(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4900982号
(P4900982)

(45)発行日　平成24年3月21日(2012.3.21)　(24)登録日　平成24年1月13日(2012.1.13)

(51)Int.Cl.　F I
G06F 11/20 (2006.01)　G06F 11/20 310C
G06F 9/50 (2006.01)　G06F 9/46 465Z

請求項の数 15　(全 14 頁)

(21)出願番号　特願2010-528436 (P2010-528436)
(86)(22)出願日　平成20年10月15日 (2008.10.15)
(65)公表番号　特表2011-501254 (P2011-501254A)
(43)公表日　平成23年1月6日 (2011.1.6)
(86)国際出願番号　PCT/EP2008/063850
(87)国際公開番号　WO2009/050187
(87)国際公開日　平成21年4月23日 (2009.4.23)
審査請求日　平成23年8月4日 (2011.8.4)
(31)優先権主張番号　11/872,235
(32)優先日　平成19年10月15日 (2007.10.15)
(33)優先権主張国　米国 (US)

早期審査対象出願

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
(74)代理人　100108501
弁理士　上野　剛史
(74)代理人　100112690
弁理士　太佐　種一
(74)代理人　100091568
弁理士　市位　嘉宏

最終頁に続く

(54)【発明の名称】サーバ・クラスタにおいてフェイルオーバを管理するための方法、フェイルオーバ・サーバ及びコンピュータ・プログラム

(57)【特許請求の範囲】
【請求項１】
サーバ・クラスタにおいてフェイルオーバを管理するための方法であって、
前記サーバ・クラスタのうちの一つのサーバが分散型ネットワーク内の前記サーバ・クラスタにおける障害のあるサーバ（以下、障害サーバという）を検出したことに応答して、前記障害サーバを検出した前記サーバ（以下、フェイルオーバ・サーバという）が、当該フェイルオーバ・サーバのサブスクリプション・キューについてのサブスクリプション・メッセージ処理を停止するステップと、
前記フェイルオーバ・サーバが、前記障害サーバのサブスクリプション・キューを開くステップと、
前記フェイルオーバ・サーバが、前記障害サーバの前記サブスクリプション・キューを現在管理しているフェイルオーバ・サーバの識別子を含むマーカ・メッセージを、特定のメッセージング・トピックをサブスクライブしている全てのサブスクライバにパブリッシュするステップであって、前記特定のメッセージング・トピックは、サブスクライバに対して関心のある主題を示す、前記パブリッシュするステップと、
前記フェイルオーバ・サーバが、前記障害サーバのサブスクリプション・キュー内のメッセージを処理するステップと、
前記フェイルオーバ・サーバが、前記障害サーバの前記サブスクリプション・キューにおけるメッセージが前記マーカ・メッセージであるかどうかを判断するステップと、
前記フェイルオーバ・サーバが、前記障害サーバの前記サブスクリプション・キューに

おけるメッセージが前記マーカ・メッセージであることに応答して、前記障害サーバの前記サブスクリプション・キューを閉じるステップと、
前記フェイルオーバ・サーバが、当該フェイルオーバ・サーバの前記サブスクリプション・キューについての前記サブスクリプション・メッセージ処理を再開するステップと
を含む、前記方法。
【請求項２】
前記フェイルオーバ・サーバが、前記障害サーバの前記サブスクリプション・キューにおけるメッセージが前記マーカ・メッセージではないことに応答して、前記メッセージを読み取り済みとして記録するステップと、
前記フェイルオーバ・サーバが、前記障害サーバに対する関連のサブスクリプション・セッションにおける処理を遂行するステップと
を更に含む、請求項１に記載の方法。
【請求項３】
前記フェイルオーバ・サーバが、当該フェイルオーバ・サーバの前記サブスクリプション・キューにおけるメッセージが前記マーカ・メッセージであるかどうかを判断するステップと、
前記フェイルオーバ・サーバが、当該フェイルオーバ・サーバの前記サブスクリプション・キューにおけるメッセージが前記マーカ・メッセージであることに応答して、正規のオペレーションを再開するステップと
を更に含む、請求項１又は２に記載の方法。
【請求項４】
前記フェイルオーバ・サーバが、当該フェイルオーバ・サーバの前記サブスクリプション・キューにおけるメッセージがマーカ・メッセージではなく且つ前記メッセージが読み取られてないことに応答して、前記フェイルオーバ・サーバに対する関連のサブスクリプション・セッションにおける処理を遂行するステップを更に含む、請求項１～３のいずれか一項に記載の方法。
【請求項５】
前記フェイルオーバ・サーバが、サーバの起動に応答して、前記起動したサーバに対する一意的なサブスクリプション識別子を生成するステップと、
前記フェイルオーバ・サーバが、新しいセッションを作成したことに応答して、前記一意的なサブスクリプション識別子を前記新しいセッションに格納するステップと
を更に含み、
前記一意的なサブスクリプション識別子は、前記新しいセッションのライフタイムの間前記新しいセッションにおいて継続する、請求項１～４のいずれか一項に記載の方法。
【請求項６】
前記一意的なサブスクリプション識別子は、メッセージング・トピックに接続するために使用される、請求項５に記載の方法。
【請求項７】
前記分散型ネットワークが、セッション・アフィニティを利用する分散型パブリッシュ・サブスクライブ・ネットワークである、請求項１～６のいずれか一項に記載の方法。
【請求項８】
前記マーカ・メッセージが、前記障害サーバの前記サブスクリプション・キュー及び前記フェイルオーバ・サーバの前記サブスクリプション・キューにおいて同時に現れる、請求項１～７のいずれか一項に記載の方法。
【請求項９】
前記マーカ・メッセージが、前記フェイルオーバ・サーバが前記正規のオペレーションを再開する前に、前記障害サーバの前記サブスクリプション・キュー及び前記フェイルオーバ・サーバの前記サブスクリプション・キューを同期させるために使用される、請求項３に記載の方法。
【請求項１０】

