

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2006年2月9日 (09.02.2006)

PCT

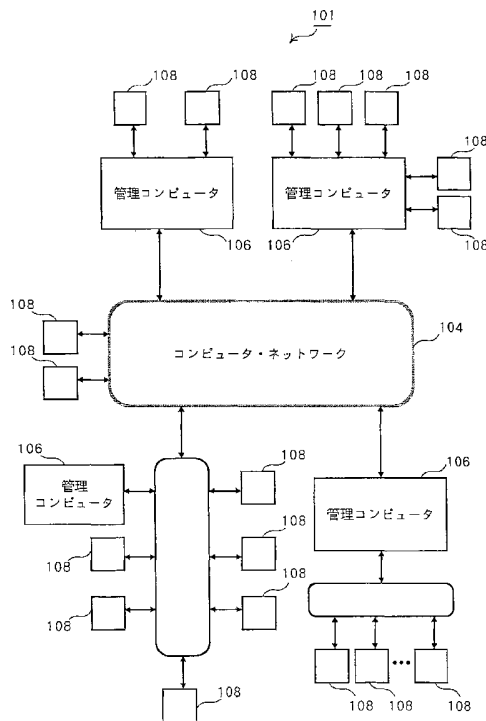
(10) 国際公開番号  
WO 2006/013992 A1

- (51) 国際特許分類: 1070062 東京都港区南青山二丁目6番21号 Tokyo (JP)  
G06F 9/50 (2006.01)
- (21) 国際出願番号: PCT/JP2005/014467
- (22) 国際出願日: 2005年8月1日 (01.08.2005)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ: 特願2004-225315 2004年8月2日 (02.08.2004) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社ソニー・コンピュータエンタテインメント (SONY COMPUTER ENTERTAINMENT INC.) [JP/JP]; 〒
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 和田 信也 (WADA, Shinya) [JP/JP]; 〒1070062 東京都港区南青山二丁目6番21号 株式会社ソニー・コンピュータエンタテインメント内 Tokyo (JP).
- (74) 代理人: 鈴木 正剛 (SUZUKI, Seigoh); 〒1050014 東京都港区芝三丁目22番7号 芝N Kビル4階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KM, KP, KR, KZ, LC, LK,

[続葉有]

(54) Title: NETWORK SYSTEM, MANAGEMENT COMPUTER, CLUSTER MANAGEMENT METHOD, AND COMPUTER PROGRAM

(54) 発明の名称: ネットワークシステム、管理コンピュータ、クラスタ管理方法およびコンピュータプログラム



106- MANAGEMENT COMPUTER  
104- COMPUTER NETWORK

(57) Abstract: There is provided a network system capable of effectively executing a process of jobs, the size, the type, and the number of which are expected to change every moment. A management computer (106) manages connection enabled/disabled information indicating whether a computer (108) under its control can be connected to another computer and easiness information indicating the easiness of the connection with the another computer. The management computer (106) searches for a computer whose latest connection enabled/disabled information managed by a local and another management computer indicates the connection enabled state and according to the easiness information on the computer specified, forms a cluster including the specified computer. The management computer (106) updates the connection enabled/disabled information on the computer (108) clustered, to the connection disabled state and upon completion of execution of the job, returns the connection enabled/disabled information on all the computers (108) belonging to the cluster to the state before formation of the cluster.

(57) 要約: 投入されるジョブのサイズ、種類、数が様々で、それらが時々刻々と変化していくことが予想される処理を効率的に実行するネットワークシステムを提供する。管理コンピュータ106は、自己の傘下のコンピュータ108が他のコンピュータとの接続が可能な状態にあるかどうかを表す接続可否情報と、他のコンピュータとの接続の容易度合いを表す容易性情報とを管理する。そして、自己及び他の管理コンピュータで管理している最新の接続可否情報が接続可を表すコンピュータを探し、これにより特定されたコンピュータの容易性情報に基づいて当該特定されたコンピュータを含むクラスタを形成する。クラスタ化されたコンピュータ108についての接続可否情報を接続不能に更新し、ジョブの実行が終了した時点でクラスタに属するすべてのコンピュータ108についての接続可否情報をクラスタ形成前の状態に

WO 2006/013992 A1

復帰させる。



LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR),

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 国際調査報告書
- 請求の範囲の補正の期限前の公開であり、補正書受領の際には再公開される。

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

## 明 細 書

ネットワークシステム、管理コンピュータ、クラスタ管理方法およびコンピュータプログラム

5

## 技術分野

本発明は、広帯域環境のコンピュータ・ネットワークに接続される複数のコンピュータによって行われる効率的な分散コンピューティング技術に関する。

より詳しくは、分散コンピューティングが可能なネットワークシステムおよび  
10 その構成要素、ならびにネットワークにおけるクラスタ管理方法に関する。

## 発明の背景

今日、ネットワークに接続された複数のコンピュータが協働して一つのジョブを分散処理することが一般的になっている。ジョブを分散処理する場合、従来は、  
15 どのコンピュータにどのジョブを割り当てるかを決めるために、予めネットワークに接続可能なすべてのコンピュータの処理能力を把握しているサーバの存在が不可欠となる。

サーバは、ジョブの負荷の大きさと、分散処理を行おうとするときにネットワークに接続されている各コンピュータの余剰処理能力（計算資源）とを特定し、  
20 負荷に応じた余剰処理能力を有するコンピュータを逐次割り当てていき、その割り当てたコンピュータからジョブの実行結果を受け取る。

サーバを必要とする従来の分散処理方法では、任意の時点でネットワークに接続されたり、非接続になったりするコンピュータの余剰処理能力を、サーバにおいて、迅速に把握することは、非常に困難である。また、サーバが、ジョブの分散処理を依頼したコンピュータからその実行結果を受け取って、ジョブの依頼元に転送しなければならないため、サーバのオーバーヘッドが大きくなる。そのため、ジョブの実行に必要な時間と、ネットワークを介するデータ伝送に必要な時間とが実質的に増加してしまうという問題がしばしば生じていた。

本発明は、このような従来の問題点を解決することができる分散処理の仕組み

を提供することを、主たる課題とする。

#### 発明の開示

本発明は、実行すべきジョブのサイズ、種類、数が様々で、それらが時々刻々と変化することが予想される情報処理を複数のコンピュータにより効率的に分散  
5 実行することができるネットワークシステムおよびその構成要素、ならびにクラスタ管理方法により、上記の課題を解決する。

ここにいう「コンピュータ」は、コンピュータプログラムによって動作するプロセッサを含む装置をいうが、装置の形式をとらないデバイス、プロセッサボード、  
10 プロセッサチップ、あるいはプロセッサ自体も「コンピュータ」の概念に含まれる。例えば、マルチプロセッサシステムのような個々のプロセッサも「コンピュータ」となり得る。

本発明が提供する第1構成のネットワークシステムは、それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムである。  
15

このネットワークシステムは、各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブルと、既に一又は複数のコンピュータにより形成されているクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2の  
20 テーブルとを含んでいる。

いずれかの前記コンピュータは、前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表す自己および他のコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報を、クラスタ化不能の状態を表す情報に更新し、さらに  
25 形成したクラスタについての前記追加容易性情報を前記第2テーブルに記録するクラスタ形成手段を有している。また、クラスタを形成した前記コンピュータは、当該クラスタへの追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段を有するものである。

## 3

このような構成のネットワークシステムでは、いずれかのコンピュータが、第1のテーブルに記録されている個々のコンピュータのクラスタ化可否情報をもとに自律的にクラスタを形成し、そのクラスタにおける追加容易性情報を第2のテーブルに記録する。クラスタに含めるコンピュータの数は、クラスタが形成されたタイミングで定まるデフォルト数であっても良く、実際に投入された、あるいは、近い将来投入が予想されるジョブのサイズ等に応じてフレキシブルに定まる数であっても良い。いずれにしても、クラスタ化可否情報がクラスタ化可能な状態であるコンピュータが選定されてクラスタが形成される。

クラスタが形成された後、追加候補となる候補コンピュータがあるときは、そのクラスタを形成したコンピュータが、第2テーブルに記録されている追加容易性情報に基づいてその候補コンピュータをクラスタに追加するかどうかを決定する。

追加容易性情報は、クラスタを形成したコンピュータが、候補コンピュータを追加することが適切かどうかを自律的に決定するための情報であって、様々な種類の情報を用いることができる。最先に追加を申し込んだいくつかのコンピュータのみを追加させる、乱数により発生した番号を識別情報とするコンピュータのみを追加させる・・・といった条件等も追加容易性情報として採用することができる。但し、追加させるかどうかの決定の処理を簡略化する観点からは、追加のし易さといった尺度を数値により定量化しておき、この数値との大小比較を行えるようにして、数値が高くなるほど候補コンピュータを当該クラスタへ追加させ易くするようにすることが望ましい。この場合の数値は、一例を挙げれば、不定期に発生する5回の追加可否の問い合わせに対して1回だけランダムに応答して追加を認める（この場合の数値は $1/5 : 20\%$ となる）、合計で1日のシステム運用時間の10%の時間帯の受付分だけ、候補コンピュータの追加を認める（この場合の数値は $24（時間） \times 60（分） \times 60（秒） \times 0.1 \times \alpha（確率値）$ となる）という類のものである。

なお、数値は、候補コンピュータの追加に拘わらず、一定値に保持させるようにしても良いし、第2のテーブルに記録した後に変動可能にしても良い。数値を変動可能にする場合は、候補コンピュータをクラスタに追加したことを契機に変

## 4

動させることにより、候補コンピュータが集まり易いクラスタは、さらに他の候補コンピュータを集めやすくなり、実際の需要に即したサイズのクラスタを容易に形成できるようになり易いという利点が生じる。

- このようにして、候補コンピュータがクラスタに逐次追加されることにより、
- 5 様々なサイズのクラスタがネットワークシステムの複数のコンピュータ上で形成され、成長していく。そのため、実行すべきジョブのサイズ、種類、数が様々で、それらが時々刻々と変化していくことが予想される情報処理を効率的に実行することができる分散コンピューティングが実現され、従来の問題点が解消される。

- 第1構成のネットワークシステムにおける、ある実施の態様では、クラスタを
- 10 形成したコンピュータが、当該クラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルに記録されている前記追加容易性情報および消滅したクラスタに所属していたすべてのコンピュータについて前記第1のテーブルに記録されていたクラスタ化可否情報をクラスタ形成前の状態に復帰させるクラスタ消滅手段をさらに有している。

- 15 クラスタは、ジョブが実行されるまでは、上述した追加容易性情報に基づいて成長するが、ジョブの実行が終了した時点で消滅するので、固定的にクラスタを用意しておく場合に比べて、コンピュータの有効活用が図れるようになる。

- 本発明が提供する第2構成のネットワークシステムは、それぞれ、他のコンピュータとのクラスタ化が可能な一又は複数のコンピュータを自己の傘下とする複数
- 20 数の管理コンピュータが自由に参加および離脱することができるネットワークシステムである。

「一又は複数のコンピュータを自己の傘下のコンピュータとして管理する」とは、自己に接続されている複数のコンピュータの動作を制御するとともに、それらの動作状態を監視することをいう。

- 25 このネットワークシステムにおいて、各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブルと、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録した第2のテーブルとを含んでいる。

また、少なくとも一つの前記管理コンピュータは、自己の傘下にあるコンピュータおよび他の管理コンピュータの傘下にあるコンピュータのうち前記第1のテーブルのクラスタ化可否情報がクラスタ化可能の状態を表すコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新するクラスタ形成手段を有している。さらに、クラスタを形成した前記管理コンピュータは、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段を有している。

10      このような構成のネットワークシステムでは、いずれかの管理コンピュータが、第1のテーブルに記録されているコンピュータのクラスタ化可否情報をもとに自律的にクラスタを形成し、そのクラスタにおける追加容易性情報を第2のテーブルに記録する。クラスタが形成された後、追加候補となる候補コンピュータがあるときは、そのクラスタを形成した管理コンピュータが、第2テーブルに記録されている追加容易性情報に基づいてその候補コンピュータをクラスタに追加するかどうかを決定する。クラスタ化可否情報、追加容易性情報の扱い、クラスタ化、クラスタ成長の際の基準は、第1構成のネットワークシステムと同じである。

15      第2構成のネットワークシステムにおける、ある実施の態様では、前記管理コンピュータは、前記形成したクラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルおよび消滅したクラスタに所属していたすべてのコンピュータについての前記第1のテーブルの記録情報をクラスタ形成前の状態に復帰させるクラスタ消滅手段をさらに有するものである。

25      クラスタは、ジョブが実行されるまでは、上述した追加容易性情報に基づいて成長するが、ジョブの実行が終了した時点で消滅するので、固定的にクラスタを用意しておく場合に比べて、コンピュータの有効活用が図れるようになる。

第2構成のネットワークシステムにおいて、第2のテーブルは、任意の箇所に存在しても良いが、通常は、第2のテーブルを生成することになるいずれかの管理コンピュータが保有することになる。この場合、その管理コンピュータは、最大で自己の傘下にあるコンピュータの数の前記第2のテーブルを保有することに

なる。

第2テーブルは、それを保有する管理コンピュータが自己の傘下にある第1コンピュータを含む第1クラスタを形成するときに当該第1コンピュータ用として生成するマスターテーブルと、それを保有する管理コンピュータが、他の管理コンピュータが形成した第2クラスタに追加される、自己の傘下にある第2コンピュータの動作を監視制御するときには当該第2コンピュータ用に生成するスレーブテーブルのいずれかを少なくとも含む。追加容易性情報はマスターテーブルに記録される。マスターテーブルを保有する管理コンピュータは、第1クラスタの形成、第1クラスタ内のコンピュータ数の変更並びに第1クラスタの消滅に関わる情報処理を主導的に行うマスター管理コンピュータとして振る舞い、スレーブテーブルを保有する管理コンピュータは、第2クラスタに対するスレーブ管理コンピュータとして振る舞う。

クラスタ化を促進する観点からは、マスター管理コンピュータは、いずれかの管理コンピュータに対してクラスタ化可能状態のコンピュータがあるかどうかを問い合わせることにより第1クラスタに追加させるための候補コンピュータを探索する探索手段を有するものとする。また、いずれかの管理コンピュータは、マスター管理コンピュータにより形成された第1クラスタについての前記追加容易性情報に基づいて自己の傘下の候補コンピュータを当該第1クラスタに追加させるかどうかを決定するように構成する。

本発明が提供する第3構成のネットワークシステムは、それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムであって；既に一又は複数のコンピュータにより形成されているクラスタ毎に、当該クラスタに所属するコンピュータが関係付けられている他のクラスタの識別情報をリストアップしたテーブルを含み；クラスタを形成した前記コンピュータは、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを自己のクラスタに関係付けられている他のクラスタに追加するかどうかを前記テーブルにリストアップされた識別情報に基づいて決定するクラスタ成長手段を有しているネットワークシステムである。「関係付けられている」とは、例えば互いに連絡が可能で連携処理等ができる状態を

