

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2010-509677

(P2010-509677A)

(43) 公表日 平成22年3月25日 (2010.3.25)

|                                |                       |             |
|--------------------------------|-----------------------|-------------|
| (51) Int.Cl.                   | F I                   | テーマコード (参考) |
| <b>G 0 6 F 12/00 (2006.01)</b> | G O 6 F 12/00 5 4 5 A | 5 B O 8 2   |
|                                | G O 6 F 12/00 5 3 1 M |             |

審査請求 未請求 予備審査請求 未請求 (全 18 頁)

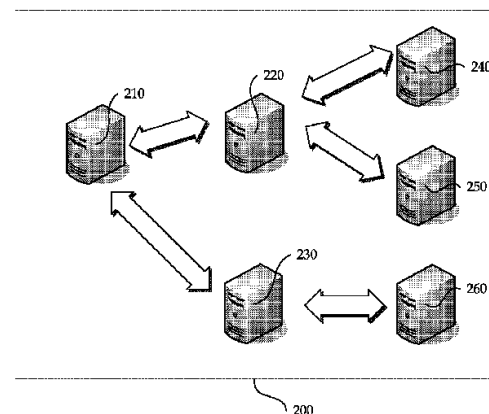
|               |                              |          |  |
|---------------|------------------------------|----------|--|
| (21) 出願番号     | 特願2009-536366 (P2009-536366) | (71) 出願人 | 500046438<br>マイクロソフト コーポレーション<br>アメリカ合衆国 ワシントン州 9805<br>2-6399 レッドモンド ワン マイ<br>クロソフト ウェイ                 |
| (86) (22) 出願日 | 平成19年8月29日 (2007. 8. 29)     | (74) 代理人 | 100077481<br>弁理士 谷 義一  |
| (85) 翻訳文提出日   | 平成21年5月11日 (2009. 5. 11)     | (74) 代理人 | 100088915<br>弁理士 阿部 和夫   |
| (86) 国際出願番号   | PCT/US2007/077114            | (72) 発明者 | ロバート エイチ. ガーバー<br>アメリカ合衆国 98052 ワシントン<br>州 レッドモンド ワン マイクロソフト<br>ウェイ マイクロソフト コーポレーシ<br>ョン インターナショナル パテンツ内 |
| (87) 国際公開番号   | W02008/057649                |          |  |
| (87) 国際公開日    | 平成20年5月15日 (2008. 5. 15)     |          |  |
| (31) 優先権主張番号  | 11/558, 887                  |          |  |
| (32) 優先日      | 平成18年11月10日 (2006. 11. 10)   |          |  |
| (33) 優先権主張国   | 米国 (US)                      |          |  |

最終頁に続く

(54) 【発明の名称】 分散型サーバシステムにおいてバックアップマネージャを転送するメッセージ

## (57) 【要約】

分散型サーバシステムは、主マネージャ、主マネージャに通信接続された1つまたは複数のバックアップマネージャ、および1つまたは複数のバックアップマネージャに通信接続された1つまたは複数のワーカーを含むことができる。1つまたは複数のバックアップマネージャは、分散型サーバシステムの状態をバックアップするコマンドを受信することができる。バックアップマネージャは、ワークタスクを1つまたは複数の通信接続されたワーカーに転送することができる。バックアップマネージャはさらに、1つまたは複数の通信接続されたワーカーから受信した応答を集め、集められた応答を主マネージャに送信することができる。



**【特許請求の範囲】****【請求項 1】**

分散型サーバーコンピュータシステムにおいて、  
分散型サーバー管理ソフトウェアサービスを提供する主マネージャと、  
前記主マネージャに通信接続された 1 つまたは複数のバックアップマネージャであって、  
前記主マネージャによって要求された情報をバックアップし、および前記主マネージャから受信した要求に回答して少なくとも 1 要求を転送するバックアップマネージャと、  
前記 1 つまたは複数のバックアップマネージャに通信接続された 1 つまたは複数のワーカーであって、前記 1 つまたは複数のバックアップマネージャによって前記主マネージャから転送された少なくとも 1 要求を受信するワーカーと  
を備えたことを特徴とする分散型サーバーコンピュータシステム。

10

**【請求項 2】**

前記 1 つまたは複数のバックアップマネージャはさらに、前記 1 つまたは複数のワーカーからの回答を受信することを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

**【請求項 3】**

前記 1 つまたは複数のバックアップマネージャはさらに、前記 1 つまたは複数のワーカーから受信したそれぞれの回答を集めることを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

**【請求項 4】**

前記 1 つまたは複数のバックアップマネージャはさらに、前記 1 つまたは複数のワーカーから受信したそれぞれの回答を集めること、およびさらに、前記 1 つまたは複数のワーカーからのワーカーが回答していない場合、回答に代わって前記 1 つまたは複数のワーカーからのワーカーの接続状態に関する情報を代用することを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

20

**【請求項 5】**

前記主マネージャおよび前記 1 つまたは複数のバックアップマネージャは、物理的コンピュータシステム上で実行することを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

**【請求項 6】**

前記主マネージャおよび前記 1 つまたは複数のワーカーは、物理的コンピュータシステム上で実行することを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

30

**【請求項 7】**

前記 1 つまたは複数のワーカーの 1 つが前記 1 つまたは複数のバックアップマネージャからの回答しないバックアップマネージャに割り当てられた場合、前記主マネージャはさらに、前記 1 つまたは複数のワーカーの 1 つを前記 1 つまたは複数のバックアップマネージャからの回答するバックアップマネージャに再度割り当てることを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

**【請求項 8】**

前記主マネージャはさらに、前記 1 つまたは複数のワーカーからの回答しないワーカーに割り当てられたワークを、前記 1 つまたは複数のワーカーからの回答するワーカーに再度割り当てることを特徴とする請求項 1 に記載の分散型サーバーコンピュータシステム。

40

**【請求項 9】**

分散型サーバーシステムのバックアップマネージャによる実行方法であって、  
前記分散型サーバーシステムの状態と関連付けられた ID を閉じ始める要求を主マネージャから受信するステップと、  
前記主マネージャに対する回答を送信して前記 ID に対応する前記状態が閉じられたことを示すステップと、

