

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4292693号  
(P4292693)

(45) 発行日 平成21年7月8日(2009.7.8)

(24) 登録日 平成21年4月17日(2009.4.17)

(51) Int. Cl. F 1  
G 0 6 F 9/50 (2006.01) G 0 6 F 9/46 4 6 2 Z

請求項の数 1 (全 18 頁)

<p>(21) 出願番号 特願2000-211980 (P2000-211980)                  (22) 出願日 平成12年7月7日(2000.7.7)                  (65) 公開番号 特開2002-24192 (P2002-24192A)                  (43) 公開日 平成14年1月25日(2002.1.25)                  審査請求日 平成19年7月3日(2007.7.3)</p>	<p>(73) 特許権者 000005108                  株式会社日立製作所                  東京都千代田区丸の内一丁目6番6号                  (74) 代理人 100100310                  弁理士 井上 学                  (72) 発明者 玉置 由子                  東京都国分寺市東恋ヶ窪一丁目280番地                  株式会社日立製作所中央研究                  所内                  (72) 発明者 庄内 亨                  東京都国分寺市東恋ヶ窪一丁目280番地                  株式会社日立製作所中央研究                  所内</p>
---	---

最終頁に続く

(54) 【発明の名称】 計算機資源分割装置および資源分割方法

(57) 【特許請求の範囲】

【請求項1】

複数のユーザのシステムとネットワークを介して結合された計算機システムであって、前記複数のユーザのシステムからの処理要求パケットにより処理を行うための相互に接続された複数の計算機を含む計算機システムにおいて、各ユーザに異なる計算機群を割り当てる計算機資源分割方法であって、

各ユーザ毎に、使用する計算機数の上限値、及び下限値、割り当てられた計算機群の負荷の閾値を含むサービスレベル条件を入力させる、

上記サービスレベル条件の範囲内で各ユーザへの計算機群の割り当てを設定し、各ユーザからの処理要求パケットを分割投入し、

各ユーザに割り当てられた計算機群の負荷状態をモニタして、あるユーザに割り当てた計算機群の負荷が前記閾値よりも高く、該ユーザに割り当てた計算機群の計算機数が前記サービスレベル条件の計算機数の上限値より小さく、かつ空き計算機が前記計算機システムに存在していれば該ユーザに計算機を追加割り当てすることを特徴とする計算機資源分割方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、計算機システム内の資源をユーザごとに分割する機構および方法に係わり、特に相互にネットワークで結合される複数の計算機からなるシステムにおいて、複数ユーザ

の要求を処理する際に、各ユーザとの間であらかじめ取り決めたサービスレベルを維持するために必要な計算資源をリアルタイムに提供し、かつユーザ間のセキュリティを保持するための資源分割機構および方法に関する。

【0002】

【従来の技術】

情報部門のコスト削減のために、企業内情報システム運用や企業のホームページ管理をASP（アプリケーション・サービス・プロバイダ）業者にアウトソースする事業形態が増加している。ASP業者はさらにデータセンタ業者に計算機資源供の供給、運用管理をアウトソースする形態が多い。

【0003】

データセンタ業者は、多数の計算機資源を用意して複数のユーザ企業に分割使用させることにより、自身の運用コストを削減し、低価格のサービスをユーザ企業に提供する。ユーザ企業間のセキュリティを保持するため、一般には各ユーザ企業ごとに異なる計算機資源およびストレージ資源を割り当てることが多い。

【0004】

またユーザ企業の負荷は時間帯、季節等により変動するため、負荷に応じて割当て資源を増減させる契約形態も多い。特に企業のホームページ管理を請け負っている場合などは、不特定多数の消費者がインターネット経由でアクセスするため負荷の予測がつきにくい。これに対してはユーザ企業側でたとえば新製品発表による負荷増を予測してあらかじめ決めた台数の計算機資源をある一定期間増強するようデータセンタ業者と契約する。データセンタ業者は、増強した計算機資源を、別の期間は別のユーザ企業に割り当てることにより資源の有効活用を図る。このような構成変更を容易にするために、負荷分散装置を複数の計算機資源の前段に配置し、ある期間はユーザ企業Aに、別の期間はユーザ企業Bに計算機資源を使用させるように、データセンタを構成する。負荷分散装置の例としては、アルテオン社のACE directorなどが挙げられる（日経オープンシステムズ 1999.12 no. 81 pp. 128 - 131）。負荷分散装置の設定は上記の契約に基づきあらかじめ人手で設定する。またストレージ資源を増強する必要がある場合は、ストレージの内容をミラーリングする必要がある。

【0005】

さらに、データセンタでは多数のユーザ企業に異なる計算機資源を提供するため、多数の計算機資源を管理する必要が生じ管理コストが増大する。そこで1台あたりの性能が高い計算機資源、たとえば高多重SMP計算機を数少なく導入し、それを複数ユーザ企業が共有するように制御する方法が考えられる。ユーザ企業間のセキュリティを保持するため仮想計算機の機能を利用する。仮想計算機の例としては、日立のプロセッサ資源分割管理機構PRMFがあげられる（HITACマニュアル 8080-2-148-60）。PRMFでは、1つの計算機上で複数のOS（オペレーティング・システム）が動作し、OSごとに独立の資源、たとえば主記憶、ネットワークアダプタなどが割り当てられる。OS間で資源を共有しないため、異なるOS上で実行される異なるユーザ企業のプログラム間のセキュリティは保持される。またPRMFでは、OSごとに割り当てるCPU資源の比率を制御できるように構成されているが、あらかじめ計画した比率変更のみ可能である。

【0006】

ASP、ISP（インターネット・サービス・プロバイダ）業者とユーザの間ではサービスレベル契約を結ぶことが一般的になりつつある。接続性、可用性、レイテンシ性能などのサービスレベル保証を契約する。さらに保証レベル未達の場合の補償契約を結ぶ形態も多い。

【0007】

【発明が解決しようとする課題】

上記の従来技術では、以下の問題点がある。

【0008】

負荷分散装置の設定を契約に基づきあらかじめ人手で設定する方式では、ユーザ企業側で

10

20

30

40

50

予測できない急速な負荷変動にリアルタイムに対応することは困難である。これはユーザ間で異なる計算機を割り当てる場合も仮想計算機を使用する場合も同様である。またストレージ資源を増強する場合、ミラーリングに伴うデータコピーのオーバーヘッドにより迅速な増強が困難である。さらにデータセンタの場合処理内容が定型でなく、ひとつのユーザからの処理要求が複数の計算機にまたがって処理される場合など、レイテンシ性能などは定義、測定しにくい。

【0009】

【課題を解決するための手段】

上記を解決するため、本発明では、データセンタの計算機資源およびストレージ資源をユーザ負荷変動に基づきリアルタイムにユーザ企業ごとに分割割当てする資源分割手段および方法を提供する。

