

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2013-4000  
(P2013-4000A)

(43) 公開日 平成25年1月7日(2013.1.7)

(51) Int.Cl.  
G06F 9/50 (2006.01)

F I  
G06F 9/46 465D

テーマコード (参考)

審査請求 未請求 請求項の数 7 O L (全 14 頁)

(21) 出願番号 特願2011-137163 (P2011-137163)  
(22) 出願日 平成23年6月21日 (2011.6.21)

(71) 出願人 000004226  
日本電信電話株式会社  
東京都千代田区大手町二丁目3番1号  
(74) 代理人 100064414  
弁理士 磯野 道造  
(74) 代理人 100127720  
弁理士 大石 恵  
(74) 代理人 100162374  
弁理士 中村 新二  
(72) 発明者 金子 雅志  
東京都千代田区大手町二丁目3番1号 日  
本電信電話株式会社内  
(72) 発明者 入江 道生  
東京都千代田区大手町二丁目3番1号 日  
本電信電話株式会社内

最終頁に続く

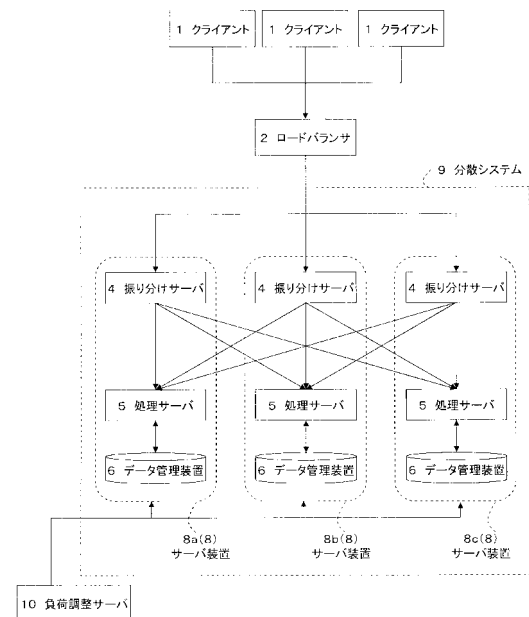
(54) 【発明の名称】 負荷調整方法、負荷調整サーバ、負荷調整用サーバ装置、および、負荷調整プログラム

(57) 【要約】 (修正有)

【課題】分散システムにおける処理サーバ間の負荷の偏りを平滑化するとともに、その平滑化のための処理負荷を軽減する。

【解決手段】負荷調整サーバ10は、各サーバ装置8間の負荷値を比較して、相対的に負荷値が他のサーバ装置8よりも偏っている所定のサーバ装置8を選択し、所定のサーバ装置8上で稼働しているセッションを他のサーバ装置8へと引き継いで、所定のサーバ装置8を分散システム9から離脱させ、データ配置マップにおいて所定のサーバ装置8に対応する配置IDの範囲を変更し、変更されたデータ配置マップが示す配置IDの範囲における要求を分散させる対象に所定のサーバ装置8を追加することで、所定のサーバ装置8を分散システム9へと組み込む。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムで実行される負荷調整方法であって、

前記各サーバ装置は、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、

前記分散システムの前記各サーバ装置と接続されている負荷調整サーバは、

前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、

前記各サーバ装置上でそれぞれ計測された前記各サーバ装置内のリソース使用率を、前記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、

前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサーバ装置を前記分散システムから離脱させ、

前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、

変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする

負荷調整方法。

**【請求項 2】**

クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムで実行される負荷調整方法であって、

前記各サーバ装置は、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、

前記分散システムの前記各サーバ装置と接続されている負荷調整サーバは、

前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、

前記データ配置マップに示される配置IDの範囲の大きさを、前記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、

前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサーバ装置を前記分散システムから離脱させ、

前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、

変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする

負荷調整方法。

**【請求項 3】**

クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムに対して負荷調整を行う負荷調整サーバであって、

前記各サーバ装置は、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、

前記分散システムの前記各サーバ装置と接続されている前記負荷調整サーバは、

前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、

前記各サーバ装置上でそれぞれ計測された前記各サーバ装置内のリソース使用率を、前

10

20

30

40

50

記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、

前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサーバ装置を前記分散システムから離脱させ、

前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、

変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする

10

負荷調整サーバ。

【請求項4】

クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムに対して負荷調整を行う負荷調整サーバであって、

前記各サーバ装置は、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、

前記分散システムの前記各サーバ装置と接続されている前記負荷調整サーバは、

前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、

前記データ配置マップに示される配置IDの範囲の大きさを、前記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、

20

前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサーバ装置を前記分散システムから離脱させ、

前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、

変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする

30

負荷調整サーバ。

【請求項5】

請求項3または請求項4に記載の負荷調整サーバに対して、前記所定のサーバ装置を選択する処理、前記所定のサーバ装置を前記分散システムから離脱させる処理、前記所定のサーバ装置に対応する配置IDの範囲を変更する処理、および、前記所定のサーバ装置を前記分散システムへと組み込む処理を実行させるための負荷調整プログラム。