50

前記 ID と関連付けられたタスクを送信する要求を前記主マネージャから受信するステップと、

前記 ID と関連付けられた前記タスクを 1 つまたは複数のワーカーに送信するステップと

を備えることを特徴とする方法。

【請求項 10】

前記分散型サーバーシステムの前記状態を持続するステップをさらに備えることを特徴とする請求項 9 に記載の方法。

【請求項 11】

前記ワークが実行されたことを示す応答を前記 1 つまたは複数のワーカーから受信するステップをさらに備えることを特徴とする請求項 9 に記載の方法。

【請求項 12】

前記 1 つまたは複数のワーカーから応答を集めるステップをさらに備えることを特徴とする請求項 9 に記載の方法。

【請求項 13】

前記 ID と関連付けられた前記タスクを前記 1 つまたは複数のワーカーに送信するステップはさらに、前記分散型サーバーシステムの前記状態と関連付けられた前記 ID を閉じる要求を含むことを特徴とする請求項 9 に記載の方法。

【請求項 14】

実行すべき新しいワークを含む前記 1 つまたは複数のワーカーから応答を受信するステップをさらに備えることを特徴とする請求項 9 に記載の方法。

【請求項 15】

前記分散型サーバーシステムの状態と関連付けられた前記 ID が閉じられたことを示す応答を前記主マネージャに送信するステップをさらに備えることを特徴とする請求項 9 に記載の方法。

【請求項 16】

前記 1 つまたは複数のワーカーから受信した 1 つまたは複数の新しいワーク要求を含む応答を前記主マネージャに送信するステップをさらに備えることを特徴とする請求項 9 に記載の方法。

【請求項 17】

ステップを実行する装置実行可能な命令を記録する 1 つまたは複数の装置読み取り可能な媒体であって、

分散型サーバーシステムの状態と関連付けられた ID を閉じるコマンドを 1 つまたは複数のバックアップマネージャに送信するステップと、

第 1 の既定の時間間隔の間待機し、前記 1 つまたは複数のバックアップサーバーのそれぞれから応答を受信するステップと、

応答するバックアップサーバーの数が既定の数よりも多いかどうかを判定するステップと、

応答しなかった 1 つまたは複数のバックアップサーバーに割り当てられた 1 つまたは複数のワーカーを応答した 1 つまたは複数のバックアップサーバーに再度割り当てるステップと

を備えたことを特徴とする装置読み取り可能な媒体。

【請求項 18】

応答に対して応答しなかった前記 1 つまたは複数のバックアップサーバーのために第 2 の既定の時間間隔の間待機するステップをさらに備えたことを特徴とする請求項 17 に記載の 1 つまたは複数の装置読み取り可能な媒体。

【請求項 19】

前記 1 つまたは複数のワーカーの応答性について前記 1 つまたは複数のバックアップサーバーから情報を受信するステップをさらに備えたことを特徴とする請求項 17 に記載の 1 つまたは複数の装置読み取り可能な媒体。

10

20

30

40

50

**【請求項 20】**

前記バックアップマネージャは、前記分散型サーバーシステムの前記状態と関連付けられたIDを閉じるコマンドを前記1つまたは複数のワーカーに転送することを特徴とする請求項17に記載の1つまたは複数の装置読み取り可能な媒体。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、分散型サーバーシステムにおいてバックアップマネージャを転送するメッセージに関する。

**【背景技術】**

10

**【0002】**

分散型サーバーシステムは、ワークタスクを任意の数の通信接続されたコンピューターシステムに分散するセントラルサーバーを含むことができる。セントラルサーバーは、典型的なサーバークラスソフトウェアアプリケーションまたはソフトウェアサービスを実行するサーバーコンピューターとすることができる。通信接続されたコンピューターシステムは、ディスクリットサーバークラスコンピューターシステム、パーソナルコンピューターシステム等とすることができる。通信接続されたコンピューターシステムを、ネットワークによって同じ物理的場所に配置することができ、または異なる物理的場所に分散することができる。

**【0003】**

20

セントラルサーバーによって分散されるワークタスクは、コンピューターシステムがコンピューターソフトウェアにおいて実行することができる任意の種類のタスク、コマンドまたは関数呼び出しとすることができる。例えば、ワークタスクは、計算、データを記憶するコマンド等とすることができる。セントラルサーバーは、任意の典型的なコンピューター通信方法を使用して、ワークタスクを通信接続されたコンピューターシステムに送信することができる。例えば、セントラルサーバーは、HTTP(hypertext transport protocol)、HTTP経由のRPC(remote procedure call)、DCOM(distributed component object model)システム等を使用して、ワークタスクを送信することができる。

**【0004】**

セントラルサーバーはさらに、ソフトウェアアプリケーションまたはソフトウェアサービスを実行して、分散型サーバーシステムの現在の「状態」を管理および/または維持することができる。より詳細には、セントラルサーバーは、分散型サーバーシステム外部のエンティティに対して分散型サーバーシステム全体を表すことができる。従って、分散型サーバーシステムの状態に関する情報を探索する外部エンティティは、セントラルサーバーと通信を行い、セントラルサーバーは、分散型サーバーシステム全体の現在の状態に回答することができる。通信接続されたコンピューターシステムは、分散型コンピューターシステムの現在の状態を維持することもできるが、現在の状態はセントラルサーバーによって管理されるため、通信接続されたコンピューターシステムが分散型コンピューターシステムの現在の状態の維持を行うことは必須ではない。さらに、セントラルサーバーは、通信接続されたコンピューターシステムが互いに通信接続されたコンピューターシステム

30

40

**【0005】**

そのような分散型サーバーシステムによって実行されるワークタスク量は、セントラルサーバーに通信接続されたコンピューターシステムの数に関連することができる。例えば、各通信接続されたコンピューターシステムは、ワークタスクを実行することができる定率を有することができる。分散型サーバーシステムに割り当てられたワークタスク総量が通信接続されたコンピューターシステムが実行することができる量よりも多い場合、分散型サーバーシステムは、全体のワークタスクを完了させる際に低速化を起こす可能性がある。通信接続されたコンピューターシステムを分散型サーバーシステムにさらに追加することによって、分散型サーバーシステムのワークタスク容量を増加させ、全体のワークタ

50

スクを完了させる際のそのような低速化を抑制することができる。

【0006】

しかし通信接続されたコンピューターシステムを分散型サーバーシステムにさらに追加した結果、1つまたは複数の通信の問題が起こる場合がある。例えば、セントラルサーバーは、効率的にワークタスクを送信し、少数の通信接続されたコンピューターシステムから受信することができるかもしれない。分散型サーバーシステムの仕事容量は、通信接続されたコンピューターシステムのそれぞれを追加するにつれて、増加することができる。従って、セントラルサーバーによって実行される通信量も増加することができる。結果として、セントラルサーバーは、増加した数の通信を効率的に管理できないかもしれないし、そのような分散型サーバーシステムによって実行することができるワークタスク全体の数が減少するかもしれない。