10

【0010】

そのため本発明では、ユーザ識別表を用意し、データセンタへのユーザ要求パケットからユーザ企業を特定し、パケットにユーザ識別子を付加する。さらにユーザ識別子ごとに処理を実行すべき計算機の組を定義する表を管理サーバで生成し、本表を負荷分散装置に動的に設定する。負荷分散装置は設定された計算機の組からいずれかを選択してユーザ要求を実行させる。複数の負荷分散装置がある場合は、管理サーバは負荷分散装置間で本表が整合するよう制御する。さらに、管理サーバでは各計算機の稼動状況をモニタし、サービスレベル契約を満たしているか調べ、必要であれば計算機資源の削減・増強を行う。具体的には上記の計算機の組の定義表を変更して、負荷分散装置に再設定する。さらに管理サーバはユーザ識別子対応に割当てた計算機資源量やサービスレベル契約を守れたかの履歴を作成し、課金情報を作成する。またデータセンタ全体の処理スループットを計測するために、データセンタに入力されるユーザ要求数と返答数をユーザ識別子ごとに測定し集計する。

20

【0011】

また本発明の別の実施形態では、仮想計算機機構を有する計算機でデータセンタを構成する。ユーザ企業ごとに1つのOSで制御される仮想計算機機構を与え、管理サーバは各計算機に各計算機機構のCPU時分割使用割当て%を動的に設定する。さらに、管理サーバでは各計算機の稼動状況をモニタし、サービスレベル契約を満たしているか調べ、必要であればCPU時分割割当て%の削減・増強を行う。

30

【0012】

【発明の実施の形態】

以下、図面を参照しながら、本発明の実施の形態を説明する。

【0013】

(1) 第1の実施形態

図1は、本発明で対象とするデータセンタが、インターネット(I I 0)を介してユーザ企業A(A A 0)、ユーザ企業B(B B 0)、A社やB社のホームページにアクセスする一般消費者(c 0、c 1)と接続されている例を示す。ここでクライアントa 0、a 1、a 2はA社システムのプライベート・ネットワーク・アドレスを有しており、ゲートウェイA 0、A 1を介してVPN(バーチャルプライベートネットワーク)経由でデータセンタ内のゲートウェイD 0と接続する。なお、クライアントc 0、c 1からの要求に関しては第3の実施形態において述べる。

40

【0014】

図2は、データセンタD D 0の構成を示す。本例ではユーザ要求に対しwebブラウザインターフェースを提供するwebサーバ群、webサーバを経由して発生するアプリケーションプログラムを動作させるAPサーバ群、アプリケーションプログラムからデータベースアクセス要求が出たときにそれを処理するDBサーバ群の3層構成になっている場合を示す。

【0015】

図22は、ユーザ企業Aが本データセンタと使用条件契約を結ぶときの入力画面の一例を

50

示している。本例では、ゲートウェイD0に入力されるアクセス要求が、ユーザ企業Aに属するユーザからのアクセス要求であることを識別するために、要求パケットのアクセス元IPアドレスがA0またはA1であることを利用する契約を結んでいる。またユーザ企業Aはデータセンタ内のwebサーバ群、APサーバ群、DBサーバ群のすべてを使用し、ユーザ企業Aのユーザ要求により起動されるプログラムは、webサーバのIPアドレスとしてa100、APサーバのIPアドレスとしてa200、DBサーバのIPアドレスとしてa300を使用する契約を結んでいる。

【0016】

図23は、ユーザ企業Aが本データセンタとサービスレベル契約を結ぶときの入力画面の一例を示している。本例では、ユーザ企業Aのために、webサーバ、APサーバ、DBサーバいずれも必ず最低2台を割当て、またそのすべてをCPU稼働率50%未満で稼働させる。稼働率が50%以上となる場合は各webサーバ、APサーバ、DBサーバの割当て数を最大8台までは増加する契約を結んでいる。本例では入力画面中にチェック印が入力されていないが、たとえばデータセンタの出口における出力トランザクションのスループット、出力トランザクションと入力トランザクションのスループット比、トランザクション処理のレイテンシをサービスレベル契約とすることも可能である。

10

【0017】

上記のような入力画面による契約に基づき、A社にはwebサーバa10、a11、APサーバa20、a21、DBサーバa30、31、が与えられ、B社にはwebサーバb10、b11、APサーバb20、b21、DBサーバb30、b31が与えられるものとする。またストレージS1はボリューム単位にA社、B社に割り当てられる。ボリュームV0がA社用、V1がB社用である。ストレージS2、S2も同様に割り当てられているが記述を省略している。サーバ群y10～y31は、A社、B社の負荷が大きくなったときに割り当てるための予備サーバ群である。

20

【0018】

A社が使用するIPアドレスは、webサーバがa100、APサーバがa200、DBサーバがa300とする。またB社も同様の入力画面により、webサーバb100、APサーバb200、DBサーバb300というIPアドレスを使用するものとする。

【0019】

以下図を参照しながら、サーバ群および負荷分散装置が、ユーザAの要求をサーバ群a0～a31を使用して如何に処理するかを示す。

30

【0020】

図1においてクライアントa0がゲートウェイA0に送出した要求パケットの構成は、図7(1)1200に示すとおりである。パケットの先頭が先サーバのアドレス、次項が送り元クライアントのアドレスである。ゲートウェイA0はパケットをインターネットI10に送出するときにVPNのカプセル化を行い、図7(2)のパケット1201を生成する。ゲートウェイD0ではこのパケットをアンカプセル化する。本技術は公知である。

【0021】

図3はデータセンタの入口にあるゲートウェイD0の構成図である。ゲートウェイD0は信号線I0から入力された図7(2)のパケットに対しアンカプセル化を行うとともに、ユーザ識別表T10を参照してユーザ識別子#0を得、パケットに#0を付加し、図7(3)のパケット1202を生成して信号線L10に送出する。ユーザ識別表T10は、あらかじめ図22のユーザ条件入力画面に基づき管理サーバC0により生成され、信号線L0を介して設定されている。すなわち、本データセンタに送り元アドレスA0またはA1でアクセスした要求はユーザ識別子#0、すなわちA社ユーザとみなされる。

40

【0022】

ゲートウェイD0はパケット生成と同時に、入出力結果格納表T11に、ユーザ識別子#0の入力要求が通過したことを計数手段1003を用いて積算して記録する。

【0023】

50

パケット1203を信号線L10を介して受け取った負荷分散装置d100は図5(1)に示すサーバアドレス対応表T30が保持されている。本表は、ユーザ識別子ごとにユーザアプリケーションで認識しているサーバへの要求を、どの実サーバに送出すべきかを記録しているものである。パケット1202のユーザ識別子は#0、あて先a100なので、負荷分散装置d100は表T30を参照してあて先サーバをa10かa11のいずれかに変更し、図7(4)のパケット1203を生成する。あて先の選択および変更技術自身は公知である。