いう。このような構成のネットワークシステムでは、追加容易性情報によらずともベキ分布に従うクラスタリングが可能になる。

本発明は、また、例えば上記第2構成のネットワークシステムを構成するための管理コンピュータを提供することにより、上記課題の解決を図る。

- 5 この管理コンピュータは、それぞれ、他のコンピュータとのクラスタ化が可能な一又は複数のコンピュータを自己の傘下とする管理コンピュータであって；他の同種機能を有する管理コンピュータと共に自由に参加および離脱することができるコンピュータネットワークに接続するためのネットワーク接続手段と；各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブル、および、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルへのアクセスを可能にするテーブル管理手段と；自己の傘下にある前記コンピュータおよび他の管理コンピュータの傘下にある前記コンピュータのうち前記第1のテーブルのクラスタ化可否情報
- 10 報がクラスタ化可能の状態を表すコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタについての前記追加容易性情報を前記第2のテーブルに記録するクラスタ形成手段と；追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段と；を有するものである。前記形成したクラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルに記録されている追加容易性情報および消滅したクラスタに所属していたすべてのコンピュータについて前記第1のテーブル
- 15 20 25 に記録されているクラスタ化可否情報をクラスタ形成前の状態に復帰させるクラスタ消滅手段をさらに有するものとすることもできる。

本発明は、また、例えば上記の第1構成および第3構成のネットワークシステムに含まれる複数のコンピュータにより実行されるクラスタ管理方法により、上記課題の解決を図る。

## 8

第1のクラスタ管理方法は、それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるクラスタ管理方法であって；各コンピュータが、それぞれ自己の状態がクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を第1のテーブルに記録する段階と；いずれかのコンピュータが、自己と前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表す他のコンピュータとを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を第2のテーブルに記録する段階と；追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定する段階と；を有するクラスタ管理方法である。

前記クラスタを形成したコンピュータが、前記形成したクラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルおよび消滅したクラスタに所属していたすべてのコンピュータについての前記第1のテーブルの記録情報をクラスタ形成前の状態に復帰させる段階とをさらに含めることもできる。

第2のクラスタ管理方法は、それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるクラスタ管理方法であって；既に一又は複数のコンピュータによりクラスタを形成したコンピュータが、当該クラスタに所属するコンピュータが関係付けられている他のクラスタの識別情報を所定のテーブルにリストアップする段階と；クラスタを形成した前記コンピュータが、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを当該クラスタと関係付けられている他のクラスタに追加するかどうかを前記テーブルに記録されている識別情報に基づいて決定する段階と；を有する、クラスタ管理方法である。

本発明は、また、コンピュータに所定の機能を付与するためのコンピュータプログラムを提供することにより、上記課題の解決を図る。

第1のコンピュータプログラムは、それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるいずれかのコンピュータに読み取られて実行されることにより当該コンピュータを、各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブル、および、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルへのアクセスを可能にするテーブル管理手段；前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表す自己および他のコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記第1のテーブルのクラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタについての前記追加容易性情報を前記第2のテーブルに記録するクラスタ形成手段；追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段として機能させるためのコンピュータプログラムである。

第2のコンピュータプログラムは、それぞれ他のコンピュータとのクラスタ化が可能な一又は複数のコンピュータを自己の傘下とする管理コンピュータに読み取られて実行されることにより、前記管理コンピュータを；他の同種機能を有する管理コンピュータと共に自由に参加および離脱することができるコンピュータネットワークに接続するためのネットワーク接続手段；各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブル、および、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルへのアクセスを可能にするテーブル管理手段；自己の傘下にある前記コンピュータおよび他の管理コンピュータの傘下にある前記コンピュータのうち前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表すコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ

タ化不能の状態を表す情報に更新し、さらに、当該クラスタについての前記追加容易性情報を前記第2のテーブルに記録するクラスタ形成手段；追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて  
5 決定するクラスタ成長手段；として機能させるためのコンピュータプログラムである。

第3のコンピュータプログラムは、それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるいずれかのコンピュータに読み取られて実行されること  
10 により、当該コンピュータを、既に一又は複数のコンピュータによりクラスタを形成したときに当該クラスタに所属するコンピュータが関係付けられている他のクラスタの識別情報を所定のテーブルにリストアップするクラスタ管理手段と、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを自己のクラスタに関係付けられている他のクラスタに追加するかどうかを前記テーブルにリストアップされた識別情報に基づいて決定するクラスタ成長手段；として  
15 機能させるためのコンピュータプログラムである。

これらのコンピュータプログラムは、可搬性の記録媒体に記録されて市場を流通し、あるいは、コンピュータまたは管理コンピュータがアクセス可能なコンピュータネットワークを通じてプログラムサーバ等からダウンロードされる。

本発明によれば、様々なサイズのクラスタがネットワークシステムの複数のコンピュータのいずれかにおいて形成され、それが所定のテーブルの記録情報に基づいて成長していくので、任意のサイズのクラスタを容易に得られるようになる。そのため、実行すべきジョブのサイズ、種類、数が様々で、それらのいずれかが時々刻々と変化していくことが予想される不確実性の情報処理であっても、それ  
25 を効率的に実行することができる分散コンピューティングが実現されるという、特有の効果が得られる。

#### 図面の簡単な説明

図1は、本発明が適用されるネットワークシステムの全体図である。

- 図2は、本実施形態による管理コンピュータの概要アーキテクチャを示す図である。
- 図3は、大規模情報処理統合体の説明図である。
- 図4は、本実施形態による管理コンピュータの機能構成図である。
- 5 図5は、リンクの数を横軸に、リンク数を持つノード数を縦軸にしてプロットした実測図であり、(a)は、ランダムに接続した場合の例、(b)は選択的接続した場合の例である。
- 図6(a)はマスターテーブル、図6(b)はスレーブテーブルの内容例を示した図である。
- 10 図7は、本実施形態による分散コンピューティングの概要手順図である。
- 図8は、管理コンピュータによるクラスタの形成、成長の手順説明図である。
- 図9は、管理コンピュータによるクラスタの形成、成長の手順説明図である。
- 図10は、管理コンピュータによるジョブの投入、実行の手順説明図である。
- 図11は、管理コンピュータによるクラスタ消滅時の処理手順説明図である。
- 15 図12(a)～(d)は、本実施形態によるクラスタの形成、成長、消滅の過程を示した説明図である。
- 図13(a)はマルチプロセッサにより実現されるネットワークシステムの概要図、図13(b)はマルチ・コア・プロセッサ単体により実現されるネットワークシステムの概要図である。
- 20 図14は、デュープリケーションモードで使用されるマスターテーブルの内容例を示した図である。
- 図15は、デュープリケーションモードでの運用形態例を示した図である。
- 図16(a)、(b)はデュープリケーションモードでの他の運用形態例を示した図である。
- 25 図17は、デュープリケーションモードでの他の運用形態例を示した図である。

発明を実施するための最良の形態

<ネットワークシステムのアーキテクチャ>

まず、本発明が適用されるネットワークシステムのアーキテクチャを説明する。

## 12

図1は、本発明が適用されるネットワークシステム101の全体図である。このネットワークシステム101は、コンピュータ・ネットワーク104を含んでいる。コンピュータ・ネットワーク104の例としては、ローカル・エリア・ネットワーク(LAN)、インターネットのようなグローバルネットワーク、あるいは他のコンピュータ・ネットワークが挙げられる。

コンピュータ・ネットワーク104には、一又は複数のコンピュータを自己の傘下とする管理コンピュータ106が、それぞれ任意のタイミングで接続することができる。各管理コンピュータ106は、それぞれ、他の管理コンピュータ106、および、他の管理コンピュータ106の傘下にある他のコンピュータ108との間で双方向の通信を行うことができる。各コンピュータ108は、何時でも、いずれかの管理コンピュータ106の傘下になることができる。つまり、これらのコンピュータ106、108は、ネットワークシステム101に、何時でも自由に離脱できる形態で参加できるようになっている。

この実施形態にいう「コンピュータ」は、コンピュータプログラムによって動作するプロセッサを含む装置をいうが、装置の形式をとらないデバイス、プロセッサボード、あるいはこれらの集合も「コンピュータ」の概念に含まれる。また、「一又は複数のコンピュータ108を傘下とする」とは、自己に接続されている複数のコンピュータ108の動作を制御するとともに、それらの動作状態を監視することをいう。

管理コンピュータ106の例としては、サーバ機能を有するコンピュータ、通信機能付のゲームコンソール、コンピューティングデバイス、プロセッサボードなどが含まれる。

管理コンピュータ106の傘下にあるコンピュータ108の例としては、パーソナルコンピュータ、通信機能付のゲームコンソール及びその他の有線または無線コンピュータ、コンピューティング・デバイス、プロセッサボードなどが含まれる。

管理コンピュータ106へのコンピュータ108の接続形態は様々である。例えば、図1に示されるように、コンピュータ・ネットワーク104に接続された管理コンピュータ106を中心に複数のコンピュータ108が直接接続されるス

## 13

ター接続型、管理コンピュータ106が接続されている構内ネットワークを介して複数のコンピュータ108が接続される構内ネットワーク型、コンピュータ・ネットワーク104を介して複数のコンピュータ108が管理コンピュータ106に接続される広域ネットワーク型等が挙げられる。

- 5 構内ネットワーク型における構内ネットワークの種類も様々であり、コンピュータ・ネットワーク104とダイレクトに接続されたネットワークもあれば、コンピュータ・ネットワーク104とは管理コンピュータ106を介して接続されるネットワークもある。

各管理コンピュータ106は、図2に示されるように、バスB11を通じて相互に接続されたハードディスク等のストレージ装置1061、通信装置1063、  
10 半導体メモリ1065、プロセッサ1067を含んでいる。このようなハードウェア資源を有する管理コンピュータ106は、CD-ROM等の記録媒体を通じて半導体メモリ1065にロードされたコンピュータプログラムをプロセッサ1067が読み取って実行することにより、図4に示されるように、ストレージ装置1061に、後述する複数の管理テーブル206を格納するとともに、  
15 テーブル管理部212、クラスタ管理部216、通信制御部236、およびジョブ実行部256の機能をコンピュータ本体内に構築する。

テーブル管理部212は、管理テーブル206にアクセスして、その管理テーブル206の記録内容を更新させる。クラスタ管理部216は、自己の傘下のコンピュータ108および他の管理コンピュータ106の傘下にあるコンピュータ  
20 108が所属するクラスタの形成・成長・消滅に関わる情報処理を行う。つまり、クラスタ管理部216は、クラスタの状態変化に応じて、クラスタ形成手段、クラスタ成長手段、クラスタ消滅手段として機能する。

通信制御部236は、通信装置1063を通じて、自己の傘下のコンピュータ  
25 108および他の管理コンピュータ106との間の通信を可能にする。ジョブ実行部256は、自己の傘下のコンピュータ108を含むクラスタにジョブを投入し、それを実行させる。ジョブ実行部256には、最適なクラスタのサイズ、つまりコンピュータ数の特定を容易にするために、ジョブのサイズ、処理時間等を計算する機能を持たせるようにしても良い。

コンピュータ108は、管理コンピュータ106と同様のストレージ装置、通信装置、半導体メモリおよびプロセッサを有している。この実施形態では、ストレージ装置に、後述するステータステーブルを格納する。また、CD-ROM等の記録媒体を通じて半導体メモリにロードされたコンピュータプログラムをプロセッサが読み取って実行することにより、クラスタ管理およびジョブの実行に関する種々の機能をコンピュータ本体内に構築する。望ましくは、他のコンピュータ108と命令セット・アーキテクチャ(ISA)がすべて同じ、あるいは同じとみなされるもので、同じ命令セットに従って所要の処理を実行できるものとする。

各管理コンピュータ106における傘下のコンピュータ108の数は、任意である。ある管理コンピュータ106では、各種アプリケーションによって与えられるジョブを実行する上で必要な処理能力によって、コンピュータ108の数が割り当てられている。

管理コンピュータ106の傘下にある複数のコンピュータ108におけるISAがそれぞれ同一あるいは同一とみなされる場合には、管理コンピュータ106あるいはネットワークシステム101におけるアダプタビリティを飛躍的に改善することができる。

各管理コンピュータ106は、各々自己の傘下のコンピュータ108又は他の管理コンピュータ106の傘下のコンピュータ108のうち1つまたはそれ以上を含んでクラスタを形成し、クラスタ単位でジョブを実行できるようにする。自己の傘下のコンピュータ108と他の管理コンピュータ106の傘下のコンピュータ108との間には、少なくとも一つのクラスタを形成する上での性能上の相違、制約等は存在しない。このように扱うことにより、ジョブは、どの管理コンピュータ106のどのコンピュータ108において実行するかは、さほど重要でなくなる。ジョブの実行結果の受け取り先を、ジョブを依頼した管理コンピュータ106、その傘下のコンピュータ108あるいは後続のジョブを実行する任意のコンピュータ108として指定するだけで足りるようになる。そのため、個々のジョブは、コンピュータ・ネットワーク104に接続されている複数の管理コンピュータ106の傘下のコンピュータ108の間で分散実行することが容易になる。

上記のように各コンピュータ108が、すべて同じ、あるいは同じとみなされるISAを有するものとする場合は、コンピュータ108間の互換性を達成するためのソフトウェアの追加層の計算上の負担も回避されるし、異質なコンピュータ・ネットワークの混在という問題の多くを防ぐことができる。そのため、用途  
5 に応じて、そのようなネットワークシステム101を構成するのが望ましい。このようなネットワークシステム101では、コンピュータ・ネットワーク104に接続されている複数の管理コンピュータ106およびコンピュータ108間のアーキテクチャの相違は吸収され、図3に示されるように、各管理コンピュータ106の傘下にある多数のコンピュータ108の各々があたかも情報処理の細胞  
10 (Cell) のように機能する、広帯域処理の実現が可能な大規模情報処理統合体W0が形成される。

大規模情報処理統合体W0における個々のコンピュータ108は、物理的には自己が所属する管理コンピュータ106によって管理され、単独のコンピュータ  
15 108と共にクラスタ化されて協働で動作したりする。しかし、論理的には、管理コンピュータ106による壁はなく、異なる管理コンピュータ106の傘下の他のコンピュータ108との間でのクラスタ化も可能である。このような形態でクラスタ化される場合、一つのジョブを同じクラスタに属する複数のコンピュータ108で分散実行することができる。

#### 20 <分散コンピューティング>

本発明は、上記のネットワークシステム101のアーキテクチャを利用した効率的な分散コンピューティングの仕組みを提供する。このような仕組みを可能にするため、本発明では、例えば図3に示した大規模情報処理統合体W0のような  
25 ネットワークを構成する複数のコンピュータ108の各々がノードとなり、少数のノードからスタートして時間の経過と共にノード数が増加する「成長」、ノード間のリンクの接続手法としての「選択的接続」とを繰り返すことができるようにする。

「選択的接続」とは、新たにネットワークに参加したノードがリンクを張る際に、リンク先のノードを選ぶ基準として、より多くのリンクを既に持っているノ

ードに、より接続しやすくなるいうものである。成長と選択的接続の2つの特徴により、ネットワークは、成長しながらスケールフリーの性質を持つようになる。このような成長と選択的接続の繰り返しにより、ノードの持つリンク数の分布がベキ分布になることは、例えば、Albert Barabasi, Reka Albert, Hawoong “Mean-  
5 field theory for scale-free random networks” に詳細に示されている。