前記マーカ・メッセージが、前記障害サーバの前記サブスクリプション・キュー及び前記フェイルオーバ・サーバの前記サブスクリプション・キューを同期させるために使用される、請求項１～８のいずれか一項に記載の方法。

【請求項１１】

前記マーカ・メッセージが、フェイルオーバに参加していない他のすべてのサーバによって無視される、請求項１～１０のいずれか一項に記載の方法。

【請求項１２】

サーバ・クラスタにおいてフェイルオーバを管理するためのフェイルオーバ・サーバであって、

バス・システムと、

前記バス・システムに接続され、命令のセットを含む記憶装置と、

前記バス・システムに接続された処理ユニットと

を含み、

前記処理ユニットに、請求項１～１１のいずれか一項に記載の方法の各ステップを実行させる、前記フェイルオーバ・サーバ。

【請求項１３】

サーバ・クラスタにおいてフェイルオーバを管理するためのサーバであって、

バス・システムと、

前記バス・システムに接続され、命令のセットを含む記憶装置と、

前記バス・システムに接続された処理ユニットと

を含み、当該サーバが、分散型ネットワーク内の前記サーバ・クラスタにおける障害のあるサーバ（以下、障害サーバという）を検出したことに応答して、

前記処理ユニットに、請求項１～１１のいずれか一項に記載の方法の各ステップを実行させる、前記サーバ。

【請求項１４】

コンピュータ・プログラムであって、フェイルオーバ・サーバに、請求項１～１１のいずれか一項に記載の方法の各ステップを実行させるためのコンピュータ・プログラム。

【請求項１５】

コンピュータ・プログラムであって、分散型ネットワーク内の前記サーバ・クラスタにおける障害のあるサーバ（以下、障害サーバという）を検出したことに応答して、前記障害サーバを検出したサーバ（以下、フェイルオーバ・サーバという）に、請求項１～１１のいずれか一項に記載の方法の各ステップを実行させるためのコンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、データ処理システムの改良に関する。具体的にいえば、本発明は、セッション・アフィニティを利用する分散型ネットワーク環境において、サーバ・フェイルオーバを処理するためにコンピュータを利用する方法、システム、及びコンピュータ・プログラムに関する。

【背景技術】

【０００２】

今日、ほとんどのコンピュータが何らかのタイプのネットワークに接続される。ネットワークは、コンピュータが他のコンピュータ・システムと情報を共用することを可能にする。インターネットは、コンピュータ・ネットワークの一例である。インターネットは、コンピュータのグローバル・ネットワークであり、データ転送を処理し且つ、送信ネットワークのプロトコルから受信ネットワークにより使用されるプロトコルへのメッセージの変換を処理する、ゲートウェイによって結合されたネットワークである。インターネットでは、任意のコンピュータが、プロトコルとも呼ばれる種々の言語を通して情報がインターネット上を移動することによって、他のコンピュータと通信を行なうことを可能にする

。一般に、インターネットは、伝送制御プロトコル／インターネット・プロトコル（ＴＣＰ／ＩＰ）と呼ばれる一連のプロトコルを使用する。

【０００３】

多くの最新のインターネット・アプリケーションは、異なる組織上の境界、異機種のプラットフォーム、並びにパブリッシャ（即ち、発行元）及びサブスクライバ（即ち、引用先）の大きな動的集団の全体にわたって情報を配布することを必要とする。パブリッシュ・サブスクライブ（パブ・サブ）・ネットワーク・サービスは、潜在的に無制限の数のパブリッシャおよびサブスクライバの全体にわたる情報の配布を可能にする通信機能である。パブ・サブ・システムは、多くの場合、ピア・ツー・ピア・オーバレイ・ネットワークよりも上位の通信を行なう空間的にまったく異なるノードの集合体として実現されている。