**【0024】**

webサーバa10はパケット1203を受け取り、a200へのアクセス要求であるパケット1204(図7(5))を生成する。本パケットはバスL110を介して負荷分散装置d200に受け取られる。負荷分散装置d200は図5(2)のサーバアドレス対応表T31を保持しており、それに従い、あて先サーバa200をa20に変更したパケット1205(図7(6))を生成する。

10

**【0025】**

以下同様にAPサーバa20がパケット1206を生成し、サーバアドレス対応表T32(図5(3))を保持する負荷分散装置d300がパケット1206をパケット1207に変更し、DBサーバa30で処理が行われる。DBサーバa30からAPサーバa20、webサーバa10、クライアントa0に返答が返るのも同様であり、順次パケット1208~1214が生成される。ゲートウェイD0は返答パケットをゲートウェイA0に対して送出するときに、入出力結果格納表T11に、ユーザ識別子#0の出力要求が通過したことを計数手段1003を用いて積算して記録する。

20

**【0026】**

図示していないが、ユーザ企業Bからの要求があった場合は、上記と同じ手順でゲートウェイD0でユーザ識別子#1が付加され、同様にサーバ群b10~b31で処理されることとなる。

**【0027】**

以上により、ユーザAとBの処理を実行するサーバ群はサーバa10~a31、b10~b31に夫々分割される。

**【0028】**

ストレージへのアクセスについては図2のストレージS0を例に説明する。ストレージS0は信号線L120によりすべてのwebサーバにより共有されている。ただし、各サーバはストレージアクセス時にユーザ識別子をアクセス要求に付加する。一方ストレージS0は図6のボリュームアクセス権限表T33を保持している。本表はユーザ識別子ごとにどのボリュームのアクセスが許可されているかを示したものである。もしユーザ識別子#1のアクセス要求がボリュームV0をアクセスしようとした場合、ストレージS0は本表を参照し、アクセスを拒否する。これにより、ストレージS1がすべてのwebサーバに共有されていても、ユーザAとBの間のセキュリティが保たれる。

30

**【0029】**

図2において、管理サーバC0は信号線L100、L200、L300を介してサーバ群および負荷分散装置の動作状況をモニタリングしている。モニタリングの内容はユーザとのサービスレベル契約の内容や、モニタリングプログラムの機能による。たとえばCPU稼働率や負荷分散先履歴などである。またモニタリングプログラムは管理サーバC0で動作してもよいし、各サーバ群、負荷分散装置上で動作してもよい。また管理サーバC0は信号線L0を介してゲートウェイD0からユーザごとの入出力結果表T11の内容を入手する。

40

**【0030】**

図4は管理サーバC0の構成図である。T19はユーザ識別表であり、制御プログラムP20が図22のユーザ条件入力画面に基づき設定する。T20はユーザごとのサービスレベル契約内容表であり、制御プログラムP20が図23のサービスレベル条件入力画面に基づき設定する。この場合、ユーザ識別子#0のユーザに対してはwebサーバ、APサ

50

サーバ、DBサーバいずれも最低2台を与え、与えたすべてのサーバでCPU稼働率50%未満でプログラムを動作させ、それに違反しそうな場合は8台まではサーバ数を増加する契約となっている。またユーザ識別子#1のユーザに対しては、webサーバ、APサーバ、DBサーバいずれも最低2台を与え、データセンタのからのアクセス返答スループットは秒あたり30件以上を維持し、それに違反しそうな場合は6台まではサーバ数を増加する契約となっている。制御プログラムP20はモニタリング結果とサービスレベル契約内容表T20を照合し、現在の資源割当てがサービスレベル契約を満たしているかを調べ、その結果をサービス履歴格納表T21に格納する。サービス履歴格納表T21には、たとえばユーザ識別子#0に与えたすべてのサーバのCPU稼働率履歴を記録する。制御プログラムP20はモニタリング結果がサービスレベル契約を満たしていない場合は、割当

10

【0031】

上記制御を行うために、制御プログラムP20が最初に資源を分割する手順を図20を用いて説明する。

20

【0032】

最初に図22のユーザ条件入力画面に示される情報を入力し、ユーザ識別表T19を作成する(1901)。つづいてその情報を信号線L0を介してゲートウェイD0にも設定する(1902)。

【0033】

さらに図23のサービスレベル条件入力画面に示される情報を入力し、サービスレベル契約内容表T20、およびサーバアドレス対応表T23中の仮想addr欄を作成する(1903)。続いてサーバの割当てを、webサーバ群、APサーバ群、DBサーバ群ごとに行う。具体的には、サービスレベル契約内容表T20を参照して各ユーザに最低2台ずつサーバを与えるべきことを検出すると、サーバ割当て管理表T22およびサーバアドレス対応表T23の実addr欄を作成する(1904)。続いて作成したサーバアドレス対応表T23の必要な部分のコピーを信号線L100, 200, 300を介して負荷分散装置d100, 200, 300に設定する(1905)。

30

【0034】

さらに、サービスレベル契約内容表T23に基づき、サービス履歴格納表T21を作成する。具体的には、ユーザ#0にはCPU稼働率履歴を記録する欄を、ユーザ#1にはトランザクション出力スループット履歴(図示せず)を記録する欄を作成する。

【0035】

以上により資源分割制御に必要な情報が生成され、ゲートウェイD0、負荷分散装置d100, 200, 300に設定され、正しく資源分割された状態でシステムが動作開始できる。

40

【0036】

続いて以下に、制御プログラムP20が負荷増大時に資源割当てを変更する手順を図8を用いて説明する。

【0037】

前述したように信号線L100, 200, 300, 0を介してシステムの稼働情報をモニタし(1301)、ユーザ識別子ごとに稼働情報を集計してサービス履歴格納表T21に格納し(1302)、サービスレベル契約内容表T20と比較した後(1303)、まずサービスレベル契約に照らしてサーバを削減できないか検討する(1304)。削減可能