図5は、ノードをランダムに接続した場合と、選択的接続した場合のノードが持つリンクの数を横軸に、そのリンク数を持つノード数を縦軸にして、実測値をプロットしたグラフである。図5(a)は、ノードをランダムに接続した場合の例である。この例の場合、典型的なリンク数をもつ代表的なノードというものが現れる。図5(b)は、選択的接続した場合の例であり、代表的なノードが現れず、リンク数は、広い範囲に拡がる。つまり、典型的なスケールが現れないという意味で、スケールフリーと称される。図5(b)の分布は、以下のように表され、両対数のグラフで書くと、直線になるような非常に裾野の広い分布となる。

$$P(k) = AK^{-r} \cdots (1)$$

15 但し、k:リンク数、P(k):ノード数

本発明は、上記の2つの特徴を応用し、個々のノードをクラスタと解釈し、リンクは、互いに結ばれたノードに相当するクラスタに含まれる2つのコンピュータ108であると解釈する。これにより、任意の分布、もちろん、図5(b)に示したベキ分布をもつことも可能な形態でクラスタリングを行い、これにより、  
20 効率的な分散コンピューティングを可能にするものである。このようにして成長していくネットワークでは、リンク数は、クラスタに含まれるコンピュータ108の数、つまりクラスタサイズ、あるいはクラスタの計算能力となる。横軸にクラスタに含まれるコンピュータ108の数、縦軸にクラスタの数又はクラスタ化の頻度をプロットすると、それは、図5(b)に示したベキ分布と類似した分布  
25 になるはずである。

以下、上記のようなクラスタリングを可能にするには、個々のコンピュータ108がどのように動作し、各管理コンピュータ106がどのようにコンピュータ108又は他の同種の管理コンピュータ106の動作を監視制御していったら良  
いかを述べる。

個々のコンピュータ108は、現在、自己の状態がクラスタ化が可能な状態にあるかどうかを表すコンピュータ・クラスタ・ステータス（クラスタ化可否情報）により管理される。コンピュータ・クラスタ・ステータスは、ステータステーブルに記録され、コンピュータ108の状態の変化に追隨して、当該コンピュータ  
5 108又はそれを傘下として管理する管理コンピュータ106により更新される。ステータステーブルは、コンピュータ108又は管理コンピュータ106がアクセス可能な任意のメモリ領域に存在すれば足りるが、コンピュータ108のストレージ装置に設けるようにしても良い。

ステータステーブルに記録されるコンピュータ・クラスタ・ステータスの例として、本実施形態では、「clustered」、「run」、「free」の3種類の情報を用いる。  
10

「clustered」は既にクラスタ化されており、現在はジョブの実行を待っているコンピュータであることを示す。「run」はジョブを実行中のコンピュータであることを示す。「free」は未稼働状態にあるコンピュータであることを示す。

「clustered」、「run」のときは、そのコンピュータ108はクラスタ化ができない。そのため、「clustered」、「run」は、クラスタ化不能の状態を表す情報となる。  
15 他方、「free」はクラスタ化可能の状態を表す情報となる。

管理コンピュータ106の役割の一つは、上述したように、自己の傘下のコンピュータ108を管理することにある。管理テーブル206は、そのために用いられる。上述したように、管理テーブル206は、例えば当該管理コンピュータ  
20 106が有するストレージ装置1061に格納される。この管理コンピュータ106は、最大で、自己の傘下にあるコンピュータ108の数だけの管理テーブル206を有する。

管理テーブル206には、2種類ある。ある種類に分類される管理テーブル206は、管理コンピュータ106が、初めに自己の傘下のコンピュータ108を含むクラスタを形成する際にそのクラスタの状態に関する情報を記録するための  
25 「マスタテーブル」である。もう一つは、他の管理コンピュータ106により形成されるクラスタに追加されるコンピュータ108があるときに、そのクラスタの状態に関する情報を記録するための「スレーブテーブル」である。

管理コンピュータ106は、マスタテーブルにより自己が管理するクラスタ

に含まれるコンピュータ108を監視制御するときは「マスター管理コンピュータ」として振る舞う。他方、スレーブテーブルにより他の管理コンピュータ106が管理するクラスタに属する自己の傘下のコンピュータ108を監視制御するときは、「スレーブ管理コンピュータ」として振る舞う。つまり、1つの管理コンピュータ106でありながら、最大で、自己の傘下のコンピュータ108の数だけの種類の管理コンピュータとして振る舞うことになる。

マスターテーブルは、マスター管理コンピュータとして振る舞う管理コンピュータ106が生成し、スレーブテーブルは、スレーブ管理コンピュータとして振る舞う管理コンピュータ106が生成する。スレーブテーブルが存在するという  
10 ことは、それに対応するマスターテーブルがどこかの管理コンピュータ106に存在していることを意味する。

ここで、本実施形態におけるマスターテーブルとスレーブテーブルの内容例を具体的に説明する。図6(a)にはマスターテーブル2161、図6(b)にはスレーブテーブル2162の内容例が示されている。

15 マスターテーブル2161には、クラスタID(IDは識別情報の意、以下同じ)2361、クラスタ・サイズ2362、コンピュータリスト2363、クラスタ・ステータス2364、最大コンピュータ数2365、最小コンピュータ数2366、トータルコンピュータ数2367、クラスタ・コネクションレシオ2368の値(図6(a)の右側のフィールドの記録値)が、それぞれ、そのクラ  
20 スタの成分(値)として、マスター管理コンピュータにより記録される。

クラスタID2361は、形成したクラスタに付与される固有IDである。そのクラスタが存在し続ける限りにおいて一意の情報になれば良いので、例えば、マスター管理コンピュータのIDと初めにクラスタに含まれることになるコンピュータ108のIDとの組をクラスタIDとすることができる。

25 クラスタ・サイズ2362は、新規あるいは追加でクラスタに含まれることとなったコンピュータ108の総数値である。これは、マスター管理コンピュータが、クラスタ内のコンピュータ数を計測することにより特定することができる。

コンピュータリスト2363は、クラスタに含まれることになったコンピュータ108の識別情報のリストである。

クラスタ・ステータス 2364 は、クラスタの現在の状態を表す情報である。本実施形態では、「idle」、「run」、「wait」の3種類の状態をクラスタ・ステータスとして用いる。「idle」はクラスタがジョブを実行していない状態である。クラスタがこの状態のときは、そのクラスタへのコンピュータ 108 の追加が可能である。「run」はそのクラスタがジョブを実行している状態である。「wait」は保持されているコンピュータ 108 の数が最大コンピュータ数に達している状態である。この状態では、そのクラスタへのコンピュータ 108 の追加はできない。

最大コンピュータ数 2365 は、クラスタが保持可能な最大コンピュータ数であり、ユーザによって定義されるか、あるいはシステムの保有する定数として定義される。図示の例では、最大で 200 個のコンピュータを保有できるクラスタであることが示されている。

最小コンピュータ数 2366 は、クラスタとして保持しなければならない最小コンピュータ数であり、ユーザによって定義されるか、ネットワークシステム 101 の保有する定数として定義される。図示の例では、3 個のコンピュータ 108 によりクラスタとなることが示されている。

トータルコンピュータ数 2367 は、そのクラスタに含まれることになる可能性があるコンピュータ 108 の総数値であり、統計に基づく期待値により求めることができる。図示の例では、48 個のコンピュータ 108 が含まれる可能性のあることが示されている。

コネクションレシオ 2368 は、追加容易性情報の一例となる情報である。本実施形態では、例えば追加候補となるコンピュータ 108（候補コンピュータ）が存在するときにそれを 20% の確率でクラスタに追加させる、あるいは、必要とするクラスタ・サイズ／トータルコンピュータ数のように、追加のし易さを確率値等により定めた数値をコネクションレシオとして用いる。この数値が高いほど、クラスタにはコンピュータ 108 が追加され易くなり、クラスタが成長しやすくなる。図示の例では、現在のクラスタサイズを 23、トータルコンピュータを 48 個としたときのコネクションレシオ (= 0.48) が示されている。

コネクションレシオ 2368 は、一定値を用いることもできるし、変動値とすることもできる。一定値にしていたものを事後的に変動値に代えることもできる

## 20

し、その逆にすることもできる。これは、クラスタを形成するときのデフォルト値として自動的に設定されるようにしても良いし、クラスタにジョブを実行させるアプリケーションプログラムがパラメータの一つとしてマスターテーブル2161のコネクションレシオ2368に設定するようにしても良い。

- 5 図5のように横軸にクラスタに含まれるコンピュータ108の数、縦軸にクラスタの数又はクラスタ化の頻度をプロットしたときの分布がどのようなになるかは、このコネクションレシオをどのようにするかに因るところが大きい。例えば、図5(a)のような正規分布を得ようとするときは、コネクションレシオを一定値とする。これにより、上記の分布は、代表値を中心とする正規分布となる。コンピュータが追加されてクラスタサイズが大きくなるにつれてコネクションレシオが変動するようにしても良い。例えば、1又はいくつかのコンピュータが追加される毎に、コネクションレシオが高くなるようにしても良い。このようにすれば、
- 10 図5(b)のようなベキ分布を容易に得ることができる。

- 15 なお、マスターテーブル2161における上記の成分の持ち方は例示であり、適宜、増減することができる。

スレーブテーブル2162には、クラスタID2461と、コンピュータid2462とが記録される。クラスタID2461は、クラスタを識別するためのID(マスターテーブル2161に記録されているクラスタID2361と同じ)である。コンピュータid2462はクラスタに含まれるコンピュータ108を

20 識別するためのIDである。

スレーブテーブル2162は、クラスタID2461とコンピュータid2462さえ判明していれば、そのクラスタID2461を同一にするマスターテーブル2161とリンクしているので、それに記録されている成分(値)は、マスターテーブル2161よりも簡易なものとなっている。

- 25 なお、スレーブテーブル2162における上記の成分(値)の持ち方も例示であり、適宜、増減することができる。

#### <運用形態>

次に、本実施形態のネットワークシステム101による分散コンピューティングの運用形態の一例を説明する。

## 21

本実施形態による分散コンピューティングでは、図7に示すように、3段階のクラスタの状態変化のサイクルを繰り返しながら、ジョブを分散実行していく。第1段階の状態変化は、新しいクラスタの形成、成長により生じる（ステップS1）。第2段階の状態変化は、アプリケーションプログラム等から依頼されたジョブの投入により生じる（ステップS2）。そして、第3段階の状態変化は、ジョブ実行後のクラスタの消滅により生じる（ステップS3）。クラスタが消滅すると、それまでそのクラスタに属していたすべてのコンピュータ108は、その時点で他のクラスタへ新たに属することができる候補コンピュータ（ノード）となる。

このサイクルにおける第1段階の状態変化は、ジョブの有無とは無関係に生じる。そして、ある管理コンピュータ106がマスター管理コンピュータとなってクラスタを形成し、それをスレーブ管理コンピュータとの協働により成長させていく。クラスタがどの程度成長するかは、個々のコンピュータ108の稼働状態とコネクションレシオによって決まる。コンピュータ108の稼働状態が「free」でコネクションレシオも高いほど、他のコンピュータ108と接続し易くなる。以下、上述の3段階の状態変化の様子を詳細に説明する。

#### <第1段階の状態変化>

##### [新しいクラスタの形成、成長1]

図7は、マスター管理コンピュータ（そのように動作する管理コンピュータ106のクラスタ管理部216）によるクラスタの形成、成長の過程を示す手順説明図である。

マスター管理コンピュータは、自己の傘下のコンピュータ108の稼働状態を調べる。具体的には、各コンピュータ108が保有するステータステーブルに記録されているコンピュータ・クラスタ・ステータスを調べる（ステップS101）。そして、未稼働状態のコンピュータ108、すなわちコンピュータ・クラスタ・ステータスが「free」となっているいくつかのコンピュータ108により1つのクラスタを形成する（ステップS102）。最初に形成するクラスタにいくつかのコンピュータを含まるかは、例えば事前の設定により決める。通常は、自己の傘下のコンピュータ108のうち、いくつかのコンピュータ108を残して、クラスタ化する。残ったコンピュータ108は、他のクラスタを成長させ易くするため

に、なるべく他のクラスタに追加されるような状態にしておく。この状態を図12(a)に示す。

図12(a)の例では、3つのコンピュータ108により1つのクラスタC11が形成されている。このとき、マスター管理コンピュータは、1つのマスターテーブルを生成し、このマスターテーブルの成分(値)をセットする(ステップS104)。マスター管理コンピュータが各コンピュータ108のステータステーブルを管理することになっている場合は、このクラスタに含まれるすべてのコンピュータ108のコンピュータ・クラスタ・ステータスを「free」から「clustered」に更新する(ステップS104)。

- 10 その後、マスター管理コンピュータは、ランダムに選定した他の管理コンピュータ106の傘下のコンピュータ108のうち、自己が形成したクラスタに追加するための候補コンピュータが無いかどうか、すなわちコンピュータ・クラスタ・ステータスが「free」になっている他のコンピュータ108が無いかどうかを問い合わせることにより探索する(ステップS105)。探索は、例えば、自己を起
- 15 点として論理的な距離の短いいくつかの他の管理コンピュータ106をランダムに選定し、逐次、その範囲を拡げていくことにより行う。候補コンピュータが存在する場合は(ステップS106: Yes)、マスターテーブルのコネクションレシオを調べ、コネクションレシオに基づいて、その候補コンピュータをクラスタに追加し、当該候補コンピュータに関する情報をマスターテーブルに記録する(ス
- 20 テップS107)。その後、追加した候補コンピュータのコンピュータ・クラスタ・ステータスを「free」から「clustered」に更新する(ステップS108)。

- 追加された候補コンピュータを管理する管理コンピュータ106は、その候補コンピュータ108についてのスレーブ管理コンピュータとなり、スレーブテーブルを生成する。候補コンピュータが存在したときの状態を図12(b)に示す。
- 25 図12(b)の例では、2つのコンピュータ108が候補コンピュータとして追加されて、5つのコンピュータ108が含まれることになったクラスタC14に成長している。マスター管理コンピュータは、追加された新たなコンピュータ108の情報をマスターテーブルに追記する。

なお、ステップS106において、候補コンピュータが存在しない場合は(ス

テップS106:No)、ステップS105の処理に戻る。

[新しいクラスタの形成、成長2]

図9は、マスター管理コンピュータとならない管理コンピュータ106によるクラスタの成長過程を示す手順説明図である。

- 5 クラスタの成長は、図8に示したように、マスター管理コンピュータが能動的に振る舞うことにより実現されるのが原則であるが、マスター管理コンピュータとはならない管理コンピュータ106の方からマスター管理コンピュータにアクセスしてそれを行うこともできる。

- すなわち、クラスタ化されていないコンピュータ108を傘下に持つ管理コンピュータ106は、例えばランダムに選定した他の管理コンピュータ106に問い合わせることにより、そのコンピュータ108を追加することができる他のクラスタを探索する(ステップS201)。そして、入り込む先のクラスタの接続レシオに基づいてそのクラスタに自己の傘下のコンピュータ108を追加するかどうかを決定する(ステップS202)。追加しないことになった場合は、
- 15 ステップS201に戻る(ステップS203:No)。