10

【0007】

そのような分散型サーバーシステムは、冗長的に情報を格納するよう構成された1つまたは複数のバックアップコンピューターを使用することができる。そのようなバックアップコンピューターシステムは、バックアップマネージャまたはバックアップサーバーとしても知られている。バックアップマネージャまたはバックアップサーバーを、セントラルサーバーを冗長的に格納されてきた特定の「状態」に修復するよう構成することができる。しかし、1つまたは複数のバックアップマネージャまたはバックアップサーバーに対して通信の状態を変更する場合、セントラルサーバーは、実行される通信の総数も増加し、分散型サーバーシステム内の通信性能を再度低下させる結果となる可能性がある。

20

【0008】

通信接続されたコンピューターシステムの数を、セントラルサーバーの通信容量に悪影響を与えることなく増加させることができるシステムが有効である。

【発明の概要】

【0009】

本開示の簡略化した概要を以下に示し、基本的な理解を読者に提供する。本概要は、本開示の広範囲の要約ではない。本概要は、本発明の重要な/重大な要素を特定するものでも、本発明の範囲を明確にするものでもない。本概要の目的は、後に示すより詳細な説明の前置きとして、本明細書で開示したいいくつかの概念を簡略化した形式で示すことである。

30

【0010】

本実施例は、分散型サーバーシステムにおいてバックアップマネージャを転送するメッセージの機構および技術を提供する。サーバーは、他のコンピューターシステムに代わってソフトウェアサービスを実施し、実行するコンピューターシステムとすることができる。そのような他のコンピューターシステムは、クライアント、他のサーバー等として知られている。サーバーは、他のコンピューターシステムが要求をサーバーに伝えることができるように、典型的なコンピューター通信ネットワークを使用して他のコンピューターシステムに通信接続することができる。サーバーは、要求への適切な応答を準備し、伝えることができる。

【0011】

40

分散型サーバーは、1つまたは複数のサーバーのソフトウェアサービスを実行する単一のコンピューターシステム、またはセントラルサーバーに通信接続された複数のコンピューターシステム等を含むことができる。分散型サーバーシステムの主マネージャサーバーは、1つまたは複数の他のコンピューターシステムまたはサーバーに通信接続することができる。主マネージャは、クライアント要求を受信し、全体の動作または動作の一部を通信接続されたコンピューターシステムに委譲することができる。通信接続されたコンピューターシステムは、動作を実行し、成功または失敗を示す応答メッセージを送信することができる。そのような動作または動作の一部は、「ワーク」として知られている。

【0012】

主マネージャは、分散型サーバーシステムの「状態」を維持することもできる。そのよ

50

うな状態は、主マネージャによって管理された情報またはソフトウェアアプリケーションの現在の状態を示すことができる。例えば、セントラルマネージャは、分散型サーバーシステムによって格納されたデータベースの現在の状態を維持することができる。

【0013】

任意の数の通信接続されたコンピューターシステムは、バックアップマネージャまたはバックアップサーバーとして機能するよう指定することができる。主マネージャは、分散型サーバーシステムの状態に関連している情報を冗長的に格納する命令または情報を、1つまたは複数のそのようなバックアップマネージャに送信することができる。主マネージャは、他の情報またはデータを1つまたは複数のバックアップマネージャに送信することもできる。イベントにおいて状態情報、命令、またはデータのいずれかが主マネージャ上で失われた、または破損した場合、それらの状態情報、命令、またはデータを、バックアップマネージャのいずれかから復旧することができる。そのようなバックアップマネージャは、主マネージャにバックアップ動作が成功したことを知らせる実行可能な命令を含むこともできる。

【0014】

主マネージャを、1つまたは複数のそのようなバックアップマネージャに通信接続することができる。バックアップサーバーのそれぞれを、「ワーカー」とも呼ばれる、1つまたは複数のワーカーコンピューターシステムに通信接続することができる。主マネージャは、コマンドを1つまたは複数のバックアップマネージャに伝え、分散型サーバーシステムの現在の「状態」と関連付けられたIDを閉じる動作を開始することができる。コマンドはさらに、分散型サーバーシステムの現在の状態を持続、または格納する命令を含むことができる。

【0015】

各バックアップサーバーがコマンドと関連付けられた動作を完了した時点で、各バックアップサーバーは、成功または失敗を示すメッセージを主マネージャに送信することができる。1つまたは複数のバックアップサーバーが応答しない場合、主マネージャは、全てのバックアップサーバーが応答するまで待機するか、または、主マネージャは、応答しないバックアップサーバーに割り当てられた任意のワーカーを応答するバックアップサーバーに再度割り当てるための選択をすることができる。代替的な実施形態において、主マネージャは、バックアップサーバーがまだ応答していないことのみを認めることができ、および、応答しないバックアップサーバーに割り当てられた任意のワーカーを再度割り当てる後の時間まで待機することができる。

【0016】

主マネージャは次に、コマンドを1つまたは複数のバックアップマネージャのそれぞれに伝え、「ワーク」を送信することができる。そのような「ワーク」は、現在の状態ID (state identifier) によって表された現在のタスクまたは動作と関連付けられたタスクまたは動作とすることができる。それぞれのバックアップマネージャは次に、ワークを実行するコマンドを1つまたは複数の通信接続されたワーカーのそれぞれに転送することができる。各ワーカーがワークを完了すると、各ワーカーは、独立して新しいワークを決定するか、または、他のワーカーと通信を行って実行すべき任意の新しいワークを決定することができる。各ワーカーは次に、ワーカーに割り当てられたバックアップマネージャと通信を行うことができ、現在の状態IDと関連付けられたワークが完了したことを確認し、新しいワークを要求することができる。

【0017】

1つまたは複数のバックアップマネージャのそれぞれは次に、1つまたは複数のワーカーのそれぞれから受信した情報を集めることができる。ワーカーの1つが応答しなかったイベントにおいて、バックアップマネージャは、応答しない状態に関する情報を集められた情報に追加することができる。1つまたは複数のバックアップサーバーのそれぞれは次に、主マネージャと通信を行い、現在の状態IDが閉じていることを示すことができる。1つまたは複数のバックアップマネージャのそれぞれは、ワーカーのいずれかから受信し