50

かどうかの判断方法としては、CPU稼働率とサーバ台数の積に対して比例計算を行う方法が挙げられる。たとえばユーザ#0のサービスレベル条件はCPU稼働率50%未満であるが、現在4台がwebサーバとして与えられており、いずれもCPU稼働率が25%未満であれば、単純な比例計算としてはwebサーバ数を2台まで削減してよいと判断できる。実際はこれに経験から与えられる種々の安全計数を掛けて判断する。削減可能であれば、削減対象であるサーバへの処理停止指示を信号線L100, 200, 300のいずれかを介して通知する。通知されたサーバはプログラムの処理を終了して、使用している資源を解放する。すなわちメモリアドレス変換テーブルやキャッシュ内容の無効化などを行う。解放が終了するとサーバは管理サーバにその旨通知するので、管理サーバはそれを待って、負荷分散装置d100~d300にサーバアドレス対応表T30~T32の変更を指示する。続いてすべての負荷分散装置の内容が一致して変更されたことを確認する。さらに課金計算式を変更する。本例の場合、割当てサーバ数と割り当てた時間の履歴を記録している。課金計算では、単位時間あたりの1サーバ割当てあたりに単価を決めておきそれを課金する。すなわち、総割当てサーバ数と割当て時間履歴と単価を乗じて課金を計算する(1305)。なお、本例では、webサーバ群、APサーバ群、DBサーバ群の割当て履歴を区別して記録しているの、群ごとに単価を変えておき群ごとの割当て数と割当て時間履歴と各単価の積で課金を計算することも可能である。また、本例では図示していないが、サーバごとに実効性能が異なる場合は、サーバ数×実効性能×割当て時間履歴×単価で課金計算できることも明らかである。また、本例ではゲートウェイD0を要求パケットが通過した個数と返答パケットが通過した個数を記録しているが、要求パケットのゲートウェイ通過スループットが比較的安定している時は、返答パケットのゲートウェイ通過スループットがデータセンタ処理能力の目安となる。そこで信号線L0を介してゲートウェイから返答パケットの通過スループットを受け取り、予め契約で定めた基準スループットと比較して課金を計算することも可能である。例えば、基準スループットを満たした時間は規定課金を行ない、満たさなかった時間は違約金分を課金から減ずることができる。また基準スループットの単価を決めておき、(計測スループット/基準スループット×単価)で課金を計算することもできる。また要求パケットの入力スループットが大きく変動する場合は、(返答パケットスループット/要求スループット)に基づき課金を行なう方法もある。

#### 【0038】

図8の説明に戻ると、つづいてサーバ数を増強する必要があるかを検討する(1306)。何台増強すべきかの判断は、削減時と同様に比例計算で行えばよい。増強する必要があるかをサーバ割当て管理表T22を参照して調査する(1307)。もし空きサーバがなければ運用管理者に通知する(1308)。空きサーバがあれば割当てサーバを選択し(1309)、負荷分散装置d100~d300にサーバアドレス対応表T30~T32の変更を指示する。すべての負荷分散装置の内容が一致して変更されたことを確認してから、課金計算式を変更する(1310)。

#### 【0039】

以上は管理サーバC0上の制御プログラムP20の手順の一例であるが、必ずしもこのすべてをP20が実施しなくてもよいことは明白である。たとえば稼働情報の収集や集計は本プログラムが行わずに他のプログラムからの受け取ってもよい。また本プログラムがかならず実施すべき1305, 1310の処理の内容についても、図9の(1)に示すように課金情報を変更しない1401, 1402に夫々置き換えてもよい。またサーバ側に処理停止指示後は新しいユーザ要求を受け付けない機能がある場合などは処理停止完了を待たずにサーバアドレス対応表T30~T32を変更を指示する1403, 1404に夫々置き換えてもよい。

#### 【0040】

なお上記ではストレージ資源のボリュームアクセス権限表を変更していないが、各プログラムはユーザ識別子を付加してストレージにアクセスするので、サーバ割当てが変更され

10

20

30

40

50

てもアクセス権限のないボリュームにアクセスすることは避けられる。

【 0 0 4 1 】

( 2 ) 第 2 の実施形態

続いて仮想計算機機能 P R M F を備えた高多重 S M P サーバを用いてデータセンタを構成する場合の実施形態を説明する。

【 0 0 4 2 】

データセンタとユーザの接続図は図 1 に同じである。

【 0 0 4 3 】

図 1 0 は、w e b サーバ、A P サーバ、D B サーバとして仮想計算機機能 P R M F を備えたサーバをおのおの 1 台ずつ接続した場合を示している。A P サーバ 1 5 0 1、D B サーバ 1 5 0 2 の内部構成は w e b サーバ 1 5 0 0 と同様であり記述省略する。

10

【 0 0 4 4 】

ユーザ条件入力画面は図 2 2 と同じである。すなわち、ユーザ要求パケットの送り元 I P アドレスが A 0 , A 1 である要求のみがユーザ企業 A のパケットとみなされる契約である。またユーザ企業 A が使用する I P アドレスは w e b サーバが a 1 0 0、A P サーバが a 2 0 0、D B サーバが a 3 0 0 である。

【 0 0 4 5 】

図 2 4 は、サービスレベル契約条件入力画面の一例である。本例では、ユーザ企業 A のために、w e b サーバ、A P サーバ、D B サーバいずれも必ず P R M F 機能による C P U 割当てが 5 0 % 以上となるよう制御する契約である。

20

【 0 0 4 6 】

図 1 0 に戻ると、w e b サーバ 1 5 0 0 は、制御ユニット 1 5 0 3、L P A R 制御レジスタ 1 5 0 4、C P U 群 1 5 0 5 および 1 5 0 6、メモリ 1 5 0 7、ネットワークアダプタ a 1 0 0、b 1 0 0、y 1 0 0 から構成される。L P A R とは L o g i c a l P A R t i t i o n ( 論理資源分割 ) の略であり、L P A R 制御レジスタ群 1 5 0 4 は、各 O S に与える資源の分割方法を保持している。

【 0 0 4 7 】

図 1 1 は L P A R 制御レジスタ群 1 5 0 4 が保持する情報の例である。従来技術である P R M F では、本情報のうち U I D の欄以外の情報を有している。L P A R # とは、各 O S ごとに与える資源群に統一的に与える識別子である。ネットワークアダプタは L P A R ごとにあたえる。ネットワークアダプタアドレスは、後述するが制御プログラム P 2 0 により、ユーザ条件入力画面で契約した各ユーザに与えられる I P アドレスと一致するよう設定される。これによりあるネットワークアダプタに入ってきたユーザ要求のパケットは、対応する L P A R の O S 上のプログラムに引き渡される。メモリ割当て欄は、各 L P A R がメモリ 1 5 0 7 のどの部分を使用するかを示す情報である。C P U 割当て % 欄は各 L P A R に属する O S およびその上のプログラムをどのような割合で C P U 上で動作させるかを示す。制御ユニット 1 5 0 3 が本情報を参照して、L P A R の動作割合を制御する。

30

【 0 0 4 8 】

本実施例では U I D 欄を追加し、L P A R とユーザ識別子を一意に対応させる。P R M F の制御により、異なる L P A R 間では資源が共有されないため、ユーザ間のセキュリティを保つことができる。

40

【 0 0 4 9 】

第 1 の実施形態と同様、ユーザ要求がクライアント a 0 w e b サーバ a 1 0 0 A P サーバ a 2 0 0 D B サーバ a 3 0 0 A P サーバ a 2 0 0 w e b サーバ a 1 0 0 クライアント a 0 と伝わる場合を考える。クライアント a 0 は図 1 2 ( 1 ) のパケット 1 2 0 0 を生成する。ゲートウェイ A 0 によりパケット 1 2 0 1 が生成され、ゲートウェイ D 0 によりパケット 1 2 0 2 が生成されるのは第 1 の実施形態と同じである。