【０００４】

そのような環境では、パブリッシャはイベントの形で情報をパブリッシュし、サブスクライバは、パブ・サブ・ネットワークにサブスクリプション・フィルタを送ることにより或るイベント又は或るパターンのイベントに対する関心事を表す能力を有する。パブ・サブ・ネットワークは、すべての活動中のサブスクリプションに各パブリケーションを動的に対抗させるためにコンテンツ・ベースの経路指定方法を使用し、イベントがそれらの登録済みの関心事と一致する場合、しかもその場合にのみ、そのイベントをサブスクライバに通知する。

【０００５】

集中サービス(converged service)は、高いレベルの機能を提供するために多数のネットワーク・プロトコル及びプロトコル・セッション上の通信にまたがるアプリケーションである。ハイパーテキスト転送プロトコル（ＨＴＴＰ）及びセッション開始プロトコル（ＳＩＰ）の場合、集中サービスは、ＨＴＴＰ及びＳＩＰの両方からのセッション情報を結合し、1つのプロトコルによる対話がそのプロトコルの制約次第で他のプロトコルによる通信に影響を及ぼすことがある。集中サービスは、これらのプロトコルの各々にまたがって多数のプロトコル・セッションに及ぶこともある。

【０００６】

コードの構造化および高い有効性サービスを簡単にするために、セッション・アフィニティと呼ばれるメカニズムが集中サービスと関連して使用される。セッション・アフィニティは、そのセッション内の要求をサーバ・クラスタ内の特定のサーバに関連付けるためのクラスタ構成された環境における機構である。この関連付けは、セッションを管理サーバにマップする経路指定メカニズムを介して達成される。集中サービスと共にセッション・アフィニティを使用するとき、集中セッション・データは、単一のアプリケーション・サーバの場合、セッションのライフタイム内では生きており、集中セッションに関する要求を処理するときにアプリケーション・コードがクラスタ間通信を行なう必要のないようにし得る。

【０００７】

しかし、多くの集中アプリケーションは、多数の集中セッション全体にわたって共通の資源又はデータ構造にアクセスすること及びそれを操作することを必要とする。セッション・アフィニティによっても、集中セッションは、クラスタ内の種々のサーバに割り当てられる。その結果、クラスタにおける関連するセッションの場所に関係なくそれらのセッションに関連する共通の情報の集中セッションを通知するための方法が必要である。例えば、サーバＡ、Ｂ、及びＣを含む３つのサーバがクラスタ構成された環境を考察する。サーバＡ及びサーバＣにおけるアプリケーション資源に関する通知のためにサブスクリプションが設定される。パブリッシュ要求が入ってきてサーバＢに与えられる。サーバＢは、クラスタにおけるどのサーバがその関心のあるサブスクリプション・セッションを含むかを知らない。サーバＢはサブスクリプション・データを高い信頼性でブロードキャストできなければならない。更に、サーバＢが障害を生じた場合、サーバＢによって管理されたセッションがそのクラスタ内のどこで再活性化されるかはわからない。

【発明の概要】
【発明が解決しようとする課題】
【０００８】
従って、本発明の目的は、セッション・アフィニティを利用するパブ・サブ分散型ネットワーク環境においてサーバ・フェイルオーバを管理するためにコンピュータを利用する改良された方法、システム、及びコンピュータ・プログラムを提供することにある。
【課題を解決するための手段】
【０００９】
本実施例は、サーバ・クラスタにおいてフェイルオーバを管理するためにコンピュータを利用する方法、システム、及びコンピュータ・プログラムを提供する。分散型ネットワーク内のサーバ・クラスタにおける障害のあるサーバ（以下、障害サーバという）を検出したことに応答して、フェイルオーバ・サーバのサブスクリプション・メッセージ処理が停止される。障害サーバのサブスクリプション・キューが開かれる。特定のメッセージング・トピックのすべてのサブスクライバに対して、マーカ・メッセージがパブリッシュされる。マーカ・メッセージは、障害サーバのサブスクリプション・キューを現在管理しているフェイルオーバ・サーバの識別子を含む。障害サーバのサブスクリプション・キュー内のメッセージが処理される。障害サーバのサブスクリプション・キューにおけるメッセージがマーカ・メッセージであるかどうかが判断される。障害サーバのサブスクリプション・キューにおけるメッセージがマーカ・メッセージであるという判断に応答して、障害サーバのサブスクリプション・キューが閉じられる。しかる後、フェイルオーバ・サーバが、当該フェイルオーバ・サーバのオリジナルのサブスクリプション・キューの処理を再開する。
【００１０】
オリジナルのサブスクリプション・キューからの未検出のメッセージを処理している間、マーカ・メッセージが探索されることが望ましい。一旦そのマーカ・メッセージがオリジナルのサブスクリプション・キューにおいて見つかると、正規のオペレーションが再開されることが望ましい。
【図面の簡単な説明】
【００１１】
【図１】本発明の実施例を具体化し得るデータ処理システムのネットワークの概略図である。
【図２】本発明の実施例を具体化し得るデータ処理システムのブロック図である。
【図３】本発明の実施例に従って通常のサーバ・オペレーションのための例示的プロセスを示すフローチャートである。
【図４】本発明の実施例に従ってフェイルオーバ・サーバ・オペレーションのための例示的プロセスを示すフローチャートである。
【発明を実施するための形態】
【００１２】
次に、図面、特に図１及び図２を参照すると、本実施例を具体化し得るデータ処理環境の例示的概略図が示される。図１及び図２は単に例示的なものであって、種々の実施例を具体化し得る環境に関して如何なる制限も主張及び暗示することを意図するものではないということは明らかであろう。図示の環境に対して多くの修正を施すことが可能である。
【００１３】
図１は、本実施例を具体化し得るデータ処理システムのネットワークの概略図を示す。ネットワーク・データ処理システム１００は、本実施例を具体化し得るコンピュータのネットワークである。ネットワーク・データ処理システム１００はネットワーク１０２を含み、そのネットワークは、ネットワーク・データ処理システム１００内で接続される種々の装置及びコンピュータの間の通信リンクを提供するために使用される媒体である。ネットワーク１０２は、有線通信リンク、無線通信リンク、又は光ファイバ・ケーブルのような接続体を含み得る。