【請求項6】

クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムに対して負荷調整を行う複数台のサーバ装置のうちの負荷調整用サーバ装置であって、

40

前記各サーバ装置は、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、

前記負荷調整用サーバ装置は、

前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、

前記データ配置マップに示される配置IDの範囲の大きさを、前記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、

前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサー

50

バ装置を前記分散システムから離脱させ、

前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、

変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする

負荷調整用サーバ装置。

【請求項7】

請求項6に記載の負荷調整用サーバ装置に対して、前記所定のサーバ装置を選択する処理、前記所定のサーバ装置を前記分散システムから離脱させる処理、前記所定のサーバ装置に対応する配置IDの範囲を変更する処理、および、前記所定のサーバ装置を前記分散システムへと組み込む処理を実行させるための負荷調整プログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、負荷調整方法、負荷調整サーバ、負荷調整用サーバ装置、および、負荷調整プログラムの技術に関する。

【背景技術】

【0002】

トランザクション処理システムでは、クライアントから処理サーバへとトランザクションの要求が送信されると、処理サーバがその処理を行い、その結果を応答信号としてクライアントへと返信する。そして、複数の処理サーバが協調して動作する分散システムとすることで、クライアントの台数増加に対応する。そして、分散システムのロードバランスは、クライアントからそれぞれ発行される要求に対して、単一の出入り口として受け付けると、それらの要求を複数の処理サーバにそれぞれ分配する。

20

【0003】

そのような分散システムでは、分散システムを構成するサーバ台数を増やせば増やすほど、負荷の偏りが一部の処理サーバに発生してしまう。その結果、高負荷の処理サーバは、新たなトランザクション処理が受け付けられないなどの不都合が発生する。

【0004】

そこで、非特許文献1には、分散システムにおける処理サーバ間の負荷の偏りを平滑化するために、各処理サーバが担当するトランザクションの範囲を、複数の区分に区切るための仮想ノードIDを導入する手法が記載されている。

30

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】入江 道生、西村 豪生、金子 雅志、別所 寿一、飯尾 政美著，“コンシステント・ハッシュ法におけるデータの複製を意識した負荷分散手法”，社団法人電子情報通信学会，信学技報IN2010-77

【発明の概要】

40

【発明が解決しようとする課題】

【0006】

しかし、仮想ノードIDによって各処理サーバの役割分担を細分化する手法では、トランザクションの条件などの不確定要因により、仮想ノードIDの個数を設定することが困難である。例えば、仮想ノードIDの個数が少なすぎると、処理サーバ間の負荷の偏りを十分に平滑化することができない。一方、仮想ノードIDの個数が多すぎると、それらの仮想ノードIDを管理するための余計な負荷が生じてしまう。

【0007】

そこで、本発明は、分散システムにおける処理サーバ間の負荷の偏りを平滑化するとともに、その平滑化のための処理負荷を軽減することを、主な目的とする。

50

## 【課題を解決するための手段】

## 【0008】

前記課題を解決するために、本発明は、クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムで実行される負荷調整方法であって、前記各サーバ装置が、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、前記分散システムの前記各サーバ装置と接続されている負荷調整サーバが、前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、前記各サーバ装置上でそれぞれ計測された前記各サーバ装置内のリソース使用率を、前記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサーバ装置を前記分散システムから離脱させ、前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする。

10

## 【0009】

これにより、分散システムにおける処理サーバ間の負荷の偏りを実測値として検出し、その偏りを平滑化するとともに、その平滑化のための処理負荷を軽減することができる。

20

## 【0010】

本発明は、クライアントからの要求を受けて、その要求を処理する複数台のサーバ装置に対して要求を分散させる分散システムで実行される負荷調整方法であって、前記各サーバ装置が、自身のデータ管理装置において、要求を処理するにあたりアクセスされるデータを配置IDと対応づけて格納し、前記分散システムの前記各サーバ装置と接続されている負荷調整サーバが、前記サーバ装置ごとにその前記サーバ装置内の前記データ管理装置に格納される配置IDの範囲を示すデータ配置マップを記憶手段に格納しており、前記データ配置マップに示される配置IDの範囲の大きさを、前記各サーバ装置の負荷値として取得し、前記各サーバ装置間の負荷値を比較して、相対的に負荷値が他の前記サーバ装置よりも偏っている所定の前記サーバ装置を選択し、前記所定のサーバ装置上で稼働しているセッションを他の前記サーバ装置へと引き継いで、要求を分散させる対象から前記所定のサーバ装置を除外することで、前記所定のサーバ装置を前記分散システムから離脱させ、前記データ配置マップにおいて前記所定のサーバ装置に対応する配置IDの範囲を変更し、変更された前記データ配置マップが示す配置IDの範囲における要求を分散させる対象に前記所定のサーバ装置を追加することで、前記所定のサーバ装置を前記分散システムへと組み込むことを特徴とする。