【 0 0 5 0 】

パケット 1 2 0 2 は信号線 L 0 を介してアドレス a 1 0 0 を持つネットワークアダプタ a 1 0 0 に渡され、L P A R # 0 上のアプリケーションプログラムに渡される。すなわちユ

50



ーザAのアプリケーションプログラムである。本プログラムはあて先a200を持つパケット1204を生成し、以下同様にAPサーバ1501上のA社のアプリケーションプログラム、DBサーバ1502上のA社のアプリケーションプログラムに渡される。(図示していないが、APサーバ1501内にはネットワークアダプタa200、b200、y200があり、夫々LPAR#0,1,2に対応している。またLPAR#0,1は夫々ユーザ識別子#0,1に対応している。DBサーバ1502においても同様である。)同様にしてDBサーバ1502からAPサーバ1501、webサーバ1501、クライアントa0への返答も正しくA社に割り当てられたLPAR上のアプリケーションプログラムにより実施される。詳しくは説明しないが、上記の動作により図12のパケット1206~1214が順次生成される。

10

**【0051】**

図13は管理サーバC0の構成図である。T40はLPAR割当て管理表、T19はユーザ識別表である。T50はユーザごとのサービスレベル契約内容表である。この場合、ユーザ識別子#0のユーザに対してはwebサーバ、APサーバ、DBサーバのいずれにおいてもユーザ識別子#0のLPARをCPU割当て率を50%以上割り当てる契約となっている。またユーザ識別子#1のユーザに対しては、CPU割当て%を最低20%とし、またデータセンタのからのアクセス返答スループットは秒あたり30件以上を維持し、それに違反しそうな場合はCPU割当て%を増加する契約となっている。制御プログラムP20は信号線L100,200,300,0から得たモニタリング結果とサービスレベル契約内容表T50を照合し、現在の資源割当てがサービスレベル契約を満たしているかを調べ、その結果をサービス履歴格納表T51に格納する。たとえばユーザ識別子#0に対応するLPARの実際のCPU使用率履歴を記録する。またユーザ識別子#1でアクセス返答スループットが秒あたり30件未満の場合は、設定CPU割当て%を増やす。そのためどのユーザにどれだけCPU割当てを設定しているかを示すCPU割当て管理表T52保持している。T52は各webサーバ、APサーバ、DBサーバ内のLPAR制御レジスタ群のCPU割当て%欄と同じの内容を保持している。サービス履歴格納表T51の課金情報欄の操作は第1の実施形態と同様である。

20

**【0052】**

上記制御を行うために、制御プログラムP20が最初に資源を分割する手順を図21を用いて説明する。

30

**【0053】**

最初に図22のユーザ条件入力画面に示される情報を入力し、ユーザ識別表T19を作成する(2001)。つづいてその情報を信号線L0を介してゲートウェイD0にも設定する(2002)。

**【0054】**

さらに図24のサービスレベル条件入力画面に示される情報を入力し、サービスレベル契約内容表T50、およびLPAR割当て管理表T40中のネットワークアダプタ欄を作成する(2003)。

**【0055】**

続いてサービスレベル契約内容表T50を参照してユーザ#0に最低50%、ユーザ#1に最低20%のCPU割当てを行うべきことを検出すると、CPU割当て管理表T52をとLPAR割当て管理表T40中のCPU割当て欄を作成する(2004)。さらに信号線L100,200,300を介してサーバ1500,1501,1502内のLPAR制御レジスタ群にLPAR割当て管理表T40の内容を設定する(2005)。さらに、サービスレベル契約内容表T23に基づき、サービス履歴格納表T21を作成する(2006)。

40

**【0056】**

以上により資源分割制御に必要な情報が生成され、ゲートウェイD0、サーバ1500,1501,1502に設定され、正しく資源分割された状態でシステムが動作開始できる。

50

## 【 0 0 5 7 】

続いて以下に、制御プログラム P 2 0 が負荷増大時に資源割当てを変更する手順を図 1 4 を用いて説明する。

## 【 0 0 5 8 】

稼働情報収集 ( 1 6 0 1 )、稼働情報集計 ( 1 6 0 2 )、サービスレベル契約との比較 ( 1 6 0 3 ) は第 1 の実施形態と同様である。その後 CPU 割当て率を削可能か検討し ( 1 6 0 4 )、削減可能であれば該当するサーバの L P A R 制御レジスタ群の内容を変更するよう指示する。削減可能化の判断方法は第 1 の実施形態と同様である。変更終了を待った後、課金計算式を変更する ( 1 6 0 5 )。本例の場合、割当て CPU 使用率と割り当てた時間の履歴を記録している。課金計算では、webサーバ、APサーバ、DBサーバそれぞれについて単位時間あたりの使用単価を決めておき、単価 × CPU 使用率を合計して課金する。もちろん、webサーバ、APサーバ、DBサーバそれぞれの単価を異なるように設定してもよく、また単価をサーバの実効性能によって定めてもよい。

10

## 【 0 0 5 9 】

つづいて CPU 割当て率を増強すべきか検討し ( 1 6 0 6 )、増強する必要があるれば、該当するサーバに設定しているサーバ割当て率の合計が 1 0 0 % を超えないか調べる ( 1 6 0 7 )。超えるならば運用管理者に通知する ( 1 6 0 8 )。超えなければ該当するサーバの L P A R 制御レジスタ群の内容を変更するよう指示し、変更終了を待った後、課金情報を変更する ( 1 6 0 9 )。

## 【 0 0 6 0 】

( 3 ) 第 3 の実施形態

最後に、不特定多数の一般消費者が企業 A や企業 B の提供するホームページにアクセスする場合の実施形態を説明する。

20

## 【 0 0 6 1 】

データセンタとユーザの接続図は図 1 に同じである。ユーザはクライアント c 0、c 1 である。

## 【 0 0 6 2 】

図 1 5 はデータセンタの構成を示す。第 1 の実施形態と同じく、負荷分散装置 d 1 0 0 により負荷を複数のサーバに分散できる形態とする。説明の簡単化のため、webサーバしかない構成を示す。すべての webサーバは信号線 L 1 2 0 を介してストレージ S 4 を共有している。S 4 には企業 A のホームページ情報を含むファイル F 0 と企業 B のホームページ情報を含むファイル F 1 が格納されている。ホームページ情報はトリー構成になっていて、ルートから順次たどれる構成とする。また企業 A の提供するホームページにアクセスするためのアドレスが a 1 0 0、企業 B のためのアドレスが b 1 0 0 であるとする。