【００１４】
図示の例では、サーバ１０４及びサーバ１０６が記憶ユニット１０８と共にネットワーク１０２に接続している。更に、クライアント１１０、１１２、及び１１４もネットワーク１０２に接続している。しかし、ネットワーク・データ処理システム１００が、図示されてない更なるサーバ、クライアント、及び他の装置を含み得るということを留意されたい。クライアント１１０、１１２、及び１１４は、サーバ１０４及び／又はサーバ１０６に対するクライアントである。更に、クライアント１１０、１１２、及び１１４は、例えば、パーソナル・コンピュータ又はネットワーク・コンピュータであってもよい。
【００１５】
図示の例では、サーバ１０４及びサーバ１０６はクラスタ構成されたサーバである。更に、サーバ１０４及び１０６は、サブスクライバであるクライアント１１０、１１２、及び１１４にパブ・サブ・ネットワーク・サービスを提供する。パブ・サブ・ネットワークは、特定のメッセージング・トピックに対するメッセージのパブリッシュをサポートする。トピックは、複数のサブスクライバに対して関心のある主題を示す。一般に、メッセージは、パブリッシュ・プロセス中にトピックに割り当てられ、しかる後、その特定のメッセージング・トピックをサブスクライブしているすべての消費者によって受け取られる。ゼロ又はそれ以上のサブスクライバ・クライアントが、特定のメッセージ・トピックに関するメッセージを受け取ることに関心があることを登録し得る。
【００１６】
サブスクリプションは、それが永続的及び耐久的なものとなるように構成される。サブスクリプションは、或る消費者が或るクラスのイベントを受け取ることに関心があるということを表す。例えば、Ｊａｖａメッセージング・サービス（ＪＭＳ）のようなメッセージング・サービス環境内のサブスクリプションは、トピックに対するイベントを、そのイベントがパブリッシュされた順序で受け取るための「仮想キュー」として作用する。永続的であるということは、クライアントがサブスクリプションからメッセージを読むことを停止するとき、そのクライアントが取りやめたサブスクリプション・キュー内に未読のメッセージが残っているということを意味する。
【００１７】
更に、ネットワーク・データ処理システム１００は、セッション・アフィニティを利用する分散型ネットワーク環境である。セッション・アフィニティは、同じアプリケーション・サーバ・インスタンスに同じ集中セッションの一部である要求を経路指定するために負荷平衡化要素を利用する。集中セッションに対処するために、本実施例はセッション情報を伴う要求を符号化する。なお、セッション情報は、要求を経路指定するために使用することが可能である。このセッション情報の符号化は２つの方法で行うことが可能である。１つの方法は、クライアント・アプリケーションが、例えば、クッキーのようなセッション基準を集中アプリケーションから得るものである。この場合、クライアント・アプリケーションは、その要求におけるかかるクッキーを再生する。もしくは、クライアント・アプリケーションが、要求が適切なサーバに送られるように、その要求のユニフォーム・リソース・アイデンティファイア（ＵＲＩ）を符号化する。
【００１８】
図示の例では、ネットワーク・データ処理システム１００は、相互に通信を行なうためにＴＣＰ／ＩＰスイート・プロトコルを使用するネットワーク及びゲートウェイの世界的な集合体を表すネットワーク１０２を備えたインターネットである。そのインターネットの中心には、数千個の商業的コンピュータ・システム、政府機関のコンピュータ・システム、教育機関のコンピュータ・システム、及び、データ及びメッセージを経路指定する他のコンピュータ・システムから成る、主要ノード又はホスト・コンピュータ間の高速データ通信回線のバックボーンがある。勿論、ネットワーク・データ処理システム１００は、例えば、イントラネット、ローカル・エリア・ネットワーク（ＬＡＮ）、広域ネットワーク（ＷＡＮ）のような数多くの様々なタイプのネットワークとして具現化することも可能である。図１は種々の実施例の一例として意図され、それらの実施例に対するアーキテク