30

## 【0011】

これにより、分散システムにおける処理サーバ間の負荷の偏りを理論値として検出し、その偏りを平滑化するとともに、その平滑化のための処理負荷を軽減することができる。

## 【0012】

また、本発明は、前記負荷調整方法を実行する負荷調整サーバ、および、その負荷調整サーバを収容するサーバ装置として構成してもよい。

40

## 【発明の効果】

## 【0013】

本発明によれば、分散システムにおける処理サーバ間の負荷の偏りを平滑化するとともに、その平滑化のための処理負荷を軽減することができる。

## 【図面の簡単な説明】

## 【0014】

【図1】本発明の一実施形態に関するトランザクション処理システムを示す構成図である。

50

【図2】本発明の一実施形態に関する負荷調整サーバの詳細を示す説明図である。

【図3】本発明の一実施形態に関する負荷調整サーバの処理概要として、3つのサーバ装置の担当範囲を平滑化する処理を示す説明図である。

【図4】本発明の一実施形態に関する負荷調整サーバが実行する負荷調整処理を示すフローチャートである。

【図5】本発明の一実施形態に関する図4の処理を図3の説明図に適用した一例を示すシーケンス図である。

【発明を実施するための形態】

【0015】

以下、本発明の一実施形態について、図面を参照して詳細に説明する。

10

【0016】

図1は、トランザクション処理システムを示す構成図である。トランザクション処理システムは、クライアント1と、ロードバランサ2と、分散システム9と、負荷調整サーバ10とがネットワークで接続されて構成される。分散システム9内には、複数台のサーバ装置8が収容されており、各サーバ装置8は、振り分けサーバ4と、処理サーバ5と、データ管理装置6とから構成されている。

【0017】

なお、図1のトランザクション処理システムを構成する各装置は、CPU (Central Processing Unit) とメモリとハードディスク (記憶手段) とネットワークインタフェースを有するコンピュータとして構成され、このコンピュータは、CPUが、メモリ上に読み込んだプログラムを実行することにより、各処理部を動作させる。

20

また、図1のトランザクション処理システムを構成する各装置の台数は、図1に例示した台数に限定されず、任意の台数としてもよい。

【0018】

トランザクション処理システムで動作するプログラムの一例として、電話などの呼 (セッション) 処理を行うアプリケーションサーバを実現するための、分散処理ミドルウェアがある。分散処理ミドルウェアは、アプリケーションを分散処理ミドルウェアとは別に用意するもので、アプリケーションからは装置が分散していることを意識することなくサービスの処理を実現することができるようにするものである。このような分散処理ミドルウェアを使うことによって、外部からは単一のアプリケーションサーバのように見せながら、実態は複数台のサーバによる台数効果によって大規模な処理を行うことが可能になる。

30

【0019】

クライアント1は、分散システム9に対してトランザクションを発行するユーザの端末や対向システムである。クライアント1は、処理サーバ5に対して、トランザクションの処理を要求する。

【0020】

ロードバランサ2は、クライアント1からの要求を受け、分散システム9内の複数台の振り分けサーバ4のうち1台を選択し、その選択した振り分けサーバ4に処理を依頼する。

ここで、クライアント1からは、ロードバランサ2という出入り口を介して、分散システム9があたかも単一の装置で動作するシステムとして見えるため、分散システム9内の装置構成を柔軟に変更する (装置の追加や削除など) ことができる。よって、クライアント1は、分散システム9の内部構成を意識することなく、ロードバランサ2のアドレスのみ把握していればよい。

40

【0021】

振り分けサーバ4は、ロードバランサ2からの要求を受け、データ配置マップ23 (後記する図2 (b)) を参照して、トランザクションの要求を処理する処理サーバ5を選択し、その選択した処理サーバ5に処理を依頼する。

処理サーバ5は、振り分けサーバ4からの要求を受け、データ管理装置6内のトランザクションのデータにアクセスすることで、トランザクションの要求を処理する。

50

なお、処理サーバ5とデータ管理装置6とは1:1で接続されており、あるトランザクションのデータは、ある1台のデータ管理装置6が格納している。よって、要求されたトランザクションのデータを管理するデータ管理装置6を収容する処理サーバ5を、振り分けサーバ4が選択する必要がある。

#### 【0022】

ここで、データ配置マップ23のデータ内容によっては、複数台の処理サーバ5のうちの少数の処理サーバ5に処理負荷が集中してしまうこともある。とくに、分散システム9が大規模化し、処理するトランザクション量が多い場合、高負荷のサーバ装置8は、処理限界に達する一方で、低負荷のサーバ装置8が出てきてしまうなど、分散システム9の処理効率を十分に高めることができないケースも発生する。

よって、負荷調整サーバ10は、分散システム9に対して、サーバ装置8間の負荷の偏りを調整するための制御を実行する(詳細は、図2以下の説明で、明らかにする)。

なお、図1では、負荷調整サーバ10をサーバ装置8と独立の装置として稼働させる形態を図示したが、負荷調整サーバ10の各構成要素を、各サーバ装置8に組み込む形態(負荷調整用サーバ装置8とする形態)としてもよい。

