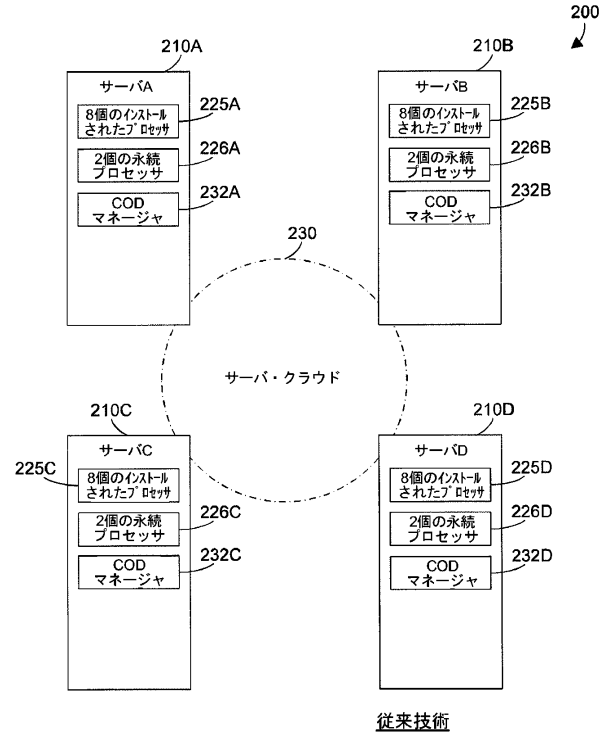
【0039】

請求項の範囲内で多くの変形が可能であることは、当業者なら理解するであろう。したがって、本開示について特に図示および上述しているが、請求項の主旨および範囲を逸脱することなく、形式および詳細におけるこれらおよび他の変更を本開示に加えることができることは、当業者には理解されるであろう。