10

20

30

40

50

た任意の新しいワーク要求を転送することもできる。

【0018】

付随する特徴の多くは、添付図面面に関連して考慮される以下の詳細な説明を参照することによって、より容易に認識されると同時に、より理解されるようになるだろう。

【図面の簡単な説明】

【0019】

本説明は、添付図面面を踏まえて読まれる以下の詳細な説明によってより理解されるだろう。

【図1】従来の分散型サーバーシステムを示すブロック図である。

【図2】分散型サーバーシステムにおいてバックアップサーバーを転送する例示的なメッセージを示すブロック図である。

【図3】図2に見られる分散型サーバーシステムの状態と関連付けられたIDを閉じることと関連付けられた例示的な通信を示す図である。

【図4】主マネージャが低速化している、または応答しないバックアップマネージャを処理する例示的な方法を示す図である。

【図5】バックアップマネージャが低速化している、または応答しないワーカーを処理する例示的な方法を示す図である。

【図6】説明したシステムおよび方法を実施する例示的なコンピューター装置を示す図である。

【0020】

同種の参照数字は、添付図面面の同種の部分を指定するのに使用される。

【発明を実施するための形態】

【0021】

添付図面に関連して以下に与えた詳細な説明は、本実施例の説明を意図しており、本実施例を構成または利用できる唯一の形態を表すことを意図していない。本説明は、本実施例の機能および本実施例を構成し、動作するステップの順序を示す。しかし、同一のもしくは同等の機能および順序を、異なる実施例によって達成することができる。

【0022】

本実施例を、分散型サーバーシステムにおいてバックアップマネージャを転送するメッセージを実施するものとして本明細書に説明し、示すが、説明するシステムは、例示として与えられており、これに限定されない。当業者が理解するように、本実施例は、さまざまな異なる種類の動的なフラグメントマッピングシステムのアプリケーションに適している。

【0023】

図1は、従来の分散型サーバーシステム100を示すブロック図である。従来の分散型サーバーシステム100は、従来型セントラルサーバー120、従来型バックアップサーバー140、従来型バックアップサーバー150、従来型サーバー160、従来型サーバー170、および従来型サーバー180を含むことができる。

【0024】

例示的な従来の分散型サーバーシステム100において、従来型バックアップサーバー140と従来型バックアップサーバー150のそれぞれを、従来型セントラルサーバー120に通信接続することができる。従来型サーバー160、従来型サーバー170、および従来型サーバー180のそれぞれを、従来型セントラルサーバー120に通信接続することもできる。そのような従来の分散型サーバーシステム100は、本来論理的であって、つまりそのような通信接続は物理的に存在しないので、その代わりに、そのような通信接続は、従来の分散型サーバーシステム100において参加者間で通信する順序を示すことができる。

【0025】

従来の分散型サーバーシステム100は、任意の種類の従来の分散型サーバーアーキテクチャとすることができる。例えば、従来の分散型サーバーシステム100は、従来の分

10

20

30

40

50

散型データベースサーバーシステム、従来の分散型媒体サーバーシステム等とすることができる。従来型セントラルサーバー１２０は、ソフトウェアサービスを他のコンピューターシステムに提供するように構成された任意の種類の従来型サーバーコンピューターシステムとすることができる。従来型セントラルサーバー１２０は、外部エンティティからの要求を受信することができる。例えば、エラー！参照ソースが見つからない（Error！Reference source not found.）等で、別のサーバーは、従来型セントラルサーバー１２０と同じネットワークドメインにおいて実行するワールドワイドウェブサーバー、またはインターネットを経由して通信接続された外部クライアント等とすることができる。

#### 【００２６】

10

従来型セントラルサーバー１２０は、要求を任意の従来型サーバー１６０、従来型サーバー１７０、および／または従来型サーバー１８０に送信し、タスクを実行することができる。従来型サーバー１６０、従来型サーバー１７０、および／または従来型サーバー１８０のそれぞれは、タスクを実行し、従来型セントラルサーバー１２０に応答してタスクの完了を示すことができる。さらに、従来型セントラルサーバー１２０は、バックアップコマンド、または冗長的に情報を格納するコマンドを任意の従来型バックアップサーバー１４０および／または従来型バックアップサーバー１５０に送信することができる。

#### 【００２７】

任意の数の従来型サーバーを従来の分散型サーバーシステム１００に追加し、従来の分散型サーバーシステム１００の全体のタスク実行容量を増加することができる。より多くの従来型サーバーが従来型セントラルサーバー１２０に通信接続されるので、従来型セントラルサーバー１２０によって送受信される通信量は、従来型セントラルサーバー１２０に通信の低速化が起こる程度まで増加する可能性がある。

20

#### 【００２８】

より多くの従来型サーバーを、付随する通信ボトルネックなしで従来型サーバーシステムに追加することができるシステムが有効である。

#### 【００２９】

図２は、分散型サーバーシステム２００においてバックアップサーバーを転送する例示的なメッセージを示すブロック図である。分散型サーバーシステム２００は、主マネージャ２１０、バックアップマネージャ２２０、バックアップマネージャ２３０、ワーカー２４０、ワーカー２５０、およびワーカー２６０を含むことができる。

30

#### 【００３０】

分散型サーバーシステム２００において、主マネージャ２１０を、バックアップマネージャ２２０およびバックアップマネージャ２３０のそれぞれに通信接続することができる。バックアップマネージャ２２０を、ワーカー２４０およびワーカー２５０のそれぞれに通信接続することができる。バックアップマネージャ２３０を、ワーカー２６０に通信接続することができる。このような分散型サーバーシステム２００は、ほんの一例である。任意の数のバックアップマネージャを、主マネージャ２１０に通信接続することができる。同様に、任意の数のワーカーを、任意の数のバックアップマネージャに通信接続することができる。例えば、バックアップサーバーの数は、エージェントサーバーの数の平方根とすることができる。そのような分散型サーバーシステム２００も、本来論理的であって、つまりそのような通信接続は物理的に存在しないので、その代わりに、そのような通信接続は、分散型サーバーシステム２００において参加者間で通信する順序を示すことができる。

40

#### 【００３１】

通信接続のそれぞれを、WAN(wide area network)、LAN(local area network)等のコンピューターネットワークを使用して行うことができる。代替的な実施例において、主マネージャ、バックアップマネージャ、およびワーカーのそれぞれは、単一のコンピューターシステム上で実行するソフトウェアサービスとすることができ、それらの通信接続のそれぞれを、マイクロソフトWindows（登録商標）等のオペレーティングシステ

50



ムに含まれるソフトウェア通信サービスを使用して行うことができる。

【0032】

主マネージャ210は、サーバークラスオペレーションシステムを実行するサーバコンピュータシステムおよびそれに関連するソフトウェアサービスとすることができる。主マネージャ210はさらに、分散型サーバシステム200の「状態」と関連付けられた1つまたは複数のIDを作成し、管理するためのソフトウェアサービスを実行することができる。そのような状態は、分散型サーバシステム200の代わりに主マネージャ210によって管理し、維持される情報のディスクリットスナップショット、または一時的な表示を示すことができる。つまり状態は、システムの任意の時点における任意の数のデータ配列、命令等を表すことができる。このように主マネージャ210は、特定の時点における分散型サーバシステム200の状態を一意的に識別することができる状態に対応するIDを割り当てることができる。