- 追加する場合は、そのコンピュータ108について、接続先のクラスタに対するスレーブ管理コンピュータとなるので、そのコンピュータ108についてのスレーブテーブルを生成し、このスレーブテーブルに各成分(値)をセットする(ステップS204)。そして、追加するコンピュータ108のコンピュータ・クラスタ・ステータスを「free」から「clustered」に更新する(ステップS205)。
- 20 さらに、追加される側のクラスタのマスター管理コンピュータに、マスターテーブルをアップデートさせる(ステップS206)。なお、このアップデートは、スレーブ管理コンピュータからマスター管理コンピュータにアクセスして自ら行うようにしても良い。

- 25 <第2段階の状態変化>

[ジョブの投入]

第2段階の状態変化は、アプリケーション等からの依頼によりジョブが投入されたいずれかの管理コンピュータ106によるアクションを起点に始まる。ジョブの実行に際しては、ジョブの実行に必要となる計算量と所要時間、例えばコン

## 24

ピュータ108の数がパラメータとなる。このパラメータは、投入されたジョブのサイズに基づいて当該管理コンピュータ106のプロセッサが自動的に生成しても良いし、アプリケーション等を利用するユーザが適宜与えるようにしても良い。アプリケーションが持っているパラメータを使用するようにしても良い。

- 5 図9は、ジョブが投入された管理コンピュータ106によるクラスタ利用の手順説明図である。管理コンピュータ106は、上記の計算量に基づいて、クラスタリスト208を参照することにより、あるいは他の管理コンピュータに問い合わせることにより、必要なサイズのクラスタを形成しているマスター管理コンピュータを探索する(ステップS301)。見つかった場合は、そのマスター管理コンピュータに、ジョブの実行を依頼する(ステップS302: Yes、S303)。
- 10 ジョブの実行依頼は、例えば、ジョブと、その実行に必要なプログラムおよびデータと、実行結果の送出手続きを自己宛とするための指定アドレスとを含むパケットをそのマスター管理コンピュータに送出手続きすることにより行う。

- マスター管理コンピュータは、そのジョブを実行するクラスタに所属している各コンピュータ108のコンピュータ・クラスタ・ステータスを「idle」又は「wait」
- 15 から「run」に更新したうえで、各コンピュータ108による分散処理によってジョブの実行を始める。図12(c)は、図12(b)のように成長したクラスタC14を管理するマスター管理コンピュータが、ジョブを実行している状態を表している。ジョブの実行結果は、マスター管理コンピュータから依頼元の管理コンピュータ106に送出手続きされる。実行結果を受領した管理コンピュータ106は、その実行結果をジョブの投入元に伝達する(ステップS304: Yes、S305)。
- 20

- なお、ジョブが、必要なサイズのクラスタを管理しているマスター管理コンピュータに直接投入された場合は、上記ステップS301~S304は、自ら行う処理手順となる。また、ステップS302において、複数のクラスタに跨ってジョブを実行する必要があると判定した場合、例えば、画像処理と音声処理とを別々のクラスタで実行する必要があると判定した場合は、各クラスタを管理するマスター管理コンピュータに、それぞれジョブの実行を依頼することになる。
- 25

<第3段階の状態変化>

[クラスタの消滅]

## 25

クラスタの消滅は、ジョブの実行を行ったクラスタを管理するマスター管理コンピュータが行う。図11は、このクラスタを消滅させるときのマスター管理コンピュータにおける処理手順図である。

5 マスター管理コンピュータは、ジョブが終了すると（ステップS401:Yes）、それまでジョブを実行したクラスタに含まれるすべてのコンピュータ108についてのコンピュータ・クラスタ・ステータスを「free」に更新する（ステップSS402）。また、そのクラスタのマスターテーブルの成分（値）をクリアする。同時に、クラスタに所属していたコンピュータを傘下にするスレーブ管理コンピュータを通じてスレーブテーブルの成分（値）をクリアする。すなわち、すべての  
10 コンピュータをクラスタ形成前の状態に戻す（ステップS403）。これにより、そのクラスタに所属していたすべてのコンピュータ108は、未稼働状態で、クラスタ化が可能な状態となり、あるコンピュータ108は、直ちにクラスタ化され、他のコンピュータ108は、他のクラスタに追加される状態となる。図12（d）はこの状態を示している。

15 このように、この実施形態によれば、投入されるジョブのサイズ、種類、数のいずれかが様々で、それらが時々刻々と変化していくことが予想される不確定要素の多いネットワークシステムであっても、複数存在する管理コンピュータ106の各々が、その傘下のコンピュータ108毎に、それぞれ、マスター管理コンピュータあるいはスレーブ管理コンピュータとして振る舞いながら、最適なサイズのクラスタを形成し、それを成長させていくので、適切なサイズのクラスタに  
20 ジョブの実行を割り当てることができ、効果的にジョブを実行していくことができる。また、大規模なプロセスから非常に小さなプロセスまで満遍なく一様の手順で対応が可能になるので、柔軟性に富む分散コンピューティングを実現することができる。特に、プロセスのサイズの分布がべき分布になるときには、追加容  
25 易性情報を変動可能にし、クラスタサイズもべき分布になるように設定することができるので、最も効果的にプロセスを処理することができ、コンピュータ資源の使用効率を高めることができる。

<変形例>

以上の例では、便宜上、ステータステーブルは、各コンピュータ108が保有

し、管理テーブル206（マスターテーブル／スレーブテーブル）は、管理コンピュータ106が保有することを前提として説明したが、これらのテーブルは、各コンピュータ108および管理コンピュータ106がそれらにアクセスできるメモリ領域に存在すれば良い。例えば、管理コンピュータ106が、自己の傘下のコンピュータ108のステータステーブルを備えていても良いし、管理コンピュータ106のテーブル管理部212がアクセス可能なサーバ等に管理テーブル206を一括して格納するようにしても良い。

また、以上の例では、スレーブ管理コンピュータの傘下にある候補コンピュータを他のクラスタに追加するかどうかは、そのスレーブ管理コンピュータが決定するようにしているが、候補コンピュータをクラスタに追加するかどうかは、すべてマスター管理コンピュータにおいて決定するようにしても良い。

#### <管理コンピュータによらない場合の実施例>

以上の説明では、複数のコンピュータ108を傘下にする管理コンピュータ106の集合から成るネットワークシステム101の例を示したが、本発明は、管理コンピュータ106によらずに、複数のコンピュータ108の各々が直接参加および離脱することができるネットワークシステムとしての実施も可能である。

このようなネットワークシステムとして実施する場合は、各コンピュータ108のステータステーブルに、コネクションレシオをも記録するようにし、ステータステーブルの管理も自ら行うようにする。例えば、他のコンピュータ108により形成されているクラスタに、自己を追加するときには、自己のコンピュータ・クラスタ・ステータスを「clustered」に更新させる。そして、そのクラスタによるジョブの実行が終了した時点で、自己のコンピュータ・クラスタ・ステータスを「free」に復帰させるようにする。また、上述した管理コンピュータ106と同様の機能を、いずれかのコンピュータ108に持たせるようにする。つまり、自らが主導的にクラスタを形成するときは、そのクラスタの内容を記録したマスターテーブルを生成する。

クラスタを形成したコンピュータ108は、他のコンピュータが保有するステータステーブルの最新のコンピュータ・クラスタ・ステータスが「free」になっている他の追加候補となる候補コンピュータを探索し、これにより特定された候

補コンピュータを、クラスタのコネクションレシオに基づいて追加するかどうかを決定するようにする。クラスタを形成したコンピュータ108は、このようにして成長したクラスタにジョブを実行させ、実行終了後は、そのクラスタを消滅させる。

5 <マルチプロセッサ/マルチ・コア・プロセッサとしての実施例>

本発明は、複数のプロセッサ又はマルチ・コア・プロセッサによるネットワークシステムとして実施することもできる。この場合は、図13(a)に例示したように、内部バスを上述したコンピュータネットワーク104として機能させ、この内部バスに接続されたいずれかのプロセッサ又はマルチ・コア・プロセッサ  
10 のいずれか一つ又はいくつかを上述した管理コンピュータ106として動作させるとともに、残りのプロセッサ等を上述したコンピュータ108として動作させるようにすれば良い。

図13(b)に例示した個々のマルチ・コア・プロセッサ自体をネットワークシステムとして実施することもできる。この場合は、内部バスに接続されたい  
15 れか一つ又はいくつかのコア(プロセッサ・コア)および入出力制御部(I/O)とキャッシュメモリとで上述した管理コンピュータ106の動作を実現するとともに、残りのコアを上述したコンピュータ108として動作させるようにすれば良い。

20 他の実施の形態

本発明のネットワークシステムは、あるクラスタのコンピュータ108が他のクラスタにアクセスしたときに、当該他のクラスタと関係付けられている別のクラスタにコンピュータ108を追加させるという実施形態も可能である。このよ  
25 うな実施形態のモードを「デュープリケーションモード(Duplication Mode)」と称する。この「デュープリケーションモードについては、下記の文献に詳細に記載されている。

タイトル : Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations

著作者 : Alexei V. Vazquez

雑誌名 : Phys Rev E Stat Nonlin Soft Matter Phys. 2003 May; 67(5 Pt2) :056104.  
Epub 2003 May 7.

デュープリケーションモードでは、上述したコネクションレシオ2368に代  
えて、アクセスしたクラスタに所属するコンピュータが関係付けられている他の  
5 クラスタの識別情報（接続アドレス等）をリストアップしたクラスタリストを用  
いる。クラスタリストは、例えばマスターテーブルに設けられる。

図14は、この場合のマスターテーブルの内容例を示している。このマスター  
テーブル2163は、図6(a)に示したマスターテーブル2161と同じ内容  
のクラスタID2361、クラスタサイズ1262、コンピュータリスト236  
10 3、クラスタ・ステータス2364、最大コンピュータ数2365、最小コンピ  
ュータ数2366のほか、クラスタリスト2369を有する。マスター管理コン  
ピュータは、自己が管理するクラスタに新たなクラスタが関係付けられる度に、  
その新たなクラスタの識別情報をクラスタリスト2369にリストアップする。  
この実施形態にいう「関係付け」は、あるクラスタに所属するコンピュータと他  
15 のクラスタに所属するコンピュータとが論理的につながって（連絡しあって）、連  
携処理等ができることをいう。

マスター管理コンピュータは、例えばスレーブ管理コンピュータからの求めに  
応じてクラスタリスト2369にリストアップされた一つのクラスタの識別情報  
を読み出し、それを当該スレーブ管理コンピュータへ通知する。セキュリティを  
20 重視しないで足りる用途では、このクラスタリスト2369を例えばスレーブ管  
理コンピュータに閲覧させるようにしても良い。

このクラスタリスト2369からの読み出し順序は、ランダムであっても良く、  
所定順であっても良い。前者の場合、クラスタリスト2369へのクラスタの識  
別情報のリストアップの順位は、任意であって良い。他方、後者の場合は、リス  
25 トアップの順位に重み付けを行う。例えば、そのマスター管理コンピュータが管  
理するクラスタの処理内容との関連性が高い他のクラスタほど先に読み出される  
ようにする。この場合の関連性は、例えば、クラスタ間のアクセス回数、クラ  
スタ間の所定時間内の交信回数、クラスタ間の分散処理の関連度合い（画像処理と  
同期をとるべき音声処理等）等により決めることができる。

## 29

デュープリケーションモードでの運用形態は、例えば以下ようになる。

いま、ネットワークシステムが、図15の上段のような初期状態であったとする。図15の例では、複数のクラスタC6～C11が既に形成されている。各クラスタC6～C11には、それぞれ1又は複数のコンピュータ（上述したコンピュータ108、以下同じ）が所属している。

この状態で、図15中段に示されるように、ネットワークシステムに新たにクラスタC20が形成されたとする。このクラスタC20を形成した管理コンピュータは、当該クラスタC20に所属するコンピュータのほか、他のクラスタを成長させ易くするためのいくつかのコンピュータ、すなわち候補コンピュータ（図示省略）をも自己の傘下として管理しているものとする。この候補コンピュータを他のクラスタに提供する場合、その候補コンピュータを管理する管理コンピュータがスレーブ管理コンピュータとなることは、前述のとおりである。

クラスタC20を管理する管理コンピュータは、クラスタC6～C11の中から、クラスタC7を探し出す。そして、このクラスタC7を管理しているマスター管理コンピュータに問い合わせ、そのマスターテーブル2163におけるクラスタC7と他のクラスタとの関係付けの情報を入手する。図15中段の場合、クラスタC7はクラスタC11とのみ関係付けられている。そこで、クラスタC20を管理する管理コンピュータは、候補コンピュータをクラスタC11に追加させ、この候補コンピュータとクラスタC20（それに所属するコンピュータ）とを関係付ける（図15下段）。また、自己が保有するスレーブテーブルの内容を更新する。

他方、クラスタC11を管理するマスター管理コンピュータは、マスターテーブル2163に、新たに追加されたコンピュータ（追加前の候補コンピュータ）のリストをコンピュータリスト2363に追加するとともに、クラスタリスト2369に、クラスタC20の識別情報を追加する。

以上は、クラスタC20とクラスタC11（各々のコンピュータ）同士が1対1で関係付けた場合の例であるが、関係付けは必ずしも1対1だけではなく、1対N（Nは2以上の自然数）であっても良い。また、この関係付けは、物理的なコンソール内のコンピュータ同士のつながりとは関係がない。

## 30

例えば、図16(a)は、ネットワークシステムに新たに加わるコンピュータの所属するクラスタC20が形成された場合において、クラスタC20を管理する管理コンピュータが、自己の傘下の候補コンピュータの追加先を探す際に、クラスタC11に到達したとする。クラスタC11のクラスタリストには、クラスタC11のコンピュータと1対1で関係付けられているクラスタC6、C7のほか、クラスタC11の複数のコンピュータと関係付けられているC8、C10がリストアップされていたとすると、候補コンピュータは、クラスタC6、C7、C8、C10のいずれかに追加される。図16(b)の場合も同様である。

10 なお、これらの場合において、スレーブテーブル、マスターテーブルの記録情報が更新されることは、図15に示した例の場合と同様である。

デュープリケーションモードでは、クラスタリスト2369とコンピュータリスト2363とを併用した運用形態も可能である。以下、この場合の運用形態例を説明する。

15 ネットワークシステムに加わる新たにクラスタC20が形成され、クラスタC20を管理する管理コンピュータが、自己の傘下の候補コンピュータの追加先を探し出す際に、クラスタC11に到達するまでは、図16(a)、(b)の例と同じである。

クラスタC11のコンピュータリスト2363によれば、図17に示されるように、コンピュータ#1~#5が所属しており、クラスタリスト2369によれば、コンピュータ#1、#3はクラスタC10、コンピュータ#4はクラスタC7、コンピュータ#5はクラスタC6に、それぞれ関係付けられており、コンピュータ#2はどことも関係付けられていないとする。必ずしもそのようにする必要はないが、コンピュータリスト2363の読み出し順もランダムに行うとすると、コンピュータリスト2363からコンピュータ#1が選ばれたときは、それ  
25 に関係付けられているクラスタC10に、候補コンピュータが追加される。なお、コンピュータ#2が選ばれたときは、それに関係付けられているクラスタがないので、再選択される。