チャ上の限定として意図されるものではない。

【００１９】

次に、図２を参照すると、本実施例を具現化し得るデータ処理システムのブロック図が示される。データ処理システム２００は、図１におけるサーバ１０４又はクライアント１１０のようなコンピュータの一例であり、そこには、本実施例のためのプロセスを具現化するコンピュータ使用可能プログラム・コード又は命令が設けられる。

【００２０】

図示の例では、データ処理システム２００は、インターフェース及びメモリ・コントローラ・ハブ（インターフェース／ＭＣＨ）２０２並びにインターフェース及び入出力（Ｉ／Ｏ）コントローラ・ハブ（インターフェース／ＩＣＨ）２０４を含むハブ・アーキテクチャを使用する。処理ユニット２０６、メイン・メモリ２０８、及びグラフィックス・プロセッサ２１０はインターフェース／ＭＣＨ２０２に接続される。処理ユニット２０６は１つ又は複数のプロセッサを含み得るし、１つ又は複数の異種のプロセッサ・システムを使用して具現化されてもよい。グラフィックス・プロセッサ２１０は、例えば、アクセラレイテッド・グラフィック・ポート（ＡＧＰ）を介してインターフェース／ＭＣＨ２０２に接続されてもよい。

【００２１】

図示の例では、ローカル・エリア・ネットワーク（ＬＡＮ）アダプタ２１２がインターフェース／ＩＣＨ２０４に接続され、オーディオ・アダプタ２１６、キーボード及びマウス・アダプタ２２０、モデム２２２、リード・オンリ・メモリ（ＲＯＭ）２２４、ユニバーサル・シリアル・バス（ＵＳＢ）及び他のポート２３２、並びにＰＣＩ／ＰＣＩｅ装置２３４がバス２３８を介してインターフェース／ＩＣＨ２０４に接続され、ハードディスクドライブ（ＨＤＤ）２２６及びＣＤ－ＲＯＭ２３０がバス２４０を介してインターフェース／ＩＣＨ２０４に接続される。ＰＣＩ／ＰＣＩｅ装置は、例えば、イーサネット・アダプタ、アドイン・カード、及びノート型コンピュータ用ＰＣカードを含み得る。ＰＣＩはカード・バス・コントローラを使用し、一方、ＰＣＩｅはそれを使用しない。ＲＯＭ２２４は、例えば、フラッシュ・バイナリ入出力システム（ＢＩＯＳ）であってもよく、ＨＤＤ２２６及びＣＤ－ＲＯＭ２３０は、例えば、統合ドライブ・エレクトロニクス（ＩＤＥ）又はシリアル・アドバンスト・テクノロジ・アタッチメント（ＳＡＴＡ）インターフェースを使用し得る。スーパＩ／Ｏ（ＳＩＯ）装置２３６がインターフェース／ＩＣＨ２０４に接続されてもよい。

【００２２】

オペレーティング・システムが処理ユニット２０６において作動し、図２におけるデータ処理システム２００内の種々のコンポーネントの制御を調整及び提供する。オペレーティング・システムは、Microsoft Windows Vistaのような市販のオペレーティングであってもよい。なお、Microsoft及びWindows Vistaは、米国におけるマイクロソフト社の登録商標である。Javaプログラミング・システムのようなオブジェクト指向プログラミング・システムが、そのオペレーティング・システムに関連して作動し得るし、データ処理システム２００において実行されるJavaプログラム又はアプリケーションからそのオペレーティング・システムにコールを行ない得る。Java及びすべてのJavaベースの商標は、米国におけるサン・マイクロシステムズ社の登録商標である。

【００２３】

オペレーティング・システム、オブジェクト指向プログラミング・システム、及びアプリケーション又はプログラムのための命令は、ＨＤＤ２２６のような記憶装置に置かれ、処理ユニット２０６による実行のためにメイン・メモリ２０８にロードされる。本実施例のプロセスは、例えば、メイン・メモリ２０８、ＲＯＭ２２４のようなメモリ或いは１つ又は複数の周辺装置に置かれるコンピュータ実装の命令を使用して処理ユニット２０６により遂行される。

【００２４】

図１及び図２におけるハードウェアは具現化態様に従って変化し得る。フラッシュ・メ

モリ、同等の不揮発性メモリ、又は光ディスク・ドライブ等のような他の内部ハードウェア又は周辺装置が、図１及び図２に示されたハードウェアに加えて、又はそのハードウェアの代わりに使用されてもよい。更に、本実施例のプロセスは、マルチプロセッサ・データ処理システムに適用されてもよい。