#### 【0023】

図2は、負荷調整サーバ10の詳細を示す説明図である。図2(a)の負荷調整サーバ10は、負荷監視部21と、サーバ装置制御部22と、データ配置マップ23と、サーバ装置リスト24とを具備する。

#### 【0024】

負荷監視部21は、サーバ装置8の負荷状況を収集し、負荷の偏り(高負荷または低負荷)を検出する。

サーバ装置制御部22は、負荷監視部21によって負荷の偏りが検出されたサーバ装置8について、分散システム9からの切り離し(離脱)を行った後、そのサーバ装置8が担当するデータ配置マップ23の範囲を変更(高負荷なら減少させ、低負荷なら増加させる)した後、再度、そのサーバ装置8を分散システム9へと組み込む。

なお、切り離し(離脱)とは、ロードバランサ2からみて、トランザクションの割当先の候補から、負荷の偏りが検出されたサーバ装置8を除外する旨を指す。また、再度組み込むとは、ロードバランサ2からみて、トランザクションの割当先の候補に対して、組み込むサーバ装置8を追加する旨を指す。

#### 【0025】

図2(b)は、データ配置マップ23を例示した表である。データ配置マップ23は、処理サーバ5ごとに、その処理サーバ5が担当する配置IDの範囲を対応づけるデータ構造である。

配置IDとは、1つ以上のトランザクションのデータに対して割り当てられるIDであり、同じ配置IDが割り当てられるトランザクションのデータは、同じ1つのデータ管理装置6内に配置IDと対応づけて格納される。

トランザクションのデータとは、トランザクションを処理するときに必要なデータであり、例えば、呼処理における呼の識別子、発信者、受信者、状態遷移を含む。トランザクションの状態遷移とは、話し中、呼び出し中などの呼の状態を示すパラメータである。

#### 【0026】

なお、図2(b)では、配置IDの値域(ID空間)を0度から360度までの360等分した度数法によって例示している。そして、各サーバ装置8が担当する配置IDの範囲は、各サーバ装置8のIPアドレスをキーとしてコンシステントハッシング法におけるハッシュ演算の結果である各サーバ装置8の配置IDを起点とし(例えば、サーバ装置8aなら40度の配置A)、時計回り方向に辿って、次のサーバ装置8の配置ID(例えば、サーバ装置8bの175度の配置B)より1つ少ない配置ID(174度)を終点とする範囲である。

#### 【0027】

つまり、サーバ装置8aの担当範囲は、配置ID「40度~174度」であり、サーバ

10

20

30

40

50

装置 8 b の担当範囲は、配置 ID 「175 度～69 度」であり、サーバ装置 8 c の担当範囲は、配置 ID 「70 度～39 度」である。例えば、配置 ID 「120 度」は、サーバ装置 8 b の担当範囲に含まれる。

さらに、各サーバ装置 8 の負荷の一例として、各サーバ装置 8 の担当範囲である円弧の角度の大きさを例示する。例えば、サーバ装置 8 c の担当範囲は、配置 ID 「70 度～39 度」なので、サーバ装置 8 c の負荷は、 $(70 \text{ 度} - 39 \text{ 度} - 1 \text{ 度}) = 30$  である。

【0028】

【表 1】

#### 24 サーバ装置リスト (図 2 (b) の時点)

サーバ装置	IP アドレス	配置 ID
サーバ装置 8 a	192.168.1.11	配置 A (40)
サーバ装置 8 b	192.168.1.21	配置 B (175)
サーバ装置 8 c	192.168.1.31	配置 C (70)

10

【0029】

表 1 は、サーバ装置リスト 24 を示す。サーバ装置リスト 24 は、トランザクションの要求を処理するサーバ装置 8 (サーバ装置 8 内の処理サーバ 5) を、振り分けサーバ 4 が選択するときの候補を格納するリストである。このサーバ装置リスト 24 は、あらかじめ管理者などによって、分散システム 9 内のサーバ装置 8 がエントリされている。

20

【0030】

サーバ装置リスト 24 は、サーバ装置 8 ごとに、そのサーバ装置 8 に割り当てられる IP アドレスと、その IP アドレスをキーとしたハッシュ値である配置 ID とを対応づけて構成される。例えば、図 2 (b) で示したように、サーバ装置 8 a の配置 ID は、「配置 A (40 度)」である。なお、サーバ装置リスト 24 の配置 ID は、サーバ装置リスト 24 の IP アドレスからいつでも計算できるので、省略してもよい。

【0031】

図 3 は、負荷調整サーバ 10 の処理概要として、3 つのサーバ装置 8 の担当範囲を平滑化する処理を示す説明図である。

まず、図 2 (b) のデータ配置マップ 23 において、負荷調整サーバ 10 は、サーバ装置 8 a (配置 A) の負荷が、他のサーバ装置 8 よりも相対的に高くなっている (負荷 = 25) ことを検出する。