デュープリケーションモードでは、コネクションレシオ2368を設定しなくとも、ベキ分布に従うクラスタリングが可能になるので、コンピュータのクラス

タへの追加が、よりスムーズになるという効果がある。

#### 産業上の利用可能性

- 本発明は、コンピュータ同士をつなぐノードの数が、例えば、正規分布あるいはベキ分布に従うクラスタリングを容易にするためのネットワークシステムに適用が可能である。また、投入されるジョブのサイズ、種類、数が様々で、それらが時々刻々と変化していく情報処理を効率的に実行することができる分散コンピューティング全般にも適用が可能なものである。

## 32

## 請求の範囲

1. それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムであって；
- 5 各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブルと；  
既に一又は複数のコンピュータにより形成されているクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルとを含み；
- 10 いずれかの前記コンピュータは、前記第1のテーブルのクラスタ化可否情報がクラスタ化可能の状態を表す自己および他のコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報を、クラスタ化不能の状態を表す情報に更新し、さらに形成したクラスタについての前記追加容易性情報を前記第2テーブルに記録する
- 15 クラスタ形成手段を有しており；  
クラスタを形成した前記コンピュータは、当該クラスタへの追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段を有している；ネットワークシステム。
- 20 2. 前記クラスタを形成した前記コンピュータは、当該クラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルに記録されている前記追加容易性情報および消滅したクラスタに所属していたすべてのコンピュータについて前記第1のテーブルに記録されていたクラスタ化可否情報をクラスタ形成前の状態に復帰させるクラスタ消滅手段；をさらに有し
- 25 ている、  
請求の範囲第1項記載のネットワークシステム。
3. 前記追加容易性情報は前記追加のし易さを定量化した数値であり、前記クラスタ成長手段は、前記数値が高くなるほど前記候補コンピュータを当該クラスタへ追加させ易くする、

## 33

請求の範囲第1項記載のネットワークシステム。

4. 前記クラスタ成長手段は、候補コンピュータの追加に拘わらず、前記数値を一定値に保持させる、

請求の範囲第3項記載のネットワークシステム。

5. 前記クラスタ成長手段が、前記数値を前記第2のテーブルに記録した後に変動可能にする、

請求の範囲第3項記載のネットワークシステム。

6. 前記クラスタ成長手段は、前記候補コンピュータを前記クラスタに追加したことを契機に前記数値を変動させる、

10 請求の範囲第5項記載のネットワークシステム。

7. それぞれ、他のコンピュータとのクラスタ化が可能な一又は複数のコンピュータを自己の傘下とする複数の管理コンピュータが自由に参加および離脱することができるネットワークシステムであって、

15 各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブルと、

既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録した第2のテーブルとを含み、

20 少なくとも一つの前記管理コンピュータは、自己の傘下にあるコンピュータおよび他の管理コンピュータの傘下にあるコンピュータのうち前記第1のテーブルのクラスタ化可否情報がクラスタ化可能の状態を表すコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新するクラスタ形成手段を有しており、

25 クラスタを形成した前記管理コンピュータは、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段を有している、

ネットワークシステム。

## 34

8. 前記クラスタを形成した管理コンピュータは、当該クラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルに記録されている追加容易性情報および消滅したクラスタに所属していたすべてのコンピュータについて前記第1のテーブルに記録されていたクラスタ化可否情報
- 5 情報をクラスタ形成前の状態に復帰させるクラスタ消滅手段をさらに有している、  
請求の範囲第7項記載のネットワークシステム。
9. いずれかの前記管理コンピュータが、最大で自己の傘下にあるコンピュータの数の前記第2のテーブルを保有している、  
請求の範囲第8項記載のネットワークシステム。
- 10 10. 前記第2テーブルは、それを保有する管理コンピュータが自己の傘下にある第1コンピュータを含む第1クラスタを形成するときに当該第1コンピュータ用として生成するマスターテーブルと、それを保有する管理コンピュータが、他の管理コンピュータが形成した第2クラスタに追加される、自己の傘下にある第2コンピュータの動作を監視制御するときには当該第2コンピュータ用に生成
- 15 するスレーブテーブルのいずれかを少なくとも含んでおり、  
前記追加容易性情報は前記マスターテーブルに記録されており、  
前記マスターテーブルを保有する管理コンピュータは、前記第1クラスタの形成、第1クラスタ内のコンピュータ数の変更並びに第1クラスタの消滅に関わる情報処理を主導的に行うマスター管理コンピュータとして振る舞い、前記スレー
- 20 ブテーブルを保有する管理コンピュータは、前記第2クラスタに対するスレーブ管理コンピュータとして振る舞う、  
請求の範囲第9項記載のネットワークシステム。
11. 前記マスター管理コンピュータは、いずれかの前記管理コンピュータに対してクラスタ化可能状態のコンピュータがあるかどうかを問い合わせることにより前記第1クラスタに追加させるための候補コンピュータを探索する探索手段
- 25 を有する、  
請求の範囲第10項記載のネットワークシステム。
12. いずれかの前記管理コンピュータは、前記マスター管理コンピュータにより形成された第1クラスタについての前記追加容易性情報に基づいて自己の傘

## 35

- 下の候補コンピュータを当該第1クラスタに追加させるかどうかを決定する、  
請求の範囲第10項記載のネットワークシステム。
13. 前記追加容易性情報は、前記第1クラスタへの候補コンピュータの追加のし易さを定量化した数値であり、
- 5 前記マスター管理コンピュータが有するクラスタ成長手段は、前記数値が高くなるほど前記候補コンピュータを当該第1クラスタに追加させ易くする、  
請求の範囲第10項記載のネットワークシステム。
14. 前記クラスタ成長手段は、候補コンピュータの追加に拘わらず、前記数値を一定値に保持させる、
- 10 請求の範囲第13項記載のネットワークシステム。
15. 前記クラスタ成長手段は、前記数値を前記第2テーブルに記録した後に変動可能にする、  
請求の範囲第13項記載のネットワークシステム。
16. 前記クラスタ成長手段は、前記候補コンピュータを前記第1クラスタに追加したことを契機に前記数値を変動させる、  
請求の範囲第15項記載のネットワークシステム。
17. それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムであって；  
既に一又は複数のコンピュータにより形成されているクラスタ毎に、当該クラスタに所属するコンピュータに関係付けられている他のクラスタの識別情報をリストアップしたテーブルを含み；  
クラスタを形成した前記コンピュータは、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを自己のクラスタに関係付けられている他のクラスタに追加するかどうかを前記テーブルにリストアップされた識別情報に基づいて決定するクラスタ成長手段を有している；  
25 ネットワークシステム。
18. 前記テーブルには、前記自己のクラスタが実行する処理との関連性が高い関連クラスタほど前記候補コンピュータが追加されるように優先的にリストアップされる、

## 36

請求の範囲第17項記載のネットワークシステム。

19. 各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブルを含んでおり、

5 いずれかの前記コンピュータは、前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表す自己および他のコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報を、クラスタ化不能の状態を表す情報に更新するクラスタ形成手段を有している、

請求の範囲第17項記載のネットワークシステム。

10 20. それぞれ、他のコンピュータとのクラスタ化が可能な一又は複数のコンピュータを自己の傘下とする管理コンピュータであって；

他の同種機能を有する管理コンピュータと共に自由に参加および離脱することができるコンピュータネットワークに接続するためのネットワーク接続手段と；

15 各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブル、および、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルへのアクセスを可能にするテーブル管理手段と；

20 自己の傘下にある前記コンピュータおよび他の管理コンピュータの傘下にある前記コンピュータのうち前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表すコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタについての前記追加容易性情報を前記第2のテーブルに記録するクラスタ形成手段と；

25 追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段と；を有する管理コンピュータ。

21. 前記形成したクラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルに記録されている追加容易性情報

## 37

および消滅したクラスタに所属していたすべてのコンピュータについて前記第1のテーブルに記録されているクラスタ化可否情報をクラスタ形成前の状態に復帰させるクラスタ消滅手段をさらに有する、

請求の範囲第20項記載の管理コンピュータ。

- 5 22. それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるクラスタ管理方法であって；

各コンピュータが、それぞれ自己の状態がクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を第1のテーブルに記録する段階と；

- 10 10 いずれかのコンピュータが、自己と前記第1のテーブルのクラスタ化可否情報がクラスタ化可能の状態を表す他のコンピュータとを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を第2  
15 のテーブルに記録する段階と；

追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定する段階と；を有するクラスタ管理方法。

- 20 23. 前記クラスタを形成したコンピュータが、前記形成したクラスタによるジョブの実行が終了した時点で当該クラスタを消滅させるとともに、前記第2のテーブルに記録されている追加容易性情報および消滅したクラスタに所属していたすべてのコンピュータについて前記第1のテーブルに記録されているクラスタ化可否情報をクラスタ形成前の状態に復帰させる段階とを有する、

請求の範囲第22項記載のクラスタ管理方法。

- 25 24. それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるクラスタ管理方法であって；

既に一又は複数のコンピュータによりクラスタを形成したコンピュータが、当該クラスタに所属するコンピュータが関係付けられている他のクラスタの識別情

報を所定のテーブルにリストアップする段階と；

クラスタを形成した前記コンピュータが、追加候補となる候補コンピュータが存在するときに当該候補コンピュータを当該クラスタと関係付けられている他のクラスタに追加するかどうかを前記テーブルに記録されている識別情報に基づいて決定する段階と；を有する、クラスタ管理方法。

25. それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるいずれかのコンピュータに読み取られて実行されることにより当該コンピュータを、
- 各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表すクラスタ化可否情報を記録した第1のテーブル、および、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルへのアクセスを可能にするテーブル管理手段；

- 前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表す自己および他のコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記第1のテーブルのクラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタについての前記追加容易性情報を前記第2のテーブルに記録するクラスタ形成手段；

- 追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段；として機能させるためのコンピュータプログラム。

26. それぞれ他のコンピュータとのクラスタ化が可能な一又は複数のコンピュータを自己の傘下とする管理コンピュータに読み取られて実行されることにより、前記管理コンピュータを；

他の同種機能を有する管理コンピュータと共に自由に参加および離脱することができるコンピュータネットワークに接続するためのネットワーク接続手段；

各々のコンピュータの状態がそれぞれクラスタ化が可能な状態かどうかを表す

## 39

クラスタ化可否情報を記録した第1のテーブル、および、既に一又は複数のコンピュータにより形成されたクラスタが他のコンピュータをどの程度追加し易いかを表す追加容易性情報を記録するための第2のテーブルへのアクセスを可能にするテーブル管理手段；

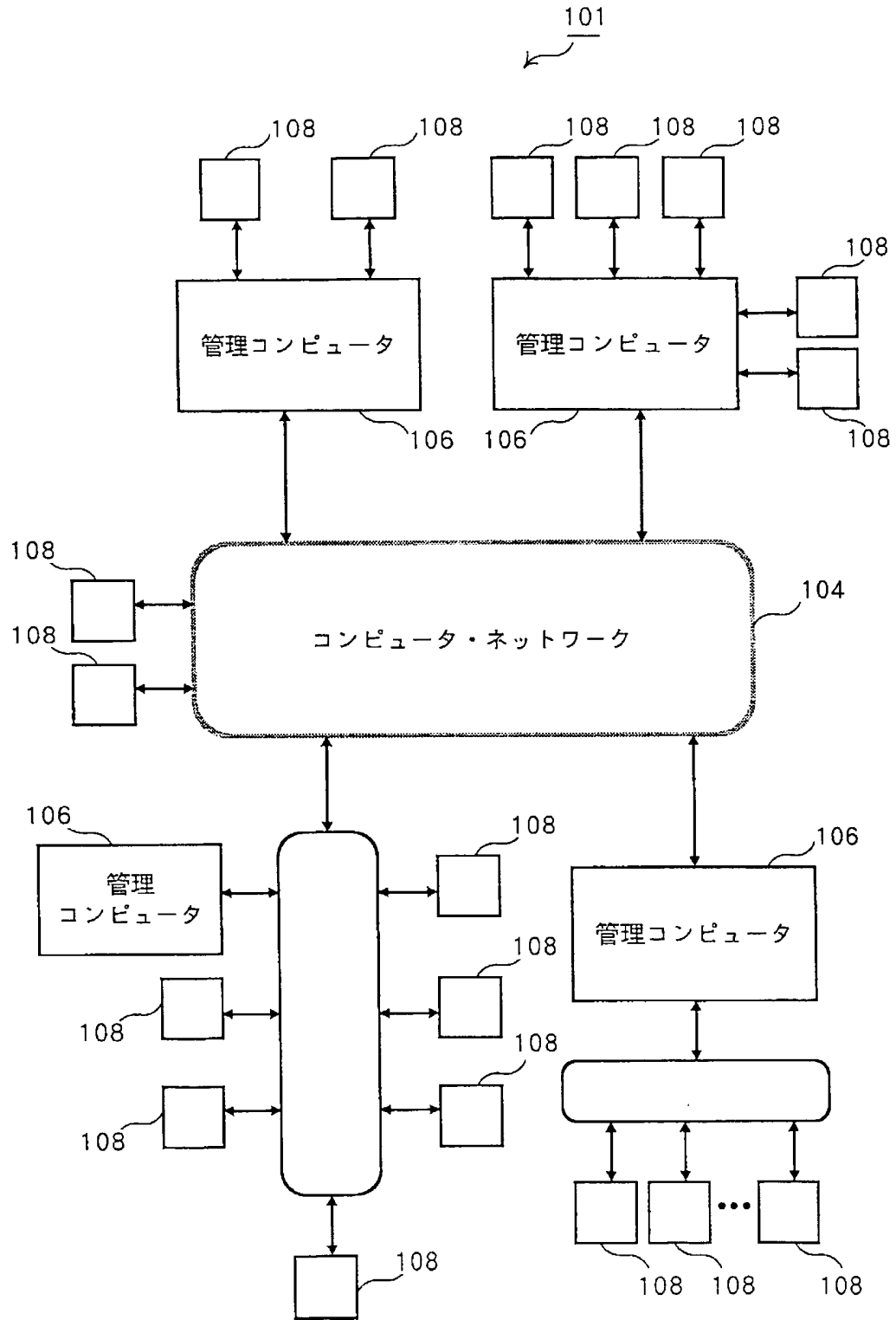
- 5 自己の傘下にある前記コンピュータおよび他の管理コンピュータの傘下にある前記コンピュータのうち前記第1のテーブルのクラスタ化可否情報がクラスタ化可能な状態を表すコンピュータを含んだクラスタを形成するとともに、形成したクラスタに含まれるすべてのコンピュータについての前記クラスタ化可否情報をクラスタ化不能の状態を表す情報に更新し、さらに、当該クラスタについての前記追加容易性情報を前記第2のテーブルに記録するクラスタ形成手段；

追加候補となる候補コンピュータが存在するときに当該候補コンピュータを前記クラスタに追加するかどうかを前記第2のテーブルに記録されている追加容易性情報に基づいて決定するクラスタ成長手段；として機能させるためのコンピュータプログラム。