10

【0033】

主マネージャ210を、ワーク「タスク」を管理し、分散するよう構成することもできる。ワーク「タスク」は、コンピュータシステムによって実行することができる任意の種類の機能性としてすることができる。例えば「ワーク」は、数値計算、データもしくは他の情報を格納するコマンド、または任意の他の種類のコンピュータソフトウェア機能としてすることができる。主マネージャ210はさらに、ワークを状態IDと関連付けることができる。特に、主マネージャ210は、現在の状態および現在の状態IDと実行されたワーク総量とを結び付けることができる。より詳細には、分散型サーバシステム200の現在の状態は、分散型サーバシステム200の状態が初期化されてから実行された「ワーク」のそれぞれのディスクリット装置の累積集計に相当することができる。

20

【0034】

主マネージャ210は、要求をバックアップマネージャ220およびバックアップマネージャ230のそれぞれに送信し、状態IDと関連付けられた状態を持続することができる。バックアップマネージャ220およびバックアップマネージャ230のそれぞれを、各状態IDおよび主マネージャ210から受信した状態IDと関連付けられた状態を冗長的に格納するよう構成することができる。この方法において、バックアップマネージャ220またはバックアップマネージャ230のいずれかは置き換えることができ、主マネージャ210の機能は、主マネージャ210に主マネージャ210が機能停止を引き起こす障害が起こることを想定することができる。さらに、バックアップマネージャ220およびバックアップマネージャ230のそれぞれが、状態を持続している主マネージャ210に伝えることができるようになった時点で、主マネージャ210は、コマンドをバックアップマネージャ220およびバックアップマネージャ230のそれぞれに伝え、現在の状態IDと関連付けられたワークをワーカー240、ワーカー250、および/またはワーカー260のそれぞれに送信することができる。

30

【0035】

ワーカー240、ワーカー250、およびワーカー260のそれぞれは、任意のバックアップマネージャ220および/またはバックアップマネージャ230から受信したワークタスクと関連付けられた機能性を実行することができるコンピュータシステムとすることができる。さらに、ワーカー240、ワーカー250、およびワーカー260のそれぞれを、通信接続することもでき、情報、データ等を互いに共有することができる。ワーカー240、ワーカー250、およびワーカー260のそれぞれが割り当てられたワークタスクを完了した時点で、それぞれのワーカーは、それらの情報をバックアップマネージャ220、またはバックアップマネージャ230のいずれかに返信することができる。

40

【0036】

バックアップマネージャ220およびバックアップマネージャ230のそれぞれは、ワーカー240、ワーカー250またはワーカー260のいずれかから受信した応答を集めることができる。バックアップマネージャ220およびバックアップマネージャ230のそれぞれは次に、集めた応答を主マネージャ210に返信することができる。ワーカー2

50

40、ワーカー250、ワーカー260のいずれかの応答が低速化しているまたは応答しないイベントにおいて、バックアップマネージャ220またはバックアップマネージャ230のそれぞれは、集められた応答内に応答が低速化しているまたは機能していないワーカーの識別を表示する情報を含むことができる。

【0037】

上述のように、バックアップマネージャ220およびバックアップマネージャ230のそれぞれは常に、分散型サーバーシステム200の状態をバックアップまたは冗長的に格納し、主マネージャ210の役割を担うことができる。従って、分散型サーバーシステム200の構成は、動的とすることができ、それらの通信接続は、分散型サーバーシステム200に含まれる低速化しているもしくは機能しないマネージャまたはワーカーを反映するように変更することができる。つまり、主マネージャ210とバックアップマネージャ220とバックアップマネージャ230との間の通信接続によって、バックアップマネージャ220、またはバックアップマネージャ230のいずれかが主マネージャ210の機能的責任を担うことができるように再構成することができる。

【0038】

上述のように、主マネージャ210は、集めた応答を分析し、ワーカーが低速化しているまたは応答していないことを判定することができる。主マネージャ210は次に、対応するバックアップマネージャに通知し、分散型サーバーシステム200から応答が低速化しているまたは機能していないワーカーを除去することができる。主マネージャ210は次に、新しい送信メッセージを残ったワーカーのそれぞれに送り、状態と関連付けられた全てのワークが完了できるように、低速化しているまたは機能していないワーカーに割り当てられたワークを実行することができる。

【0039】

説明した機能性は、任意の数のサーバーコンピューターシステム上で実施することができる。例えば、1コンピューターシステムは、セントラルマネージャ、バックアップマネージャ、ワーカー、またはそれらの任意の組み合わせに関連しているソフトウェアサービスを実行することができる。図3の、エラー！参照ソースが見つからない（Error！Reference source not found.）機能は、図2に見られる主マネージャ210の状態と関連付けられたIDを閉じることと関連付けられた例示的な通信を示す。

【0040】

図2の主マネージャ210は、コマンドを図2の任意のバックアップマネージャ220または図2のバックアップマネージャ230に送信することができる。図2について述べたように、そのようなコマンドは、状態IDと関連付けられた状態を閉じる命令を含むことができる。そのようなコマンドはさらに、現在の状態IDと関連付けられた状態を持続するコマンドを含むことができる。

【0041】

バックアップマネージャ220およびバックアップマネージャ230のそれぞれは次に、主マネージャ210によって要求された状態を持続することができる。バックアップマネージャ220およびバックアップマネージャ230のそれぞれが現在の状態IDと関連付けられた状態を持続した時点で、バックアップマネージャ220およびバックアップマネージャ230のそれぞれは、現在の状態IDと関連付けられた状態が持続していることを表示する応答320を送信することができる。主マネージャ210は、次に進む前にバックアップマネージャのそれぞれから受信すべき全ての応答を待つことができる。

【0042】

主マネージャ210は次に、コマンド330をバックアップマネージャ220およびバックアップマネージャ230のそれぞれに送信し、バックアップマネージャ220およびバックアップマネージャ230のそれぞれが、ワーカーのそれぞれに対する状態IDと関連付けられた状態に対応するワークを現在送信することができることを示す。バックアップマネージャ220およびバックアップマネージャ230のそれぞれは次に、コマンド3