【００２５】

或る実例例では、データ処理システム２００は、オペレーティング・システム・ファイル及び／又はユーザ生成のデータを格納するための不揮発性メモリを提供するために一般にはフラッシュ・メモリと共に構成される携帯情報端末（ＰＤＡ）であってもよい。バス・システムは、システム・バス、Ｉ／Ｏバス、及びＰＣＩバスのような１つ又は複数のバスで構成されてもよい。勿論、任意のタイプの通信ファブリック又はアーキテクチャを使用して、その通信ファブリック又はアーキテクチャに接続された種々のコンポーネント又は装置の間でデータの転送を行なうバス・システムが実装されてもよい。通信ユニットは、モデム又はネットワーク・アダプタのような、データを送受信するために使用される１つ又は複数の装置を含み得る。メモリは、例えば、メイン・メモリ２０８又は、インターフェース／ＭＣＨ２０２において見られるようなキャッシュであってもよい。処理ユニットは、１つ又は複数のプロセッサ又はＣＰＵを含み得る。図１及び図２における図示の例並びに上記の例は、アーキテクチャ上の限定を暗示することを意味しない。例えば、データ処理システム２００は、ＰＤＡ形式をとることの他に、タブレット・コンピュータ、ラップトップ・コンピュータ、又は電話装置であってもよい。

【００２６】

本実施例は、分散型ネットワーク内のサーバ・クラスタにおいてフェイルオーバを管理するためにコンピュータを利用する方法、システム、及びコンピュータ使用可能なプログラム・コードを提供する。サーバ・クラスタにおける障害サーバを検出したことに応答して、フェイルオーバ・サーバが、自己のサブスクリプション・メッセージング・キューのサブスクリプション・メッセージ処理を停止する。しかる後、フェイルオーバ・サーバは、障害サーバのサブスクリプション・キューを開き、特定のメッセージング・トピックの全てのサブスクライバにマーカ・メッセージをパブリッシュする。マーカ・メッセージは、その障害サーバのサブスクリプション・キューをその時点で管理しているフェイルオーバ・サーバの識別子を含む。更に、マーカ・メッセージは、フェイルオーバに参加していない他のすべてのサーバによって無視される。

【００２７】

更に、フェイルオーバ・サーバは、障害サーバのサブスクリプション・キュー内のメッセージを処理する。障害サーバのサブスクリプション・キューにおけるメッセージを処理している間、フェイルオーバ・サーバは、サブスクリプション・キューにおけるメッセージがマーカ・メッセージであるかどうかを判断する。障害サーバのサブスクリプション・キューにおけるマーカ・メッセージが見つかったことに応答して、フェイルオーバ・サーバは、障害サーバのサブスクリプション・キューを閉じ、当該フェイルオーバ・サーバのサブスクリプション・メッセージ処理を再開して自己のサブスクリプション・キューを処理する。

【００２８】

本実施例は、分散型パブ・サブ・ネットワーク環境において具現化することが可能である。開始時における各アプリケーション・サーバは、セッションのライフタイムにおいて存続する一意的なサブスクリプション識別子を発生する。各サブスクリプションは耐久性及び永続性がある。セッションがアプリケーション・サーバにおいて作成されるときには常に、アプリケーション・サーバは、セッション属性がセッション状態複製の一部として複製されるように、一意的なサブスクリプション識別子をセッション属性に格納する。フェイルオーバが生じるとき、障害サーバのセッションが別のサーバプリケーション上で活性化される。

【００２９】

一般に、例えば、Ｊ２ＥＥアプリケーション・サーバのようなアプリケーション・サー

バ環境では、フェイルオーバ・サーバ・アプリケーション・コードは、ライフサイクル・リスナを介してこの活動を信号で知らされる。しかし、ライフサイクル・リスナを介するほかに、プラットフォームによる多くの様々な方法でこの活動をアプリケーション・コードに知らせることが可能であるということに留意されたい。活動時、フェイルオーバ・サーバ・アプリケーションは、フェイルオーバしたサブスクリプションをセッションから見つける。しかる後、フェイルオーバ・サーバは、先入れ／先出し（ＦＩＦＯ）順に配布され且つ障害サーバのサブスクリプション・キューにおいても現れるマーカ・メッセージを直ちにすべてのサブスクリプションにパブリッシュする。

【００３０】

しかる後、フェイルオーバ・サーバは、自己の一次サブスクリプションに関する処理を停止し、フェイルオーバしたサブスクリプションから回復するように進み、障害サーバのセッションに関連するメッセージだけを処理する。回復している間、フェイルオーバ・サーバは、それが処理したメッセージのマップを作成する。一旦マーカ・メッセージがその障害サーバのサブスクリプション・キューにおいてヒットされると、回復の第１段階が終了する。