30

## 【 0 0 6 3 】

図 2 5 は企業 A のホームページにアクセスするユーザの条件を契約するための入力画面の一例である。本例では、ゲートウェイ D 0 に入力されるアクセス要求が、企業 A のホームページにアクセスするユーザ群を識別するために、要求パケットのアクセス先 IP アドレスが a 1 0 0 であることを利用する契約を結んでいる。また企業 A のホームページ生成のために使用される IP アドレスは a 1 0 0 である契約を結んでいる。

40

## 【 0 0 6 4 】

クライアント c 0 は企業 A のホームページにアクセスするため図 1 6 ( 1 ) のパケット 1 7 0 0 を生成する。ゲートウェイ D 0 は図 1 7 に示すようにユーザ識別表 T 6 0 を保持しており、パケット 1 7 0 0 のあて先アドレスが a 1 0 0 であることから本パケットがユーザ識別子 # 0 のホームページにアクセスしようとしていることを検知し、図 1 6 ( 3 ) のパケット 1 7 0 2 を生成する。以下負荷分散装置 d 1 0 0 により本アクセス要求は webサーバ a 1 0 または a 1 1 のいずれかに送られる。この場合はサーバ a 1 0 が選択され、パケット 1 7 0 3 ( 図 1 6 ( 4 ) ) が生成される。返答パケットも同様に、負荷分散装置 d 1 0 0 によりパケット 1 7 1 2 に変更され、さらにゲートウェイ D 0 によるパケット 1

50

714 に変形され、クライアント c0 へと返る。

【0065】

管理サーバ C0 内の構成はを図 18 に示す。ルートファイル管理表 T70 が追加されている以外は図 4 と同じである。本表は、ユーザ識別子ごとのホームページのルートページのファイル名を保持している。

【0066】

負荷が増大したときの制御プログラム P20 の手順を図 19 に示す。本手順は図 8 とほぼ同じである。図 8 との相違である手順 1800 のみを説明する。手順 1309 において割当てサーバを選択すると、選択したサーバに対して、ルートファイル管理表 T70 を参照して、これから割り当てるユーザ識別子に対応するルートファイル名を登録するよう指示する。その後は図 4 の手順 1310 と同じく、負荷分散装置 d100 にサーバアドレス対応表 T30 の変更を指示し、変更終了を待ってから、課金情報を変更する。ルートファイル名を変更することにより、新たに割当てされたサーバは正しいホームページをアクセスすることができるようになる。

10

【0067】

【発明の効果】

以上述べたように、本発明ではユーザ企業ごとにユーザ識別子を与え、それに基づき計算資源を与えるとともに、計算機の稼動状況のモニタ結果に基づき自動的にユーザ識別子ごとにサービスレベル契約と比較して計算資源の量を増減できる。これにより、ユーザ企業側で予測できない急速な負荷変動に対してもリアルタイムに計算資源の割当てを変更可能である。

20

【0068】

また計算機資源割当てを変更しても、ストレージ資源を全計算機で共有し、かつユーザ識別子に基づきストレージ側でアクセス権限チェックを行っているため、ミラーリングのオーバーヘッドなしにユーザ間セキュリティを維持することができる。

【0069】

さらにデータセンタに入力される要求と返答の単位時間あたり通過数をユーザ識別子ごとに測定、集計するため、ユーザから見たデータセンタの性能を測定しやすい。

【図面の簡単な説明】

【図 1】 インターネットを介して接続されるデータセンタとユーザの例である。

30

【図 2】 データセンタの構成の例である。

【図 3】 図 2 のゲートウェイの構成図である。

【図 4】 図 2 の管理サーバの構成図である。

【図 5】 図 2 の負荷分散装置の保持するテーブル例である。

【図 6】 図 2 のストレージが保持するテーブル例である。

【図 7】 図 2 の信号線を通過するパケットの構成図である。

【図 8】 図 4 の制御プログラムの定常動作フロー例である。

【図 9】 図 4 の制御プログラムの定常動作フローの別の例である。

【図 10】 データセンタの構成の別の例である。

【図 11】 図 9 の LPAR 制御レジスタ群の保持情報である。

40

【図 12】 図 9 の信号線を通過するパケットの構成図である。

【図 13】 図 9 の管理サーバの構成図である。

【図 14】 図 13 の制御プログラムの定常動作フロー例である。

【図 15】 データセンタの構成のさらに別の例である。

【図 16】 図 15 の信号線を通過するパケットの構成図である。

【図 17】 図 15 のゲートウェイの構成図である。

【図 18】 図 15 の管理サーバの構成図である。

【図 19】 図 18 の制御プログラムの動作フロー例である。

【図 20】 図 4 の制御プログラムの初期動作フロー例である。

【図 21】 図 13 の制御プログラムの初期動作フロー例である。

50

【図 2 2】図 2 のデータセンタを使用するユーザ条件入力画面。

【図 2 3】図 2 のデータセンタを使用するユーザのサービスレベル条件入力画面である。

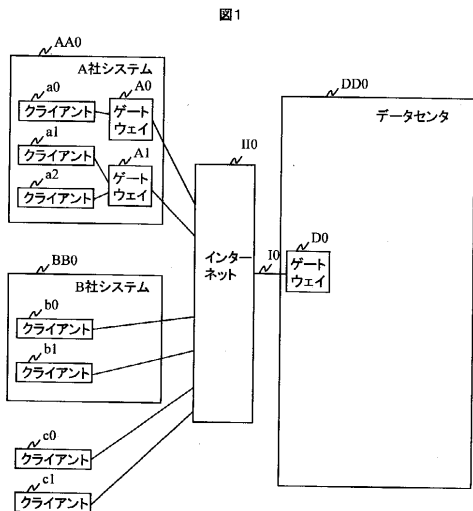
【図 2 4】図 9 のデータセンタを使用するユーザのサービスレベル条件入力画面である。