40を図2のワーカー240、図2のワーカー250、および図2のワーカー260のそれぞれに送信し、現在の状態IDと関連付けられた状態を閉じ、ワーカーに割り当てられた現在の状態IDと関連付けられたワークを実行することができる。それぞれのワーカーに送信されたコマンド340は、同一とすることができ、各ワーカーは、コマンド340の一部として送信されたワークが割り当てられていない部分を無視することができる。代替的な実施例において、コマンド340は、それぞれのワーカーに対して個々に作成することができ、コマンド340が送信されたワーカーに割り当てられたワークのみを含むことができる。

【0043】

ワーカー240、ワーカー250、およびワーカー260のそれぞれがワークを完了した時点で、それぞれのワーカーは、現在の状態IDと関連付けられた状態を局所的に閉じることができる。ワーカー240、ワーカー250、およびワーカー260のそれぞれは次に、実行すべき任意の追加のワークを決定することができる。そのような決定は、ワーカー240、ワーカー250、およびワーカー260のそれぞれが互いに通信する時にを行うことができる。代替的な実施例において、ワーカー240、ワーカー250、およびワーカー260のそれぞれは、独立して、現在の状態IDと関連付けられた状態を分析し、実行すべき任意の追加のワークを決定することができる。ワーカー240、ワーカー250、およびワーカー260のそれぞれは、メッセージ350をそれらに割り当てられたバックアップマネージャ220またはバックアップマネージャ230のいずれかのバックアップマネージャに送信することができ、ワーカー240、ワーカー250、およびワーカー260のそれぞれが、現在の状態IDと関連付けられた状態を閉じたことを示す。そのようなメッセージは、実行すべき任意の追加の新しいワークに対する要求を含むこともできる。

【0044】

バックアップマネージャ220およびバックアップマネージャ230のそれぞれは次に、メッセージ360を主マネージャ210に送信することができ、現在の状態IDと関連付けられた状態が閉じられたことを示す。また、メッセージ360は、ワーカーから受信した時に実行すべき任意の新しいワークを含むことができる。

【0045】

図4に、低速化しているまたは応答しないバックアップマネージャを処理する主マネージャの例示的な方法400を示す。そのような主マネージャは、図2の主マネージャ210とすることができる。

【0046】

ブロック410では、現在の状態IDと関連付けられた状態を閉じるコマンドを1つまたは複数のバックアップマネージャに送信する動作を示すことができる。そのようなバックアップマネージャは、図2のバックアップマネージャ220および図2のバックアップマネージャ230とすることができる。そのような現在の状態IDを閉じるコマンドは、図3のコマンド310または図3のコマンド330とすることができる。

【0047】

ブロック420では、既定の時間間隔内に応答が受信されたかどうかを判定する動作を示すことができる。そのような時間間隔は、任意の長さとするすることができる。あるいは、そのような時間間隔を、分散型サーバーシステムの予想される実行の全レベルと関連付けることができる。肯定的な判定に応じて、フローはブロック430に続く。否定的な判定に応じて、フローはブロック440に続く。

【0048】

ブロック430では、主マネージャの標準動作を継続する動作を示すことができる。標準動作は、主マネージャの機能に従って実行される任意の動作を示すことができる。

【0049】

ブロック440では、応答の最小数が受信されたかどうかを判定する動作を示すことができる。最小数は、分散型サーバーシステムに含まれるバックアップマネージャの全数、

10

20

30

40

50

バックアップマネージャに接続されたワーカーの数等を含む任意の情報に基づいて判定することができる。肯定的な判定に応じて、フローはブロック 4 5 0 に続く。否定的な判定に応じて、フローはブロック 4 6 0 に続く。

【 0 0 5 0 】

ブロック 4 5 0 では、低速化しているまたは応答しないバックアップマネージャを回避する動作を示すことができる。そのような回避を、低速化しているまたは応答しないバックアップマネージャに割り当てられた任意のワーカーを、応答するバックアップマネージャに一時的に再度割り当てることによって実行することができる。

【 0 0 5 1 】

ブロック 4 6 0 では、バックアップマネージャの数が少なすぎて既定の実行レベルを分散型データベースシステム内に確保できないこと、または分散型データベースシステムに対するバックアップマネージャの数が少なすぎて正確に機能できないことを判定された動作を示すことができる。その動作は、一定時間待機している動作を含むこともできる。追加のバックアップマネージャが待機時間内に応答しなければ、分散型データベースシステムは、例えば追加の応答するバックアップマネージャが追加される時間まで機能を停止することができる。フローは 4 2 0 に戻って継続することができる。

【 0 0 5 2 】

図 5 に低速化しているまたは応答しないワーカーを処理するバックアップマネージャの例示的な方法 5 0 0 を示す。そのようなバックアップマネージャは、図 2 のバックアップマネージャ 2 2 0 または図 2 のバックアップマネージャ 2 3 0 とすることができる。そのようなワーカーは、バックアップマネージャ 2 3 0 とすることができ、ブロック 5 1 0 では、状態に対応するワークを実行するコマンドを 1 つまたは複数のワーカーに送信する動作を示すことができる。代替的な実施形態において、コマンドが送信され、ワーカーが現在の状態 ID と関連付けられた状態を閉じなければならないことを示す。しかし、その動作は、ワーカーからの応答を要求することができる任意のコマンドを含むことができる。

【 0 0 5 3 】

ブロック 5 2 0 では、ブロック 5 1 0 のコマンドが送信された各ワーカーが既定の時間間隔内に応答しているかどうかを判定する動作を示すことができる。そのような時間間隔は、任意の時間間隔とすることができ、分散型サーバーシステムと関連付けられた実行の全レベルと関連付けることができる。肯定的な判定に応じて、フローはブロック 5 3 0 に続く。否定的な判定に応じて、フローはブロック 5 4 0 に続く。

【 0 0 5 4 】

ブロック 5 3 0 では、バックアップマネージャが全てのワーカーの応答を集めて単一メッセージを形成し、その集められたメッセージをセントラルマネージャに送信する動作を示すことができる。

【 0 0 5 5 】

ブロック 5 4 0 では、バックアップマネージャが全てのワーカーの応答を集めて単一メッセージを形成し、その集められたメッセージをセントラルサーバーに送信する動作を示すことができる。集められたメッセージは、ブロック 5 2 0 において識別された任意の低速化しているまたは機能しないワーカーの身元を示す情報を含むことができる。

【 0 0 5 6 】

図 6 に、任意の前述したサーバーコンピューターシステムまたはクライアントコンピューターシステム等の説明したシステム、方法、およびソフトウェアサービスを実施する例示的なコンピューター装置 6 0 0 を示す。それらの最も基本的な構成において、コンピューター装置 6 0 0 は典型的に、少なくとも 1 つの CPU (central processing unit) およびメモリー 6 1 0 を含む。