【００３１】

その後、フェイルオーバ・サーバ・アプリケーションは、それがマーカ・メッセージを見つけるまで一次キューの処理を再開する。マーカ・メッセージは、キューが現在同期していることを示す。一次キューを探索している間、フェイルオーバ・サーバはマップにあるメッセージをスキップするので、それは、回復されたセッションのための如何なるメッセージも処理しない。一旦フェイルオーバ・サーバがマーカ・メッセージを見つけると、フェイルオーバ・サーバはフェイルオーバしたサブスクリプションを終了し、正規の処理を再開する。しかし、フェイルオーバ・サーバが回復しているとき、障害サーバが動き出す場合、障害サーバが新しい一意的なサブスクリプション識別子を生成し、それによって、フェイルオーバ・サーバとの不一致を回避するので、フェイルオーバ・サーバは影響を受けない、ということに留意されたい。

【００３２】

次に図３を参照すると、本実施例に従って、通常のサーバ・オペレーションに関する典型的なプロセスを表すフローチャートが示される。図３に示されるプロセスは、例えば、図１におけるサーバ１０４のようなサーバにおいて具現化することが可能である。

【００３３】

プロセスは、サーバが起動するときに始まる（ステップ３０２）。起動後、サーバは、そのセッションのライフタイムの間生きている一意的なサブスクリプション識別子を生成する（ステップ３０４）。サーバは、その一意的なサブスクリプション識別子を使って、サブスクリプション・データに対する変更の通知をサブスクライバ・クライアントにパブリッシュするために使用されるメッセージング・トピックに接続する。

【００３４】

ステップ３０４において一意的なサブスクリプション識別子を生成することに続いて、サーバは、その一意的なサブスクリプション識別子を使ってメッセージング・トピックに接続する（ステップ３０６）。ステップ３０６においてメッセージング・トピックに接続した後、サーバは、これが新しいセッションであるかどうかに関して決定を行なう（ステップ３０８）。これが新しいセッションである場合（ステップ３０８の「イエス」出力）、サーバは、一意的なサブスクリプション識別子を属性としてセッション・オブジェクトに格納する（ステップ３１０）。しかる後、プロセスはステップ３１２に進む。

【００３５】

これが新しいセッションではない場合（ステップ３０８の「ノー」出力）、サーバは要求をサービスする（ステップ３１２）。サーバは要求処理を行なう（ステップ３１４）。更に、サーバは、特定のメッセージング・トピックに変更をパブリッシュする（ステップ３１６）。更に、サーバは、セッション状態複製のためにセッション・オブジェクトを複製する（ステップ３１８）。しかる後、プロセスはステップ３１２に戻り、サーバは要求

のサービスを継続する。

【００３６】

次に、図４を参照すると、本実施例に従って、フェイルオーバ・サーバ・オペレーションのための例示的なプロセスを表すフローチャートが示される。図４に示されるプロセスは、例えば、図１におけるサーバ１０６のようなサーバおいて具現化することが可能である。

【００３７】

プロセスは、サーバが、例えば図１におけるサーバ１０４のようなサーバ・クラスタ内の別のサーバの障害を検出するときに始まる（ステップ４０２）。ステップ４０２において他のサーバの障害を検出した後、別のサーバの障害を検出したサーバ（フェイルオーバ・サーバ）はそれ自身のサブスクリプション・メッセージの処理を停止する（ステップ４０４）。しかる後、フェイルオーバ・サーバは、障害サーバのサブスクリプション・キューを開く（ステップ４０６）。更に、フェイルオーバ・サーバは、現在その障害サーバのセッションにサービスを提供しているサーバに関する一意的なサブスクリプション識別子を含むマーカ・メッセージを、すべてのサブスクライバにパブリッシュする（ステップ４０８）。更に、マーカ・メッセージは、フェイルオーバが生じたとき、フェイルオーバ・サーバのサブスクリプション・キュー及び障害サーバのサブスクリプション・キューの両方においてＦＩＦＯ順に現れる。フェイルオーバ・サーバは、このマーカ・メッセージを使用して、後述するように両方のサーバのサブスクリプション・キューを同期させる。

【００３８】

しかる後、フェイルオーバ・サーバは障害サーバのサブスクリプション・キューからメッセージを取得する（ステップ４１０）。しかる後、フェイルオーバ・サーバは、そのメッセージがマーカ・メッセージであるかどうかに関して判断を行なう（ステップ４１２）。メッセージがマーカ・メッセージではない場合（ステップ４１２からの「ノー」出力）、フェイルオーバ・サーバはそのメッセージを「読取り済み」として記録する（ステップ４１４）。ステップ４１４で見られるようにメッセージを記録した後、フェイルオーバ・サーバは関連のセッションに関する処理を行なう（ステップ４１６）。しかる後、プロセスはステップ４１０に戻り、フェイルオーバ・サーバは、障害サーバのサブスクリプション・キューから別のメッセージを得る。