30

【0032】

そこで、図 3 (a) のデータ配置マップ 23 において、負荷調整サーバ 10 は、3 つのサーバ装置 8 のうちの低負荷である (負荷 = 30) サーバ装置 8 c (配置 C) の配置 ID を配置 C 2 へと変更することにより、サーバ装置 8 c の負荷を 110 へと増加させる。この変更に伴い、サーバ装置 8 b (配置 B) の負荷は、図 2 (b) の 105 から、図 3 (a) の 15 へと減少する。

【0033】

さらに、図 3 (b) のデータ配置マップ 23 において、負荷調整サーバ 10 は、3 つのサーバ装置 8 のうちの低負荷である (負荷 = 15) サーバ装置 8 b (配置 B) の配置 ID を配置 B 2 へと変更することにより、サーバ装置 8 b の負荷を 110 へと増加させる。これにより、3 つのサーバ装置 8 の負荷が、「110, 110, 140」と平滑化される。

40

【0034】



## 【表 2】

## 2 4 サーバ装置リスト (図 3 (a) の時点)

サーバ装置	I P アドレス	配置 I D
サーバ装置 8 a	192.168.1.11	配置 A (40)
サーバ装置 8 b	192.168.1.21	配置 B (175)
サーバ装置 8 c	192.168.1.32	配置 C 2 (150)

## 2 4 サーバ装置リスト (図 3 (b) の時点)

サーバ装置	I P アドレス	配置 I D
サーバ装置 8 a	192.168.1.11	配置 A (40)
サーバ装置 8 b	192.168.1.22	配置 B 2 (260)
サーバ装置 8 c	192.168.1.32	配置 C 2 (150)

10

## 【0035】

表 2 は、図 3 の各処理時点におけるサーバ装置リスト 2 4 を示す。前記した各サーバ装置 8 の配置 I D の変更は、各サーバ装置 8 の I P アドレスの変更によって行われる。つまり、ハッシュ演算のキー (入力パラメータ) である I P アドレスを変更すると、ハッシュ演算の出力パラメータである配置 I D も変更される。

20

## 【0036】

図 4 は、負荷調整サーバ 1 0 が実行する負荷調整処理を示すフローチャートである。

## 【0037】

S 1 1 において、負荷監視部 2 1 は、各サーバ装置 8 の負荷状況を集計する。具体的には、負荷監視部 2 1 は、各サーバ装置 8 の相対負荷値を計算することで、複数台のサーバ装置 8 の間での負荷が偏ったサーバ装置 8 を検出する。

「a」を (所定のサーバ装置 8 の負荷値) とし、「A」を (全サーバ装置 8 の負荷値の平均値) とすると、(所定のサーバ装置 8 の相対負荷値) は、以下のいずれかの計算式で計算できる。

30

平均からの差分値の絶対値を計算する式は、次の通りである。

$$(\text{所定のサーバ装置 8 の相対負荷値}) = |(a) - (A)|$$

平均値からの乖離率の絶対値を計算する式は、次の通りである。

$$(\text{所定のサーバ装置 8 の相対負荷値}) = |(a - A) \div A|$$

## 【0038】

なお、サーバ装置 8 の負荷値として、サーバ装置 8 の負荷の実測値として、負荷調整サーバ 1 0 が各サーバ装置 8 に計測させたりソース使用率 (CPU 使用率やメモリ使用率など) の通知を受けて、そのソース使用率を用いてもよい。

一方、サーバ装置 8 の負荷値として、サーバ装置 8 の負荷の理論値として、データ配置マップ 2 3 における配置 I D の範囲を用いてもよい。この理論値を用いることにより、他のサーバ装置 8 の稼働状況 (CPU 使用率など) を測定すること無くサーバ装置 8 の負荷値を取得できるため、各サーバ装置 8 で I D 空間の状況を見て自律的に I D 更新を行うことができる。

40

## 【0039】

S 1 2 において、負荷監視部 2 1 は、S 1 1 の相対負荷値が偏っているサーバ装置 8 が存在するか否かを判定する。S 1 2 で Yes なら S 1 3 に進み、No なら S 1 1 に戻る。

相対負荷値が高負荷に偏っているサーバ装置 8 とは、例えば、そのサーバ装置 8 の相対負荷値 (例えば、200 度) が、他のサーバ装置 8 の負荷平均値 (例えば、100 度) よりも、所定閾値 (例えば、80 度) を超える分、偏っているサーバ装置 8 である (200 > 100 + 80)。

50

相対負荷値が低負荷に偏っているサーバ装置 8 とは、例えば、そのサーバ装置 8 の相対負荷値（例えば、50 度）が、他のサーバ装置 8 の負荷平均値（例えば、100 度）よりも、所定閾値（例えば、30 度）を下回る分、偏っているサーバ装置 8 である（ $50 < 100 - 30$ ）。