【 0 0 5 7 】

正確な構成およびコンピューター装置の種類に依存する場合、メモリー 6 1 0 は、揮発性 (RAM 等)、不揮発性 (ROM、フラッシュメモリー等)、またはそれら 2 つのある組み合わせとすることができる。さらに、コンピューター装置 6 0 0 は、追加の機能 / 機

10

20

30

40

50

能性を有することもできる。例えば、コンピューター装置 600 は、複数の CPU を含むことができる。説明した方法を、コンピューター装置 600 の任意の処理装置による任意の方法で実行することができる。例えば、説明したプロセスを、同時に複数の CPU によって実行することができる。

#### 【0058】

コンピューター装置 600 は、磁気もしくは光ディスクまたは磁気もしくは光テープを含む付加的な記憶装置（取り外し可能および／または取り外し不可能）を含むこともできるがこれらに限定されない。そのような付加的な記憶装置を、図 6 の記憶装置 615 に示す。コンピューター記憶媒体は、コンピューター読み取り可能な命令、データ構造、プログラムモジュールもしくは他のデータ等の情報を格納する任意の方法または技術において実施される揮発性および不揮発性媒体、取り外し可能および取り外し不可能媒体を含む。メモリー 610 および記憶装置 615 は、コンピューター記憶媒体の全ての例である。コンピューター記憶媒体は、RAM、ROM、EEPROM、フラッシュメモリーもしくは他のメモリー技術、CD-ROM、DVD(digital versatile disk)もしくは他の光記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置もしくは他の磁気記憶装置、または所望の情報を記憶するために使用でき、コンピューター装置 600 によってアクセスすることができる任意の他の媒体を含むが、これらに限定されない。そのような任意のコンピューター記憶媒体は、コンピューター装置 600 の一部とすることができる。

#### 【0059】

コンピューター装置 600 は、通信装置 640 を含むこともでき、当該装置が他の装置と通信することを可能とする。通信装置 640 は、通信媒体の一例である。通信媒体は、典型的には、コンピューター読み取り可能な命令、データ構造、プログラムモジュール、または他のデータを、搬送波または他の転送機構等の変調データ信号として具現化したものであり、また、任意の情報配信媒体を含む。用語「変調データ信号」は、1 つまたは複数のそれらの特徴セットを有するか、または情報を信号にエンコードするような方法で変更された信号を意味する。一例として通信媒体には、有線ネットワークまたは直接有線接続、および音響、RF、赤外線ならびに他の無線媒体等の無線媒体を含むが、これらに限定されない。本明細書で使用されている用語コンピューター読み取り可能な媒体または装置読み取り可能な媒体は、コンピューター記憶媒体および通信媒体の両方を含む。説明した方法を、データ、コンピューター実行可能な命令等の任意の形態の任意のコンピューター読み取り可能な媒体にエンコードすることができる。

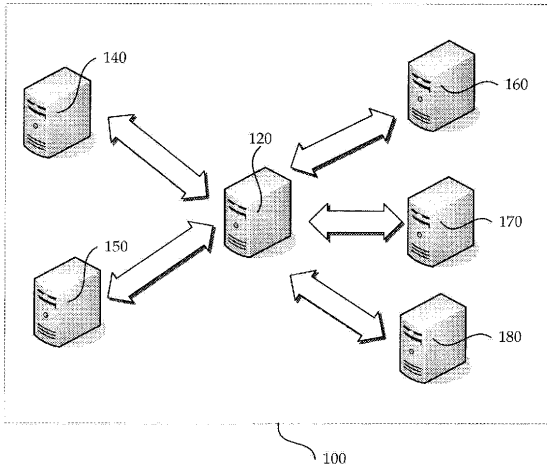
#### 【0060】

コンピューター装置 600 は、キーボード、マウス、ペン、音声入力装置、タッチ入力装置等の入力装置 635 を有することもできる。ディスプレイ、スピーカー、プリンター等の出力装置 630 も含むことができる。これら全ての装置は、当技術分野で周知であり、詳細に説明する必要はない。

#### 【0061】

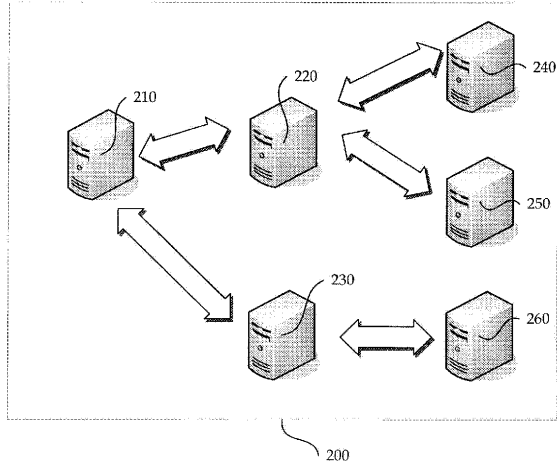
当業者は、プログラム命令を格納するのに利用される記憶装置を、ネットワーク上に分散することができることを認識するだろう。例えばリモートコンピューターは、ソフトウェアとして説明した例示的なプロセスを格納することができる。ローカルコンピューターまたはターミナルコンピューターは、リモートコンピューターにアクセスし、ソフトウェアの一部または全てをダウンロードしてプログラムを実行することができる。あるいはローカルコンピューターは、必要に応じてソフトウェアの一部をダウンロードすることができる、またはローカル端末におけるいくつかのソフトウェアの命令およびリモートコンピューター（またはコンピューターネットワーク）におけるいくつかのソフトウェアの命令を実行することによって、分散して処理することができる。当業者は、当業者に周知である従来を用いて、ソフトウェアの命令の全てまたは一部を、DSP、プログラム可能な論理アレイ等の専用回線で実行することができることも認識するだろう。