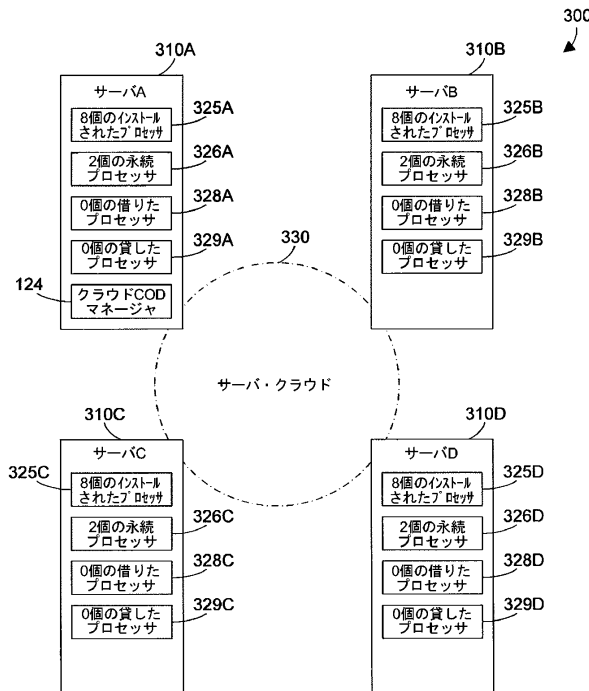
【図 1】



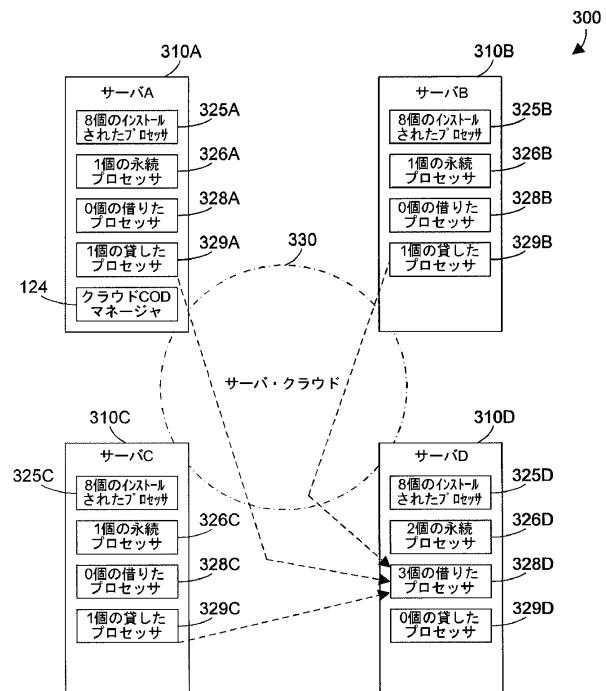
【図 2】



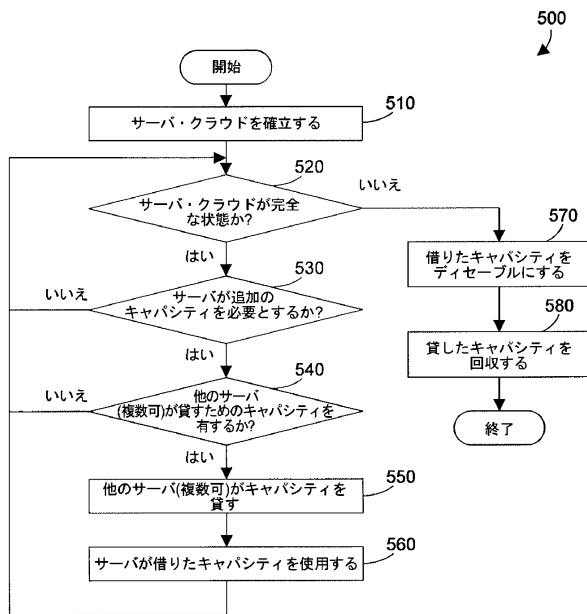
【図 3】



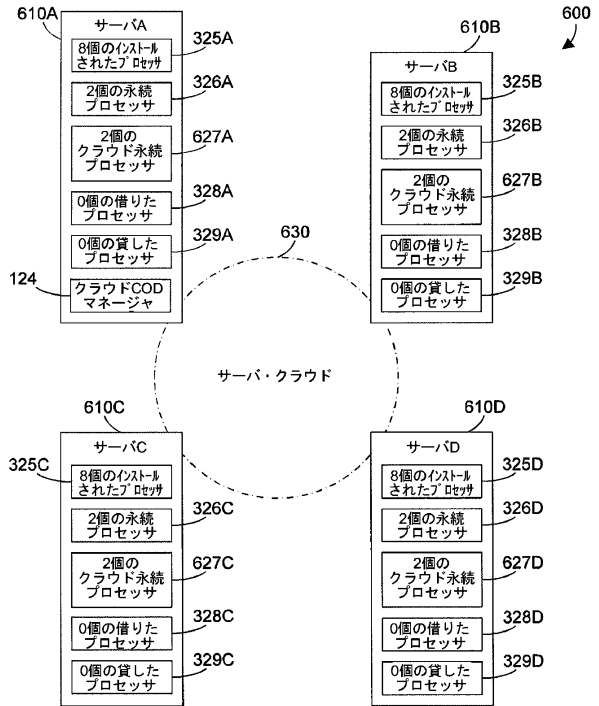
【図 4】



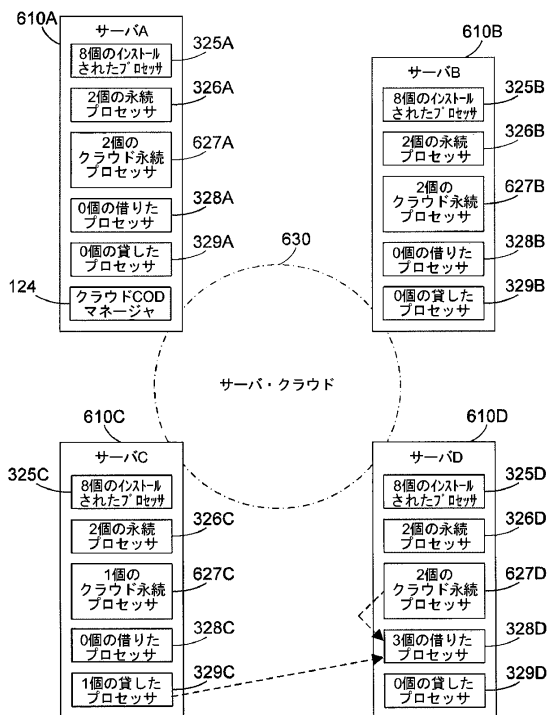
【図 5】



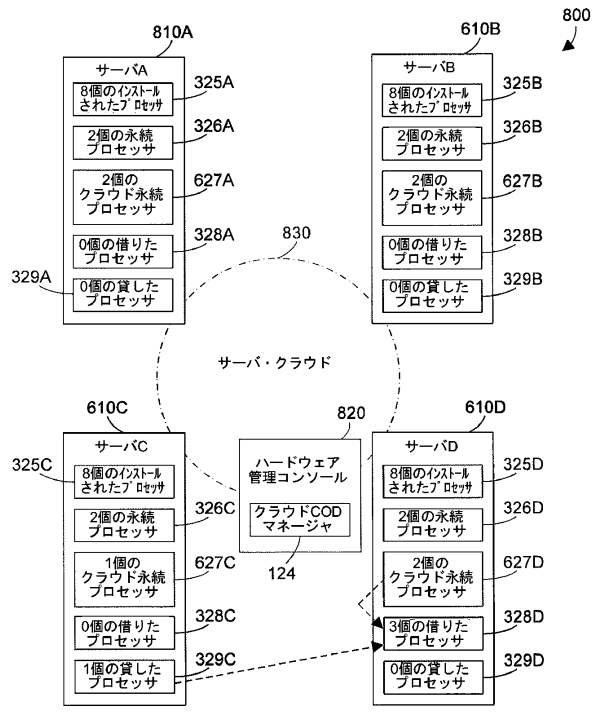
【図 6】



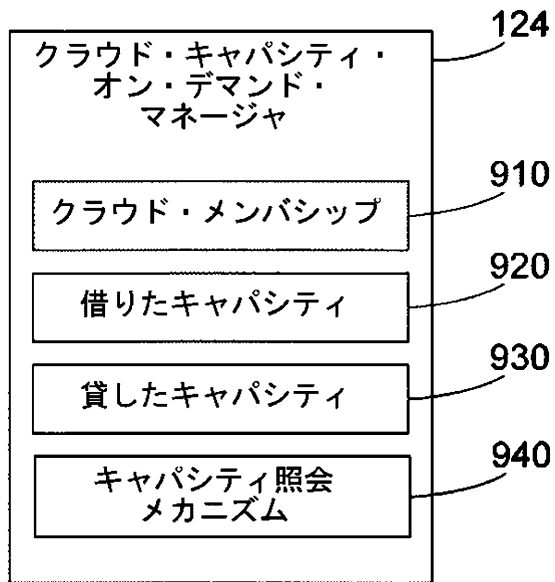
【図 7】



【図 8】



【図 9】



【図 10】

クラウド・リソース・テーブル			
キャパシティID	リソース・タイプ	所有元	貸出先
P3	プロセッサ	サーバA	
⋮	⋮	⋮	⋮
P14	プロセッサ	サーバC	サーバD

フロントページの続き

(72)発明者 オルセン、ポール、エフ

アメリカ合衆国55901 ミネソタ州 ロチェスター ハイウェイ52ノース3605

(72)発明者 シャルト、テリー、エル

アメリカ合衆国55960 ミネソタ州 オロノコ ホワイト・バーチ・コート・エヌダブリュー
1083

審査官 漆原 孝治

(56)参考文献 特開2010-178381 (JP, A)

特開2004-288183 (JP, A)

勝山 恒男, プロアクティブな運用管理を実現するセンタ／ネットワークの統合管理技術 Integrated Management of Data Centers and Networks for Proactive Operation Management, FUJITSU, 富士通株式会社, 2003年 9月10日, 第54巻, 第5号, p.396-401

河井 保博, 新紀年企画・日本型IT革命の将来 ブロードバンド時代のサイト構築 大容量のトラフィックをさばき切るシステムを求めて, 日経インターネットテクノロジー, 日本, 日経BP社, 2001年 8月22日, 第50号, p.46-49

(58)調査した分野(Int.Cl., DB名)

G06F 9/50