【0040】

S13 において、サーバ装置制御部 22 は、偏りが最大のサーバ装置 8 などの相対負荷値が偏っているサーバ装置 8 を分散システム 9 から切り離す。なお、切り離す処理とは、例えば、該当するサーバ装置 8 内の処理サーバ 5（第 1 処理サーバ 5）が、自身とクライアント 1 との間のトランザクションを消去（中止）する処理である。この消去処理に伴い、第 1 処理サーバ 5 は、配下のデータ管理装置 6 に格納されているトランザクションのデータを消去する。

10

【0041】

なお、第 1 処理サーバ 5 の切り離し処理によって、データ配置マップ 23 において第 1 処理サーバ 5 の配置 ID と隣接する他の処理サーバ 5（第 2 処理サーバ 5）の担当範囲に、第 1 処理サーバ 5 の担当範囲が追加される。

よって、第 1 処理サーバ 5 から第 2 処理サーバ 5 に対して、第 1 処理サーバ 5 で消去したトランザクションを行うためのセッションを、引き継ぐ必要がある。そのため、各処理サーバ 5 は、それぞれ自身が有しているセッションに関するデータを、他の処理サーバ 5 との間で共有（データ同期）しておく。そして、第 2 処理サーバ 5 は、引き継がれたセッションを介して、クライアント 1 との間でトランザクションを再実施する。

20

【0042】

S14 において、サーバ装置制御部 22 は、切り離したサーバ装置 8 の配置 ID を更新する。つまり、表 2 で説明したように、配置 ID を変更するために、サーバ装置 8 の IP アドレスを変更する。

そして、サーバ装置制御部 22 は、配置 ID が変更されたサーバ装置 8 を、再度、分散システム 9 に組み込む。なお、S13 で説明したように、組み込まれたサーバ装置 8 が新規に担当する分の配置 ID について、他のサーバ装置 8 からの引き継ぎ処理を行う。

【0043】

S15 において、負荷監視部 21 は、所定時間待ってから、S11 に戻る。よって、図 4 のフローチャートは、所定間隔で繰り返し実行されるので、分散システム 9 の負荷調整を随時行うことが可能となる。

30

【0044】

以上説明した図 4 の処理は、サーバ装置 8 と負荷調整サーバ 10 との間で行われるので、クライアント 1 にとっては、ロードバランサ 2 に変更がない状態が続く。よって、分散システム 9 内部の構成を隠蔽させつつ複数のサーバ装置 8 で分散処理することができ、サーバ装置 8 を追加することによって容易にシステムの性能を向上させることができる。

【0045】

図 5 は、図 4 の処理を図 3 の説明図に適用した一例を示すシーケンス図である。この図 5 では、サーバ装置 8 と負荷調整サーバ 10 との間でやりとりされるメッセージに着目して、説明を行う。

40

【0046】

まず、S101 ~ S104 は、図 3 (a) の配置 C2（サーバ装置 8c）の組み込み処理（S14）を詳細に説明するものである。

S101 において、サーバ装置 8c は、自装置を分散システム 9 へと組み込む旨の組み込み要求を、負荷調整サーバ 10 に通知する。なお、サーバ装置 8c は、S101 の開始前は、分散システム 9 へと組み込まれていない状態（遊休状態）である。よって、遊休状態のサーバ装置 8 は、負荷調整サーバ 10 に対して組み込み要求を送信するようにしてもよいし、負荷調整サーバ 10 が遊休状態のサーバ装置 8 のリストをリソースプールとして管理しておき、負荷調整サーバ 10 がリソースプールから任意のサーバ装置 8 を組み込み対象として選択することで、S101 の処理を代替してもよい。

50

## 【 0 0 4 7 】

S 1 0 2において、負荷調整サーバ10は、サーバ装置8cの配置IDをC（配置ID = 70）から、C2（配置ID = 110）へと更新し、その結果を、図3（a）のデータ配置マップ23およびサーバ装置リスト24に反映させる。

S 1 0 3において、負荷調整サーバ10は、S 1 0 1への応答としての組み込みOKの通知を、S 1 0 2で更新されたデータ（データ配置マップ23およびサーバ装置リスト24）を含めてサーバ装置8cに送信する。

S 1 0 4において、負荷調整サーバ10は、S 1 0 2で更新されたデータを含めたシステム更新通知を、ブロードキャストで各サーバ装置8（サーバ装置8bなど）に通知する。

10

## 【 0 0 4 8 】

そして、負荷調整サーバ10は、図3（a）のデータ配置マップ23で示されるように、配置B（サーバ装置8b）の低負荷を検出する（S 1 1, S 1 2）。

次に、S 1 1 1 ~ S 1 1 5は、図3（b）の配置B（サーバ装置8b）の切り離し処理（S 1 3）を詳細に説明するものである。

## 【 0 0 4 9 】

S 1 1 1において、負荷調整サーバ10は、低負荷が検出されたサーバ装置8bを分散システム9から切り離して再組み込みを行うための再組み込み実施要求を、サーバ装置8bに通知する。