【図 2 5】図 1 5 のデータセンタを使用するユーザ条件入力画面である。

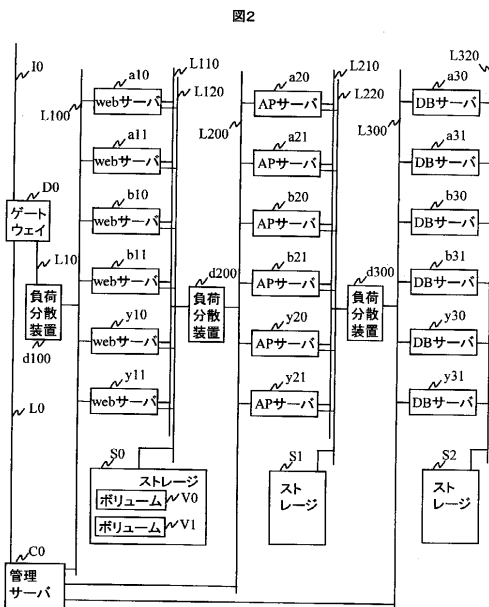
【符号の説明】

DD0 : データセンタ、	
D0 : ゲートウェイ、	
d100 ~ d300 : 負荷分散装置、	
a10 ~ a31、b10 ~ b31、y10 ~ y31 : サーバ、	
C0 : 管理サーバ、	10
S0 ~ S4 : ストレージ、	
T10、T19 : ユーザ識別表、	
T11 : 入出力結果格納表、	
1003 : 計数手段、	
T20 : サービスレベル契約内容表、	
T21 : サービス履歴格納表、	
T22 : サーバ割当て管理表、	
T23、T30 ~ T32 : サーバアドレス対応表、	
T33 : ボリュームアクセス権限表、	
1200 ~ 1214 : パケット、	20
1500 ~ 1502 : サーバ、	
1504 : LPAR 制御レジスタ、	
T40 : LPAR 割当て管理表、	
T50 : サービスレベル契約内容表、	
T51 : サービス履歴格納表、	
T52 : CPU 割当て管理表、	
1700 ~ 1714 : パケット、	
T60、T69 : ユーザ識別表、	
T70 : ルートファイル管理表、	
2100、2400 : ユーザ条件入力画面、	30
2200、2300 : サービスレベル条件入力画面。	

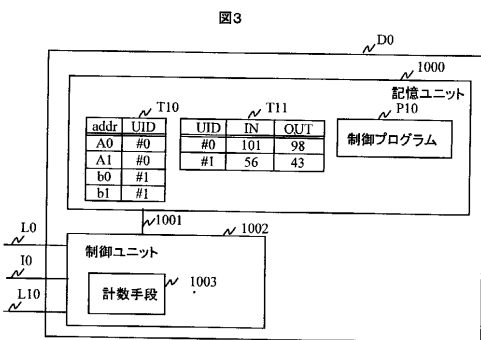
【図1】



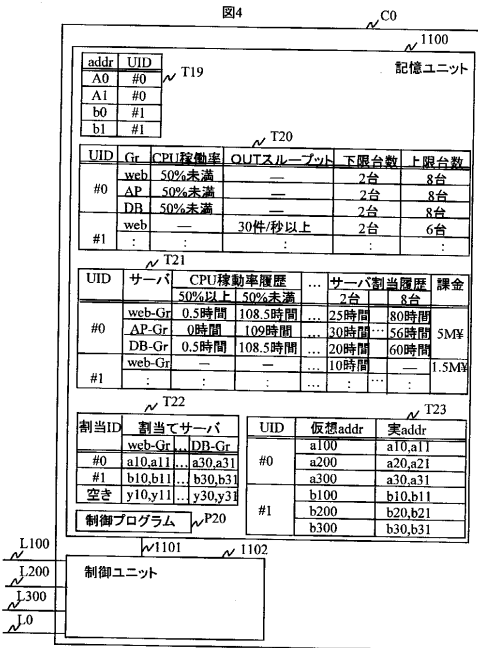
【図2】



【図3】



【図4】



【図5】

図5

(1) 

T30		
UID	仮想addr	実addr
#0	a100	a10,a11
#1	b100	b10,b11

(2) 

T31		
UID	仮想addr	実addr
#0	a100	a10,a11
	a200	a20,a21
#1	b100	b10,b11
	b200	b20,b21

(3) 

T32		
UID	仮想addr	実addr
#0	a200	a20,a21
	a300	a30,a31
	b200	b20,b21
#1	b300	b30,b31