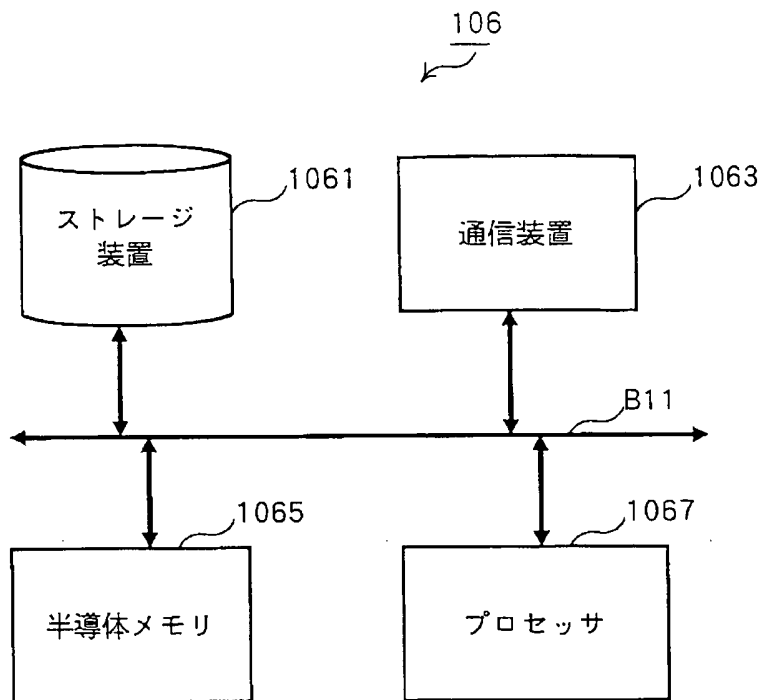
- 15 27. それぞれ、他のコンピュータとのクラスタ化が可能な複数のコンピュータが自由に参加および離脱することができるネットワークシステムにおけるいずれかのコンピュータに読み取られて実行されることにより、当該コンピュータを；

- 20 既に一又は複数のコンピュータによりクラスタを形成したときに当該クラスタに所属するコンピュータが関係付けられている他のクラスタの識別情報を所定のテーブルにリストアップするクラスタ管理手段；

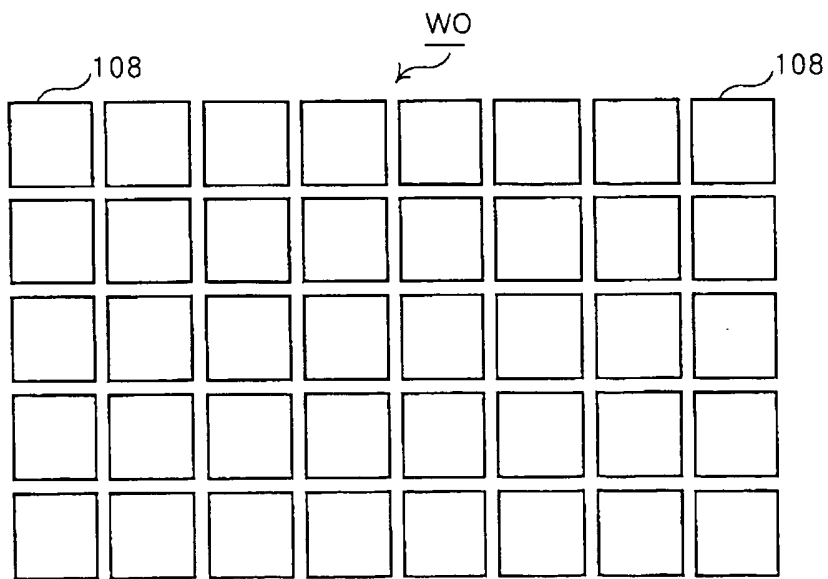
- 25 追加候補となる候補コンピュータが存在するときに当該候補コンピュータを自己のクラスタに関係付けられている他のクラスタに追加するかどうかを前記テーブルにリストアップされた識別情報に基づいて決定するクラスタ成長手段；として機能させるためのコンピュータプログラム。



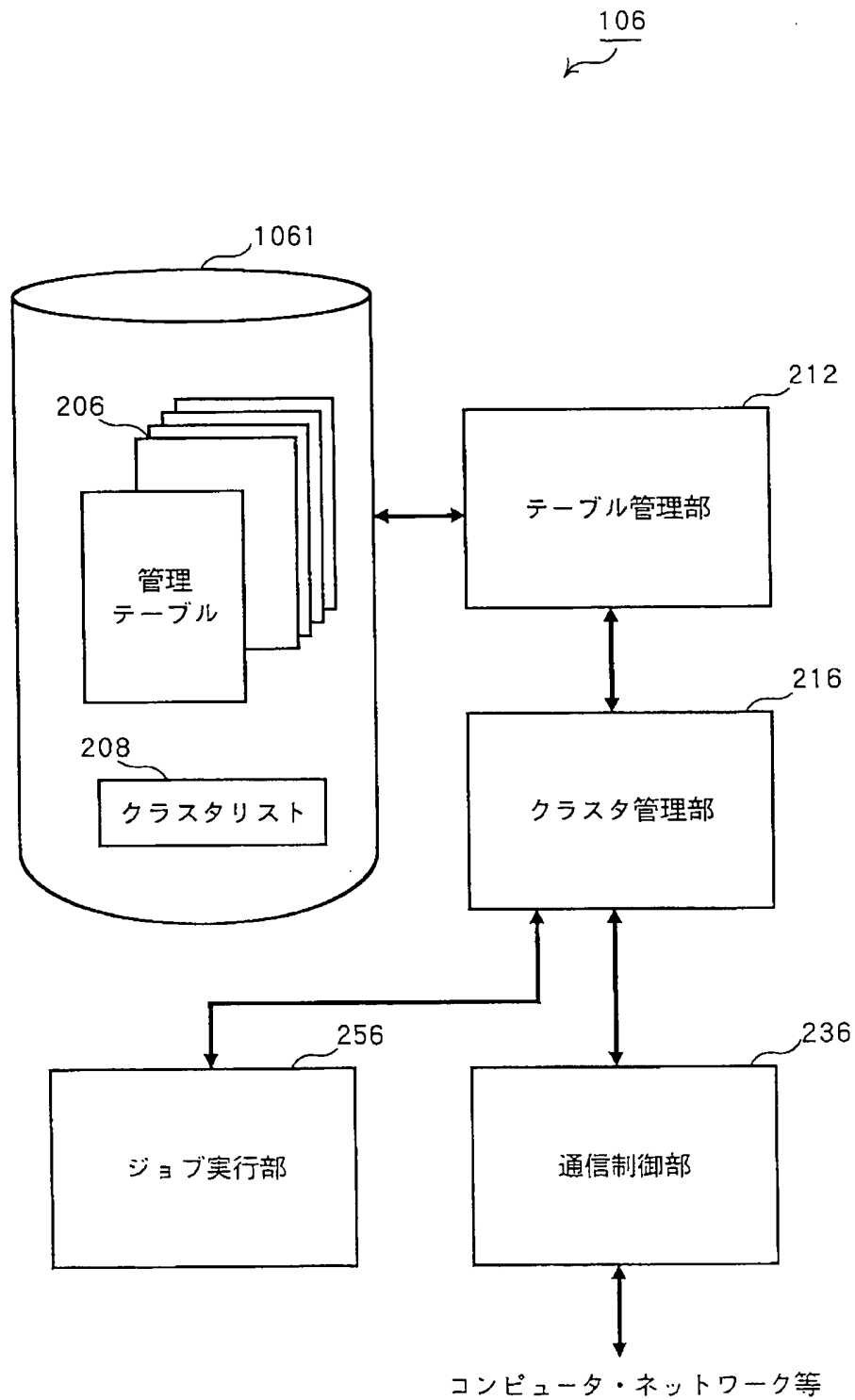
第1図



第2図



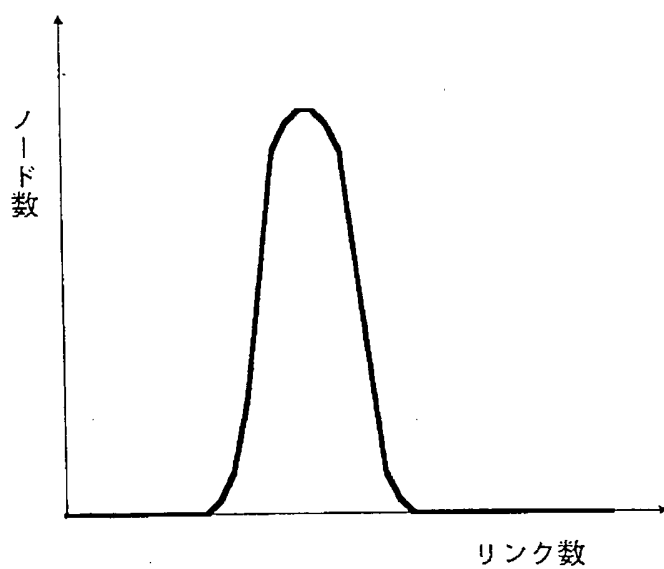
第3図



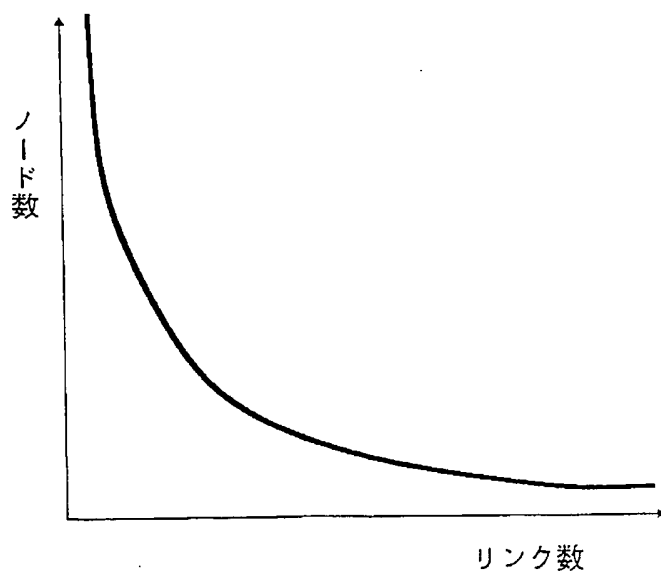
第4図

4/16

(a)



(b)



第5図

5/16

(a)

2161

2361	クラスタ ID	*****
2362	クラスタサイズ	*****
2363	コンピュータ リスト	
2364	クラスタ・ステータス	idol
2365	最大 コンピュータ数	200
2366	最小 コンピュータ数	3
2367	トータル コンピュータ数	48
2368	コネクション レシオ	0.48

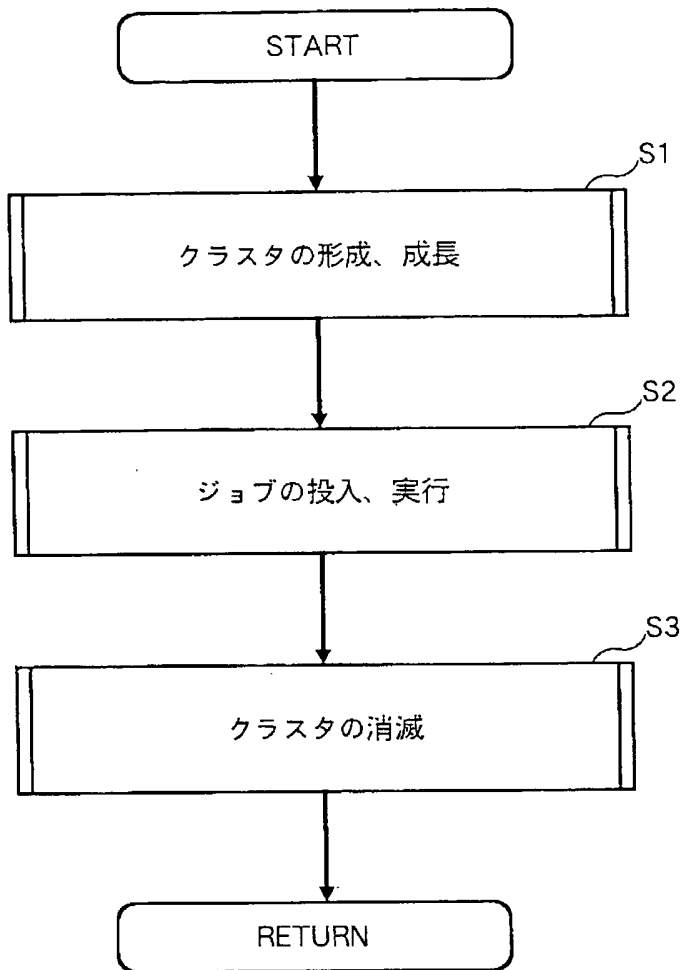
(b)