S 1 1 2において、サーバ装置8bは、S 1 1 1を受けて、自装置の分散システム9からの離脱を要求するための離脱要求を、負荷調整サーバ10に通知する。

20

S 1 1 3において、負荷調整サーバ10は、S 1 1 2で離脱要求されたサーバ装置8bを分散システム9から離脱させる。この離脱により、負荷調整サーバ10は、図3（b）で示すように、サーバ装置8bの配置B（175）をデータ配置マップ23から削除するとともに、サーバ装置リスト24のサーバ装置8bのレコードにおける「配置ID」列を空欄にする。

S 1 1 4において、負荷調整サーバ10は、S 1 1 3の結果を通知する離脱OKの通知を、サーバ装置8bに送信する。

S 1 1 5において、負荷調整サーバ10は、S 1 1 3で更新されたデータを含めたシステム更新通知を、ブロードキャストで各サーバ装置8（サーバ装置8cなど）に通知する。

30

## 【 0 0 5 0 】

さらに、S 1 2 1 ~ S 1 2 4は、図3（b）の配置B2（サーバ装置8b）の組み込み処理（S 1 4）を詳細に説明するものである。このS 1 2 1 ~ S 1 2 4の処理は、S 1 0 1 ~ S 1 0 4の処理について、サーバ装置8cからサーバ装置8bへと置き換えただけなので、説明を省略する。

## 【 0 0 5 1 】

以上説明した本実施形態では、複数台のサーバ装置8とロードバランサ2によって構成される分散システム9において、各サーバ装置8の負荷を調整する負荷調整サーバ10を説明した。

40

分散システム9は、セッション制御などの通信制御をクラスタシステムとして実施する。各サーバ装置8への処理の負荷分散には、例えば、コンシステント・ハッシュ手法が用いられる。ここで、各サーバ装置8の負荷の偏りは、主に、各サーバ装置8が担当する配置IDの範囲の大きさの偏りによって発生するので、配置IDの範囲を適宜変更する負荷調整サーバ10を導入することにより、配置IDの範囲を均等にする。

また、各サーバ装置8の負荷の偏りを求める別の方法として、各サーバ装置8のリソース使用率（CPU使用率やメモリ使用率など）を用いてもよい。

## 【 0 0 5 2 】

負荷調整サーバ10は、このようにして負荷の偏りがあるサーバ装置8を検出すると、そのサーバ装置8を分散システム9から離脱させ、配置IDの範囲を変更してから、分散

50

システム 9 へと再度組み込むことで、負荷を調整する。

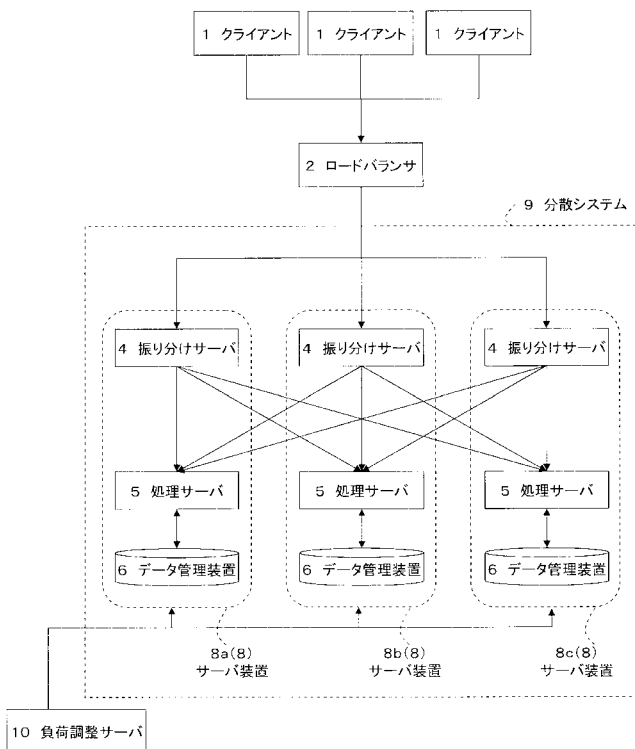
これにより、サーバ装置 8 間での負荷の偏りを是正し、サーバ装置 8 の利用効率を向上させることができる。

【符号の説明】

【 0 0 5 3 】

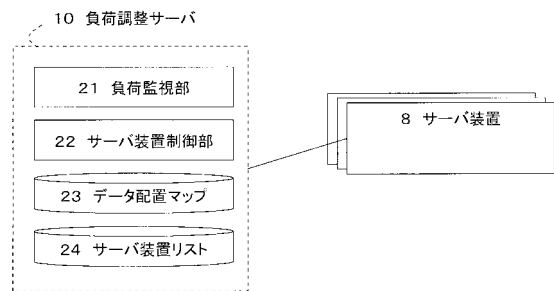
- 1 クライアント
- 2 ロードバランサ
- 4 振り分けサーバ
- 5 処理サーバ
- 6 データ管理装置
- 8 サーバ装置
- 9 分散システム
- 10 負荷調整サーバ
- 21 負荷監視部
- 22 サーバ装置制御部
- 23 データ配置マップ
- 24 サーバ装置リスト