【図 1】



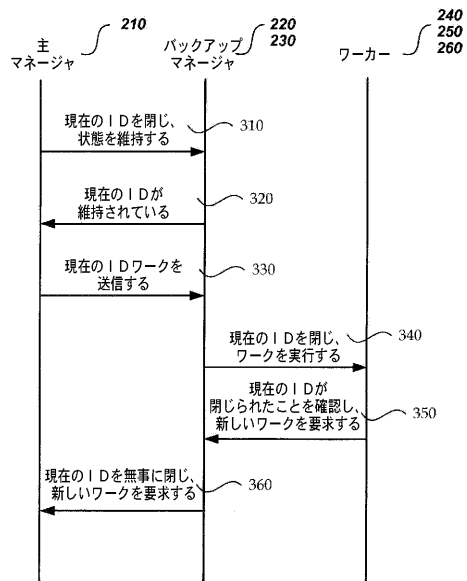
従来技術

【図 2】



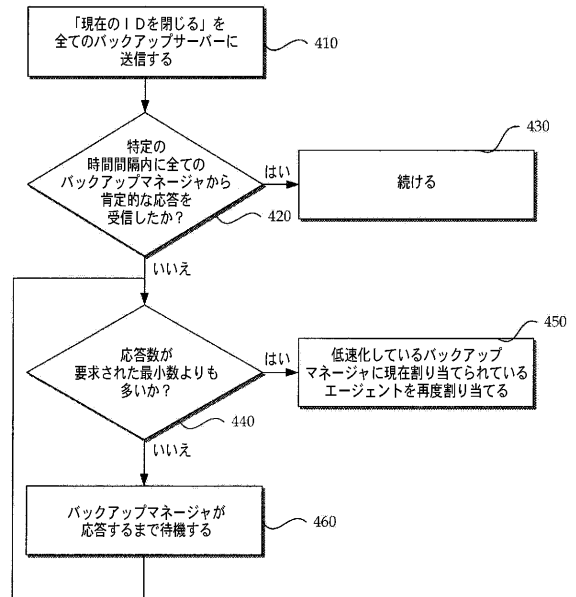
【図 3】

300



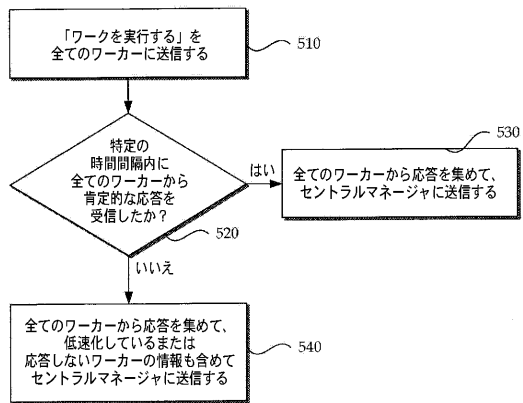
【図 4】

400

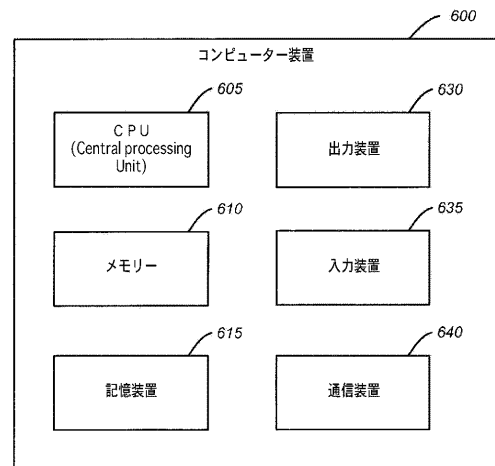


【図 5】



500



【図 6】



## 【 国際調査報告 】

|  |   |   |
|--|---|---|
| <b>INTERNATIONAL SEARCH REPORT</b>   |   | International application No.<br><b>PCT/US2007/077114</b>   |
| <b>A. CLASSIFICATION OF SUBJECT MATTER</b>   |   |   |
| <b>G06F 15/16(2006.01)i</b>  |   |   |
| According to International Patent Classification (IPC) or to both national classification and IPC  |   |   |
| <b>B. FIELDS SEARCHED</b>  |   |   |
| Minimum documentation searched (classification system followed by classification symbols)<br>IPC8 G06F 11/14, 13/00, 15/16, 17/30, H04L 12/40  |   |   |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched<br>Korean Utility models and applications for Utility Models since 1975<br>Japanese Utility models and applications for Utility Models since 1975  |   |   |
| Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br>eKIPASS(KIPO internal) "distributed computer system, server, back up, and similar terms"   |   |   |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>  |   |   |
| Category*  | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No.   |
| A  | US 6260269 B1 (MATHEW JOSEPH ANGLIN) 10 July 2001<br>See Summary of the invention, Column 3 Line 17 - Column 5 Line 8                 | 1 - 20  |
| A  | US 6026414 A (MATHEW JOSEPH ANGLIN) 15 Feb. 2000<br>See Summary of the invention, Column 5 Line 25 - Column 6 Line 27                 | 1 - 20  |
| A  | US 6611923 B1 (MADHAV MUTALIK et al.) 26 Aug. 2003<br>See Summary of the invention, Figures 1, 2, Column 3 Line 52 - Column 6 Line 37 | 1 - 20  |
| A  | JP 11353202 A2 (NTT MOBILE COMMUNICATION NETWORK INC.) 24 Dec. 1999<br>See Summary of the invention, Paragraph [0005] - [0010]        | 1 - 20  |
| A  | US 6496949 B1 (DIMITRY KANEVSKY) 17 Dec. 2002<br>See Summary of the invention, Column 2 Line 13 - Column 2 Line 65                    | 1 - 20  |
| <input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.   |   |   |
| * Special categories of cited documents:<br>"A" document defining the general state of the art which is not considered to be of particular relevance<br>"E" earlier application or patent but published on or after the international filing date<br>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)<br>"O" document referring to an oral disclosure, use, exhibition or other means<br>"P" document published prior to the international filing date but later than the priority date claimed<br>"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<br>"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<br>"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<br>"&" document member of the same patent family |   |   |
| Date of the actual completion of the international search<br>11 FEBRUARY 2008 (11.02.2008)   |   | Date of mailing of the international search report<br><b>11 FEBRUARY 2008 (11.02.2008)</b>  |
| Name and mailing address of the ISA/KR<br> Korean Intellectual Property Office<br>Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea<br>Facsimile No. 82-42-472-7140   |   | Authorized officer<br>JEONG, Jae Woo<br>Telephone No. 82-42-481-5718<br> |



**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

**PCT/US2007/077114**

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|---|---------------------|----------------------------|---------------------|
| US 6260269 B1                             | 10.07.2001          | None                       |                     |
| US 6026414 A                              | 15.02.2000          | None                       |                     |
| US 6611923 B1                             | 26.08.2003          | US 6360330 B1              | 31.03.2002          |
| JP 11353202 A2                            | 24.12.1999          | None                       |                     |
| US 6496949 B1                             | 17.12.2002          | None                       |                     |

## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(72)発明者 トーマス ジェイ・ミラー

アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ  
マイクロソフト コーポレーション インターナショナル パテント内

Fターム(参考) 5B082 DE06 HA08