2162

2461	クラスタ ID	*****
2462	コンピュータ id	*****

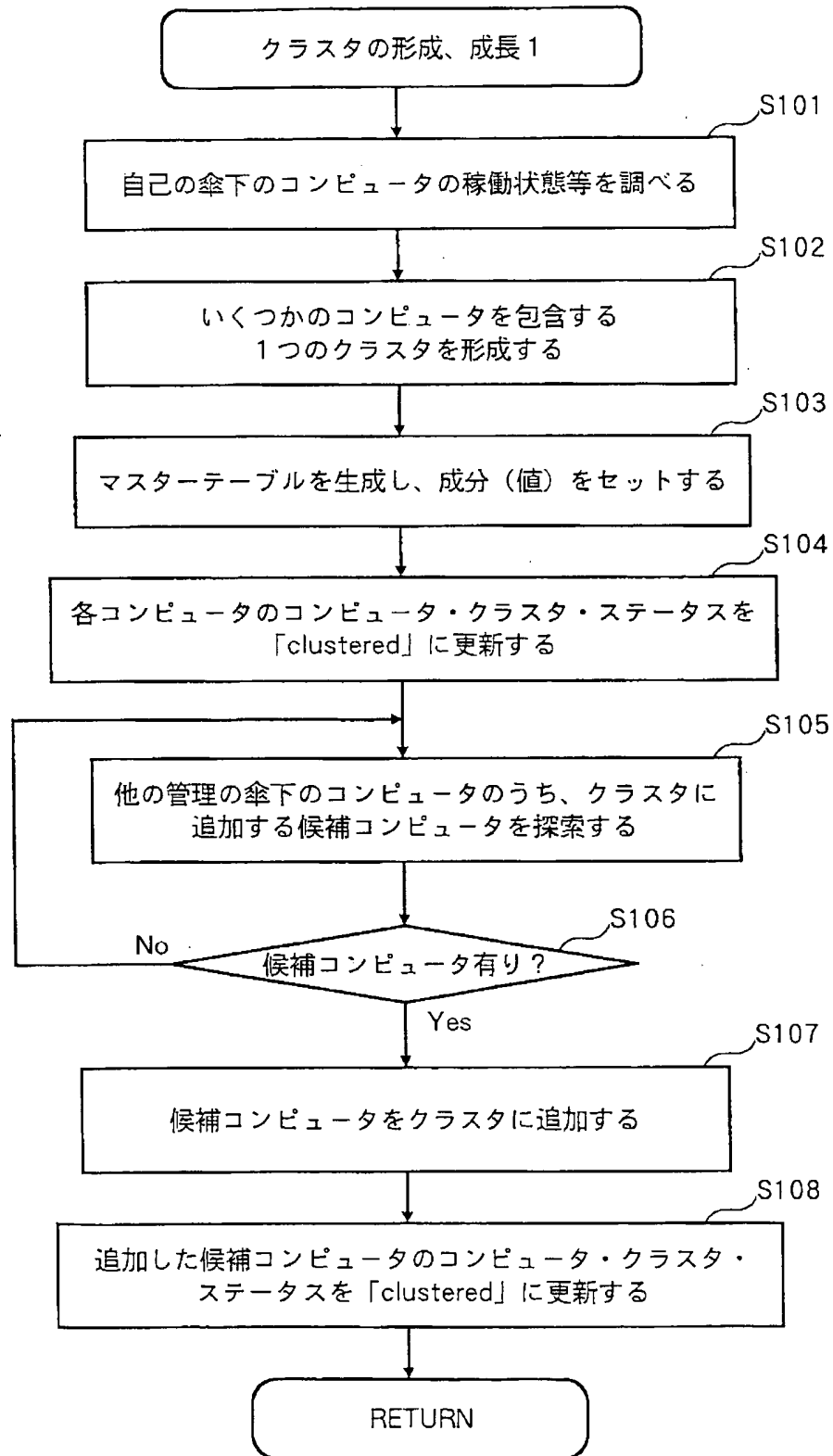
第6図

6/16

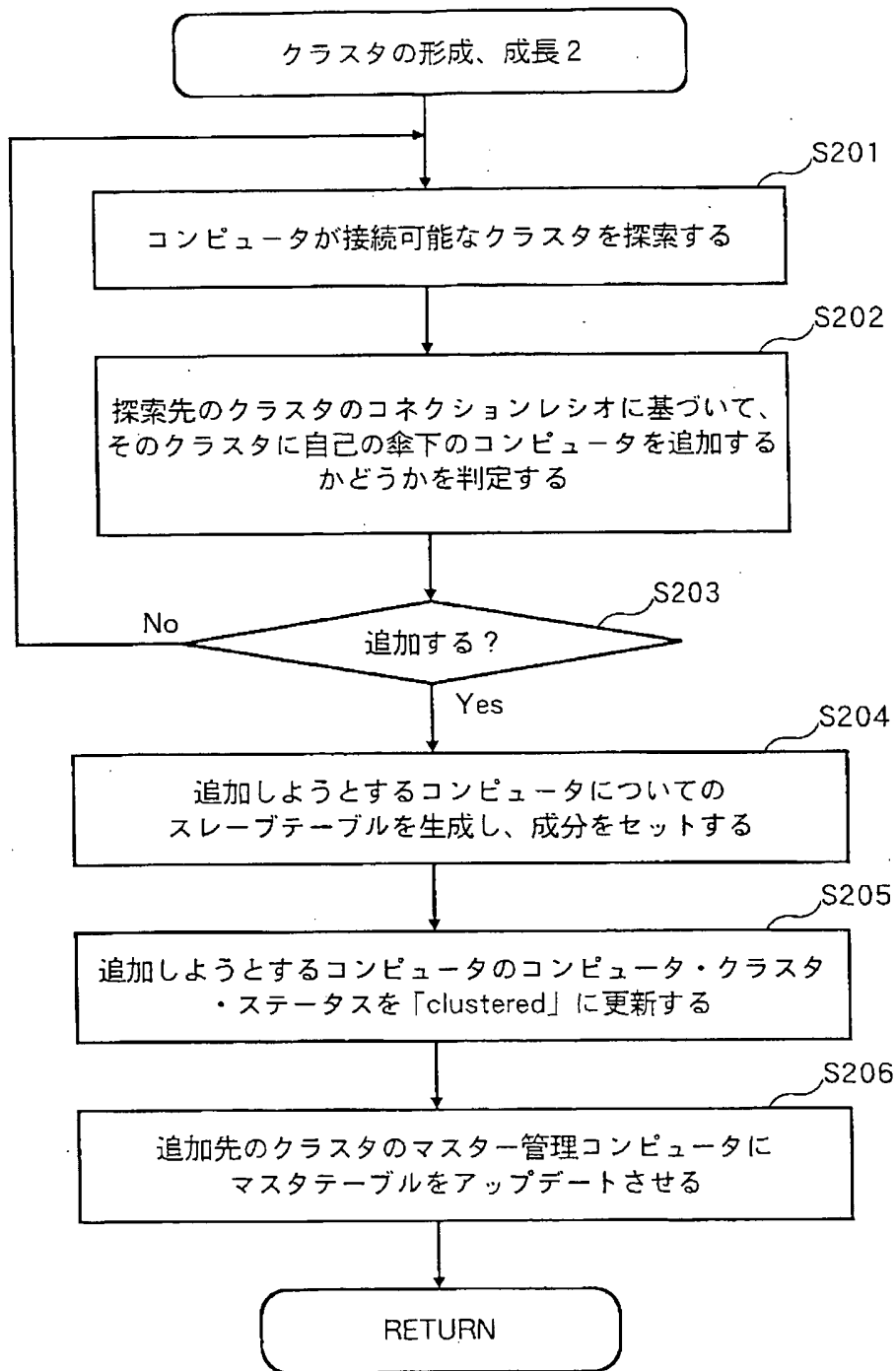


第7図

7/16

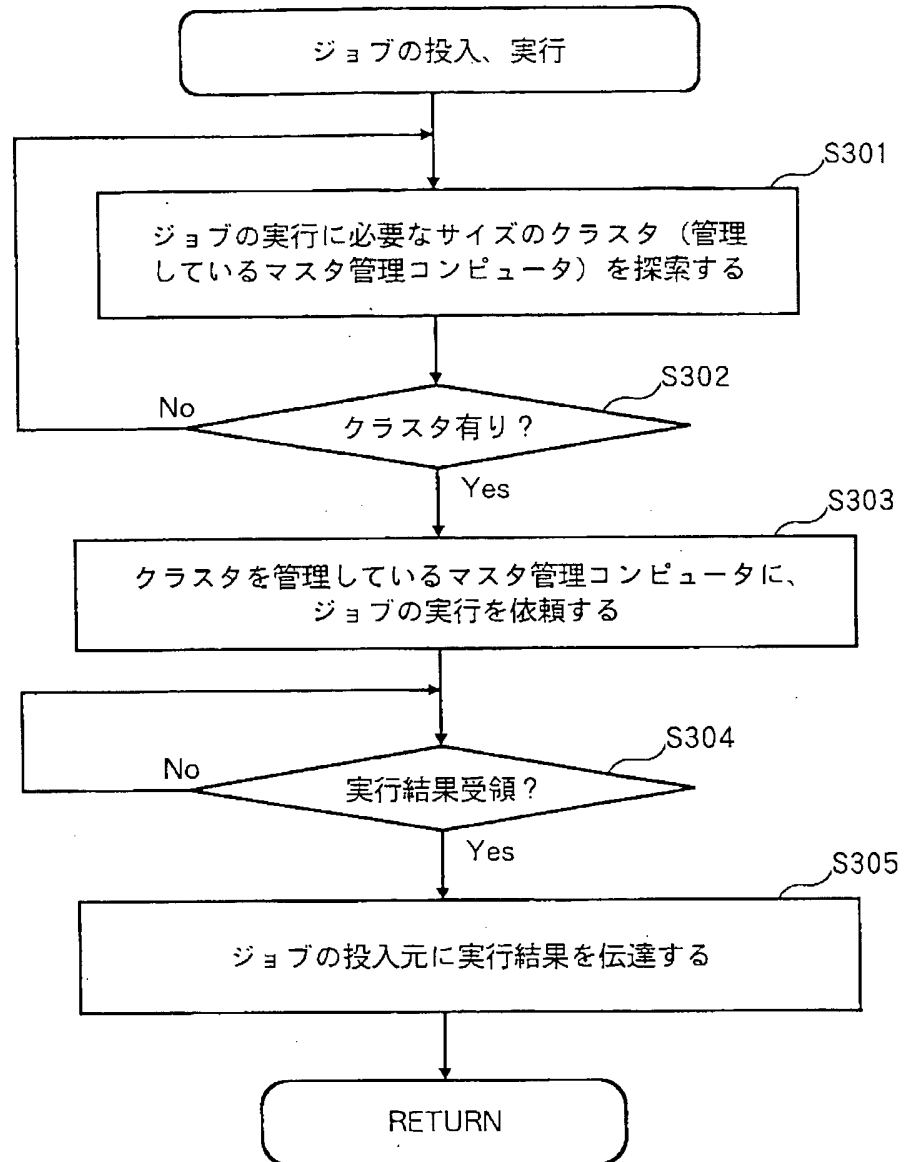


第8図



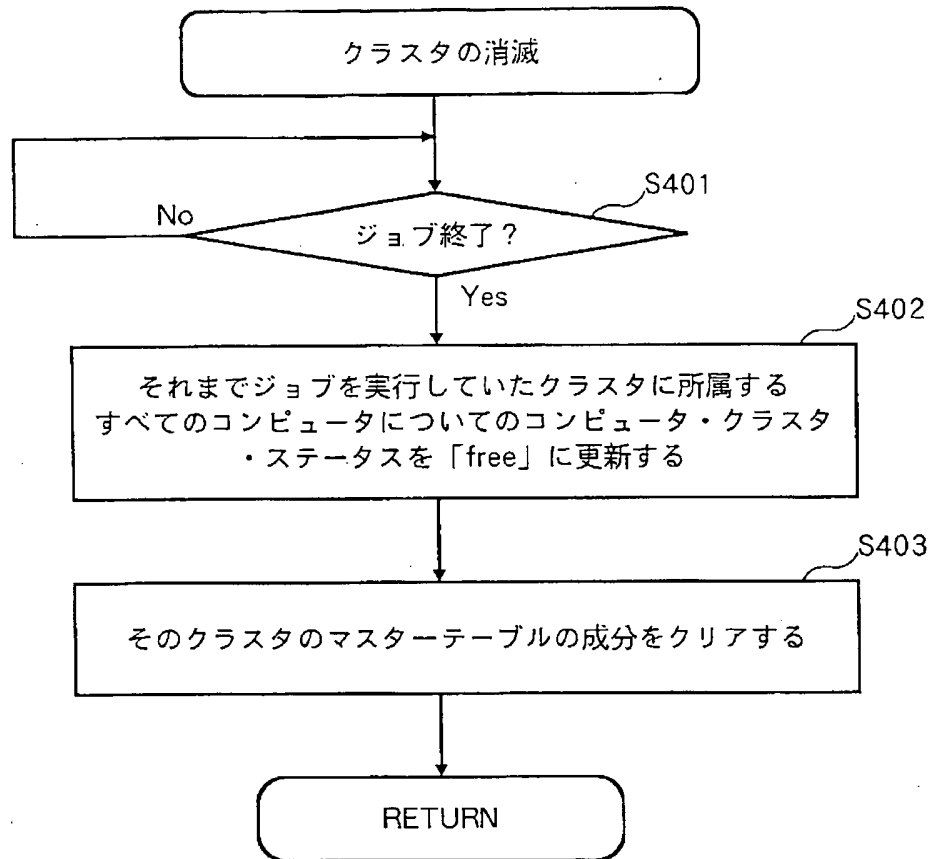
第9図

9/16



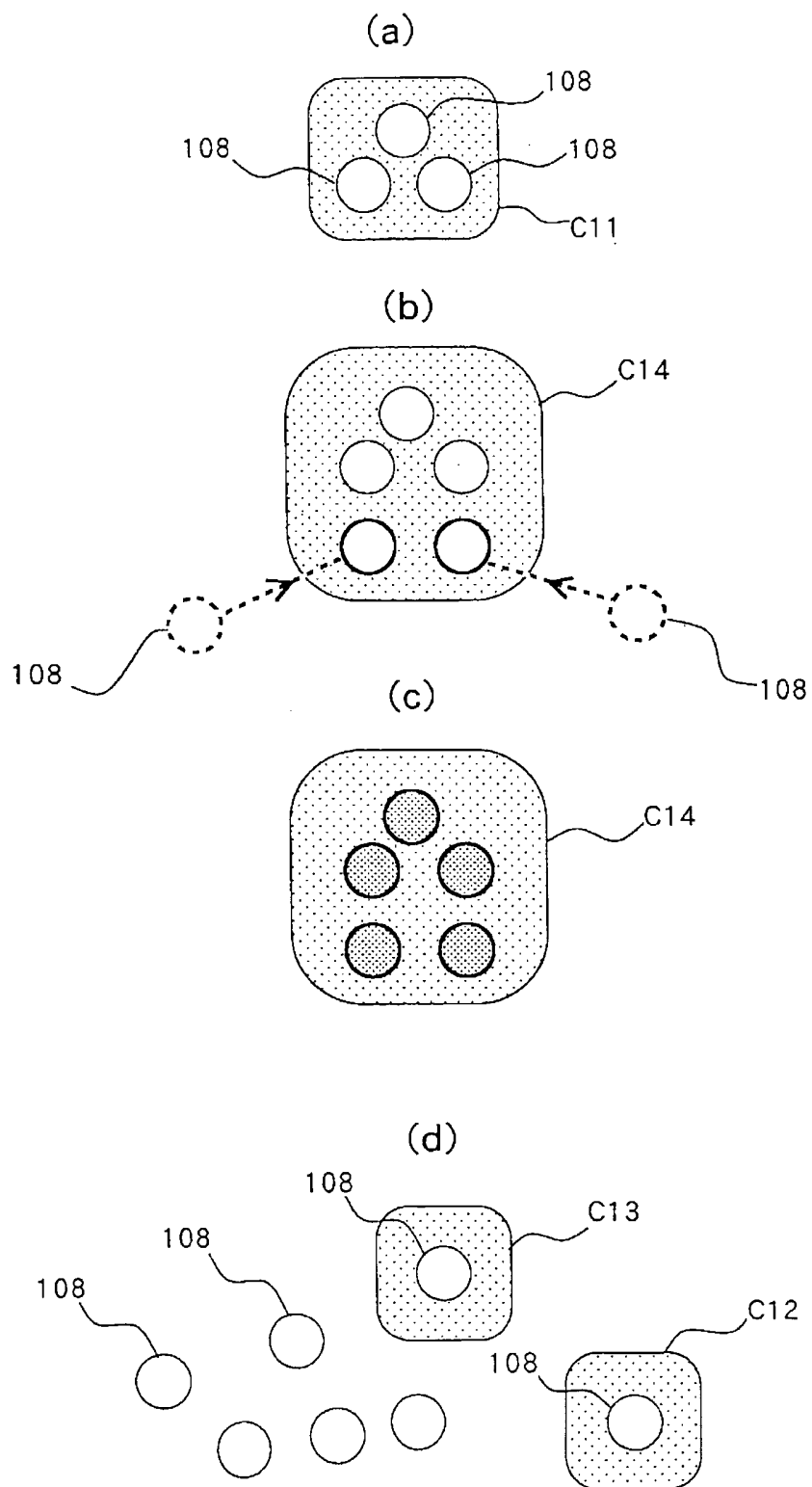
第10図

10/16



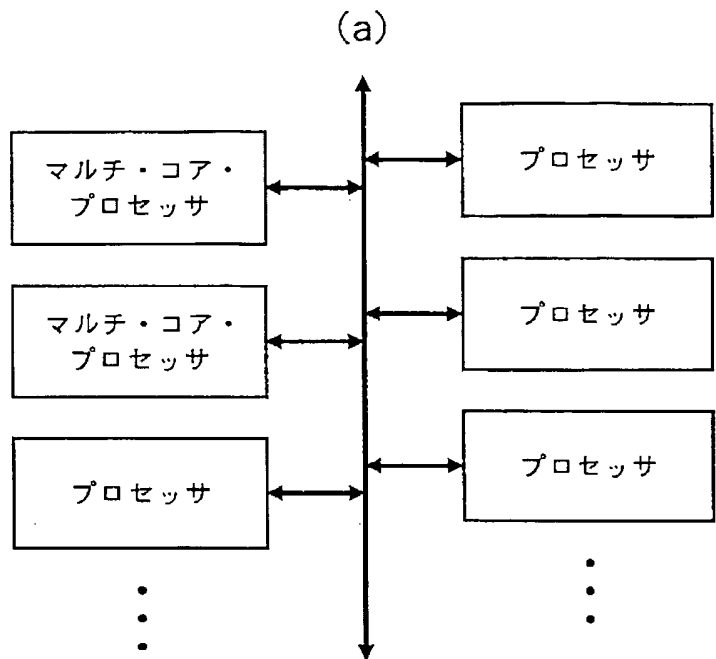
第11図

11/16

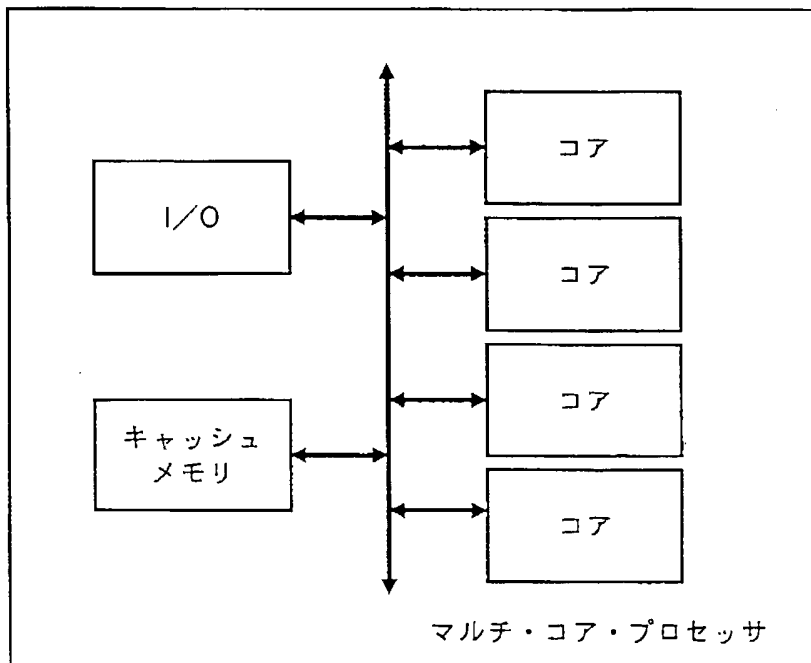


第12図

12/16



(b)



第13図

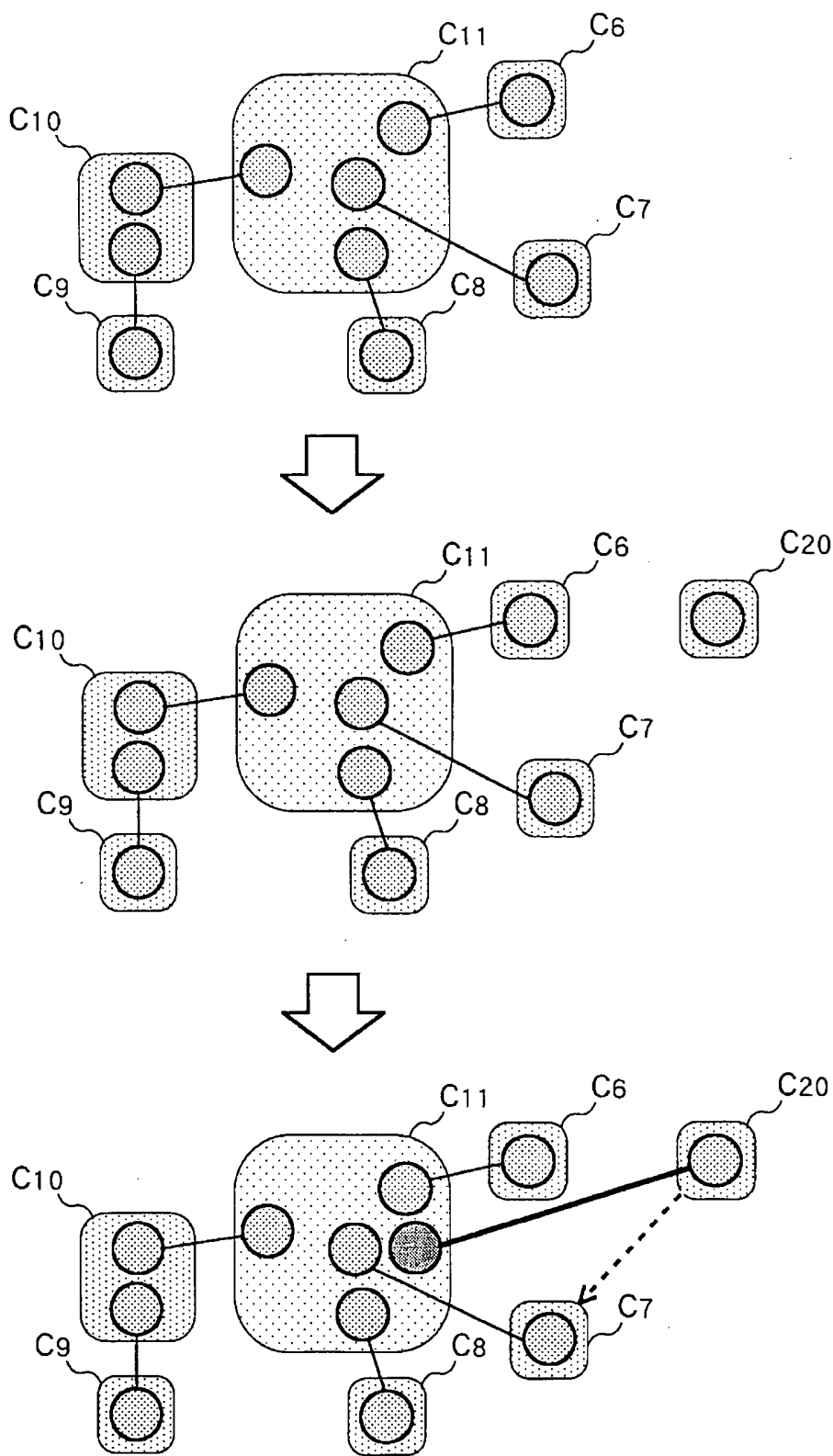
13/16

2163  
↙

2361	クラスタ ID	*****
2362	クラスタサイズ	*****
2363	コンピュータ リスト	..... ..... .....
2369	クラスタリスト	..... ..... .....
2364	クラスタ・ステータス	idol
2365	最大 コンピュータ数	200
2366	最小 コンピュータ数	3

第14図

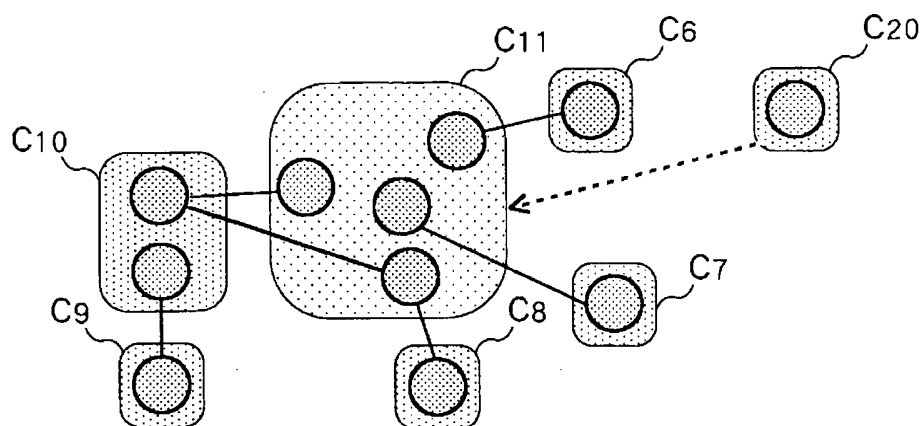
14/16



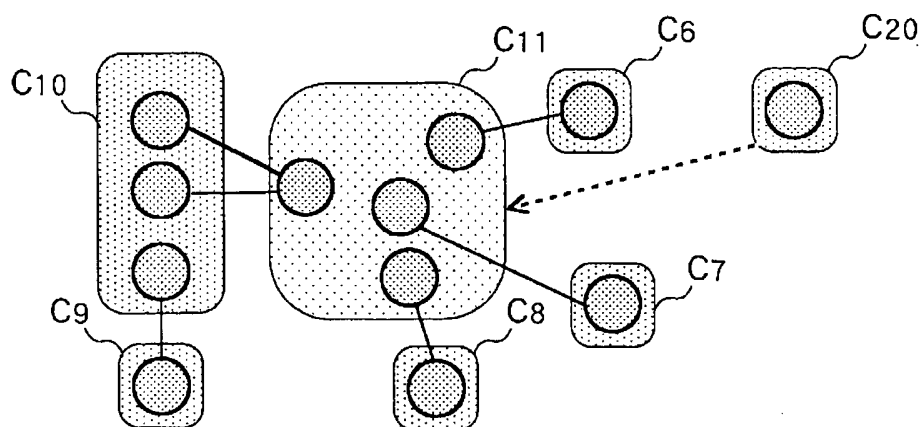
第15図

15/16

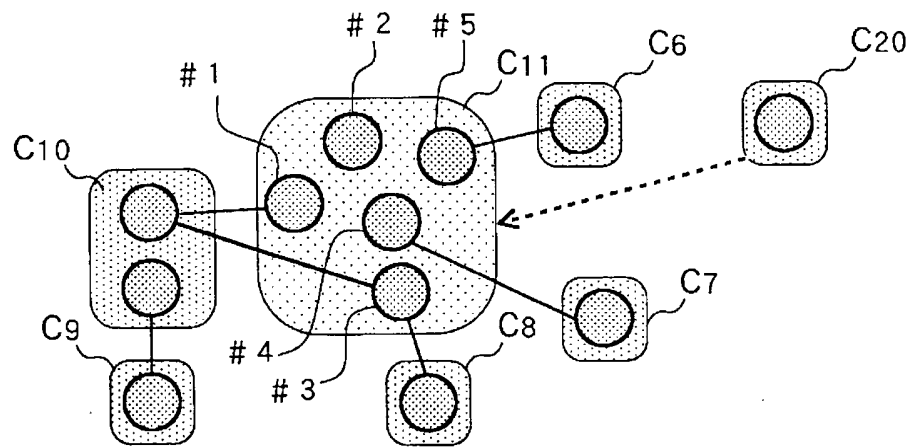
(a)



(b)



第16図



第17図

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2005/014467

A. CLASSIFICATION OF SUBJECT MATTER <b>G06F9/50</b> (2006.01)		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) <b>G06F9/46</b> (2006.01), <b>G06F9/50</b> (2006.01), <b>G06F13/00</b> (2006.01), <b>G06F15/16-15/177</b> (2006.01), <b>G06F15/00</b> (2006.01), <b>G06F12/00</b> (2006.01)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2005 Kokai Jitsuyo Shinan Koho 1971-2005 Toroku Jitsuyo Shinan Koho 1994-2005		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) IEEE, JSTplus (JOIS)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Jaesun Han and Daeyeon Park, "A lightweight personal grid using a supernode network", Proceeding of Third International Conference on Peer-to-Peer Computing (P2P'03), 03 September, 2003 (03.09.03), pages 168 to 175	1-27
A	Elth Ogston et al., "A method for decentralized clustering in large multi-agent systems", Proceedings of the second international joint conference on Autonomous agents and multiagent systems, 18 July, 2003 (18.07.03), pages 789 to 796	1-27
A	Wen Dou et al., "A P2P approach for global computing", Proceedings of International Parallel and Distributed Processing Symposium (IPDPS'03), 26 April, 2003 (26.04.03), CD-ROM	1-27
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 14 November, 2005 (14.11.05)		Date of mailing of the international search report 29 November, 2005 (29.11.05)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2005/014467