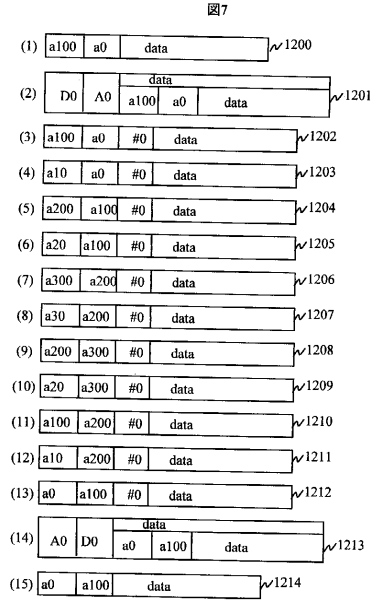
【図6】

図6

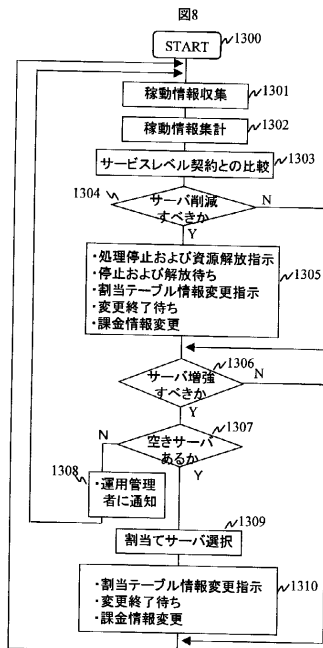
~ T33

UID	ボリューム
#0	V0
#1	V1

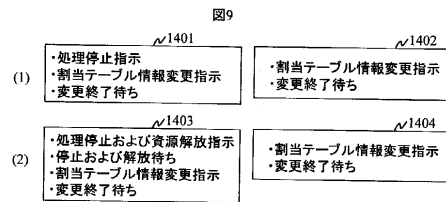
【図7】



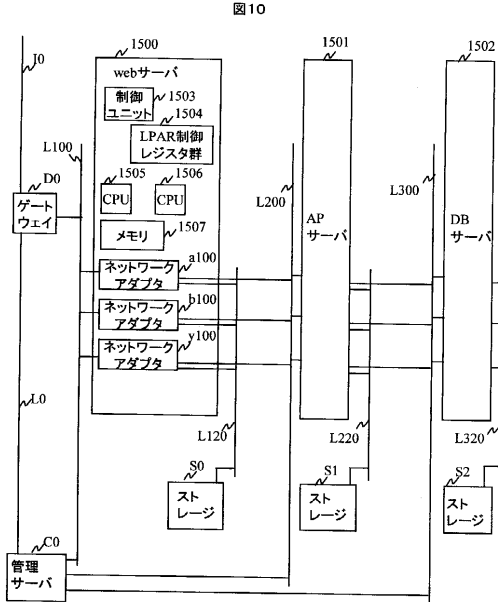
【図8】



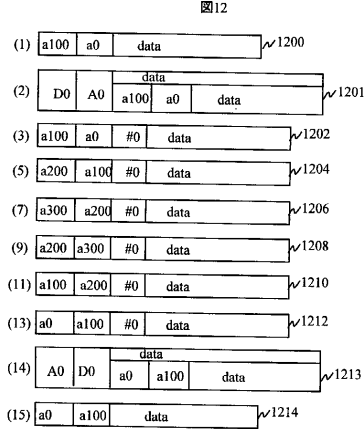
【図9】



【図10】



【図12】

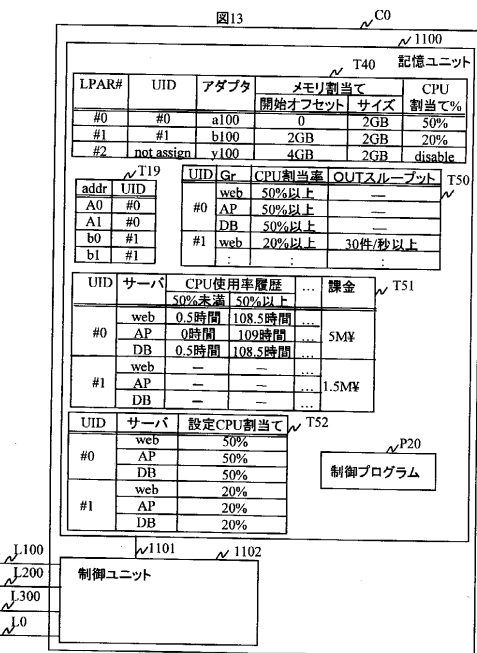


【図11】

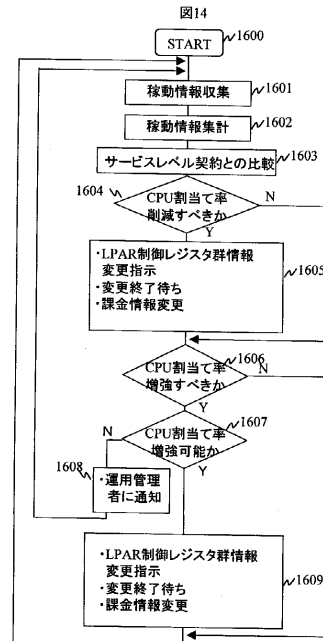
図11

LPAR#	UID	アダプタ	メモリ割当て		CPU割当て%
			開始オフセット	サイズ	
#0	#0	a100	0	2GB	50%
#1	#1	b100	2GB	2GB	20%
#2	not assign	y100	4GB	2GB	disable

【図13】



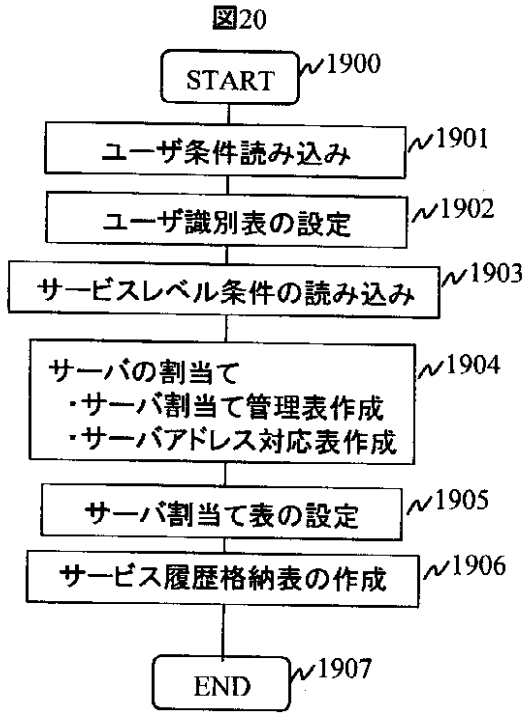
【図14】



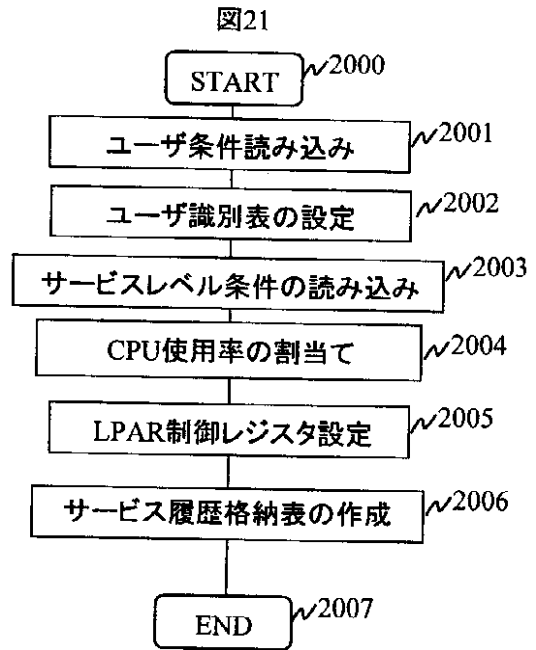




【図20】



【図21】



【図22】

図22 ~2100

ユーザ条件入力画面

ユーザ識別条件

アクセス元IPアドレス  A0,A1

アクセス先IPアドレス

使用サーバ種別 使用IPアドレス

webサーバ  a100

APサーバ  a200

DBサーバ  a300

【図24】

図24 ~2300

サービスレベル条件入力画面

使用サーバ種別 CPU割当て率

webサーバ  50%以上

APサーバ  50%以上

DBサーバ  50%以上

要求性能

出力トランザクションスループット

出力/入力スループット比

トランザクション処理レイテンシ平均

【図23】

図23 ~2200

サービスレベル条件入力画面

使用サーバ種別 サーバ数 使用条件

webサーバ  下限 2台 上限 8台 CPU稼働率 50%未満

APサーバ  2台 8台 50%未満

DBサーバ  2台 8台 50%未満

要求性能

出力トランザクションスループット

出力/入力スループット比

トランザクション処理レイテンシ平均

【図25】

図25 ~2400

ユーザ条件入力画面

ユーザ識別条件

アクセス元IPアドレス

アクセス先IPアドレス  a100

使用サーバ種別 使用IPアドレス

webサーバ  a100

APサーバ

DBサーバ

## フロントページの続き

- (72)発明者 佐川 暢俊  
東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内
- (72)発明者 河辺 峻  
東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内

審査官 殿川 雅也

- (56)参考文献 特開平07-295841(JP,A)  
特開平08-079240(JP,A)  
特開平09-265409(JP,A)  
特開平10-143381(JP,A)  
特開平11-027320(JP,A)  
特開平11-055252(JP,A)  
特開平11-282695(JP,A)  
特表平11-507785(JP,A)  
国際公開第00/014635(WO,A1)  
米国特許第5961599(US,A)  
間瀬憲一, インターネットにおけるSLAの動向と課題, 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, 2000年 7月 6日, Vol. 100, No. 175, pp. 1 - 6

- (58)調査した分野(Int.Cl., DB名)  
G06F 9/46 - 9/54  
G06Q 50/00