【 図 1 】

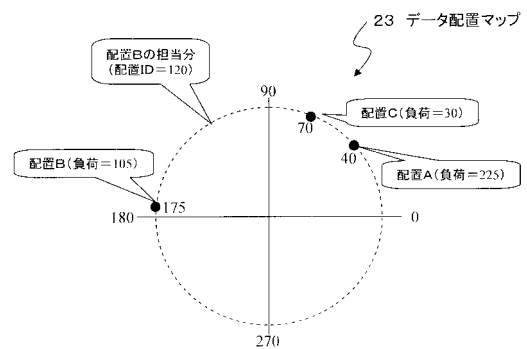


【 図 2 】

(a) 負荷調整サーバ

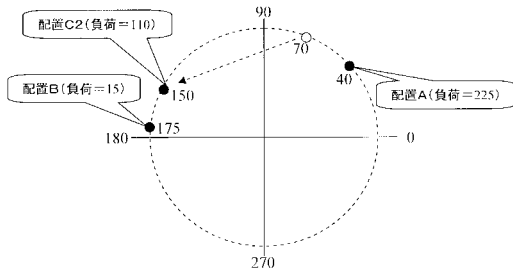


(b) データ配置マップ

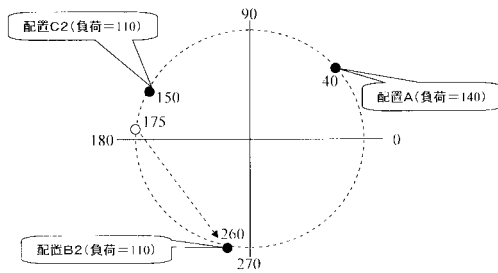


【 図 3 】

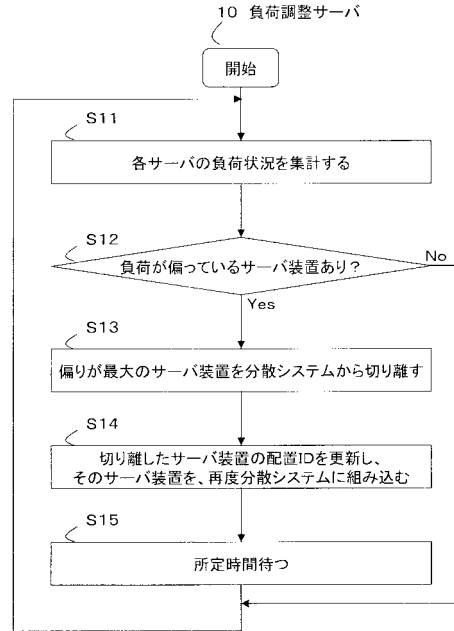
(a) サーバ装置8cの配置IDを変更



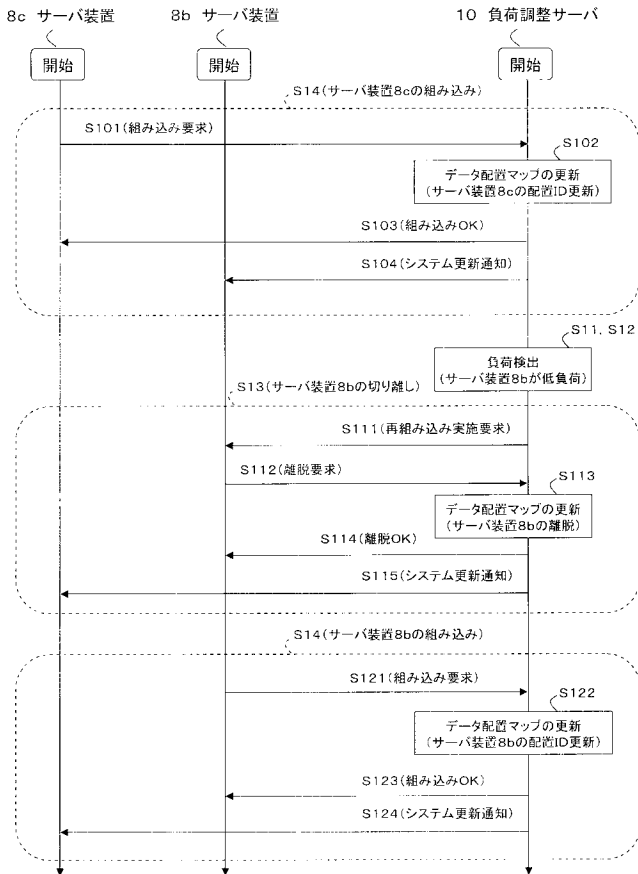
(b) サーバ装置8bの配置IDを変更



【 図 4 】



【 図 5 】



フロントページの続き

(72)発明者 西村 豪生

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内