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Alexei Vázquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations", 2003, Physical Review, Vol.67, No.5, 056104.1-056104.15	1-27
P,A	JP 2004-341962 A (Kabushiki Kaisha Toyota IT Kaihatsu Senta), 02 December, 2004 (02.12.04), Full text; all drawings (Family: none)	1-27
E,A	JP 2005-251160 A (Nippon Telegraph and Telephone East Corp.), 15 September, 2005 (15.09.05), Full text; all drawings (Family: none)	1-27

A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. <b>G06F9/50</b> (2006.01)		
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. <b>G06F9/46</b> (2006.01), <b>G06F9/50</b> (2006.01), <b>G06F13/00</b> (2006.01), <b>G06F15/16-15/177</b> (2006.01), <b>G06F15/00</b> (2006.01), <b>G06F12/00</b> (2006.01)		
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2005年 日本国実用新案登録公報 1996-2005年 日本国登録実用新案公報 1994-2005年		
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語) IEEE, JSTPlus(JOIS)		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	Jaesun Han and Daeyeon Park, "A lightweight personal grid using a supernode network", Proceedings of Third International Conference on Peer-to-Peer Computing (P2P'03), 2003.09.03, pp.168-175	1-27
A	Elth Ogston et al., "A method for decentralized clustering in large multi-agent systems", Proceedings of the second international joint conference on Autonomous agents and multiagent systems, 2003.07.18, pp.789-796	1-27
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <span style="margin-left: 200px;"><input type="checkbox"/> パテントファミリーに関する別紙を参照。</span>		
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献		
国際調査を完了した日 14.11.2005		国際調査報告の発送日 29.11.2005
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 赤川 誠一 電話番号 03-3581-1101 内線 3546
		5 B    3 2 4 0

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	Wen Dou et al. , "A P2P approach for global computing", Proceedings of International Parallel and Distributed Processing Symposium (IPDPS' 03), 2003. 04. 26, CD-ROM	1-27
A	Alexei Vázquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations", 2003, Physical Review, vol. 67, no. 5, 056104.1-056104.15	1-27
PA	JP 2004-341962 A(株式会社トヨタ I T 開発センター), 2004. 12. 02, 全文, 全図(ファミリーなし)	1-27
EA	JP 2005-251160 A(東日本電信電話株式会社), 2005. 09. 15, 全文, 全図(ファミリーなし)	1-27