【００３９】

ステップ４１２に戻ると、メッセージがマーカ・メッセージである場合（ステップ４１２の「イエス」出力）、フェイルオーバ・サーバは、障害サーバのサブスクリプション・キューを閉じる（ステップ４１８）。しかる後、フェイルオーバ・サーバは、それ自身のサブスクリプション・キューの処理を再開する（ステップ４２０）。そのサブスクリプション・キューの処理を再開した後、フェイルオーバ・サーバは、そのサブスクリプション・キューからメッセージを得る（ステップ４２２）。しかる後、フェイルオーバ・サーバは、そのメッセージがマーカ・メッセージであるかどうかに関して判断を行なう（ステップ４２４）。メッセージがマーカ・メッセージである場合（ステップ４２４の「イエス」出力）、フェイルオーバ・サーバは正規の処理を再開する（ステップ４２６）。しかる後、プロセスは終了する。

【００４０】

メッセージがマーカ・メッセージではない場合（ステップ４２４の「ノー」出力）、フェイルオーバ・サーバは、メッセージが以前に読み取られたかどうかに関して判断を行なう（ステップ４２８）。メッセージが読み取られた場合（ステップ４２８の「イエス」出力）、プロセスはステップ４２２に戻り、フェイルオーバ・サーバはそれ自身のサブスクリプション・キューから別のメッセージを得る。メッセージが読み取られなかった場合（ステップ４２８の「ノー」出力）、フェイルオーバ・サーバは、関連のセッションに関する処理を行なう（ステップ４３０）。しかる後、そのプロセスは、再びステップ４２２に戻る。

【００４１】

従って、本実施例は、セッション・アフィニティを利用するパブリッシュ・サブスクライブ分散型ネットワーク環境においてサーバ・フェイルオーバを処理するためにコンピュータを利用する方法、システム、及びコンピュータ使用可能プログラム・コードを提供する。本発明は、全体的にハードウェアの実施例、全体的にソフトウェアの実施例、又は、ハードウェア要素及びソフトウェア要素の両方を含む実施例の形式を取り得る。好ましい実施例では、本発明は、ファームウェア、駐在ソフトウェア、マイクロコード等を含むがそれに限定されないソフトウェアにおいて具現化される。

【００４２】

更に、本発明は、コンピュータ又は任意の命令実行システムによる使用、或いは、それに関連した使用のためのプログラム・コードを提供するコンピュータ使用可能媒体又はコンピュータ可読媒体からアクセスし得るコンピュータ・プログラムの形式を取り得る。説明の便宜上、コンピュータ使用可能媒体又はコンピュータ可読媒体は、命令実行システム、装置、又はデバイスによる使用、或いは、それに関連した使用のためのプログラムを含み、格納し、通信し、又は搬送し得る任意の有形の装置であってもよい。

【００４３】

その媒体は、電子的、磁気的、光学的、電磁気的、赤外線、又は半導体システム（又は装置又はデバイス）であってもよい。コンピュータ可読媒体の例は、半導体又はソリッド・ステート・メモリ、磁気テープ、取り外し可能なコンピュータ・ディスケット、ランダム・アクセス・メモリ（ＲＡＭ）、リード・オンリ・メモリ（ＲＯＭ）、固定磁気ディスク、及び光ディスクを含む。光ディスクの現在の例は、コンパクト・ディスク－リード・オンリ・メモリ（ＣＤ－ＲＯＭ）、コンパクト・ディスク－リード／ライト（ＣＤ－Ｒ／Ｗ）、又はＤＶＤを含む。

【００４４】

更に、コンピュータ記憶媒体はコンピュータ可読プログラム・コードを含み又はそれを格納し得るので、そのコンピュータ可読プログラム・コードがコンピュータ上で実行されるとき、このコンピュータ可読プログラム・コードの実行は、そのコンピュータに、通信リンクを介して他のコンピュータ可読プログラム・コードを伝送させ得る。この通信リンクは、例えば、制限なく、物理的な媒体又は無線媒体を使用し得る。

【００４５】

プログラム・コードを格納及び／又は実行するに適したデータ処理システムは、システム・バスを介してメモリ素子に直接的又は間接的に接続された少なくとも１つのプロセッサを含むであろう。メモリ素子は、プログラム・コードの実際の実行中に使用されるローカル・メモリ、大容量記憶装置、及び、プログラム・コードの実行中にそれらのコードが大容量記憶装置から検索されるべき回数を減らすために少なくとも幾つかのプログラム・コードの一時的記憶装置を提供するキャッシュ・メモリを含み得る。

【００４６】

入出力装置、即ちＩ／Ｏ装置（キーボード、ディスプレイ、ポインティング・デバイスを含むがそれらに限定されない）は、システムに、直接に又は介在するＩ／Ｏコントローラを介して接続される。

【００４７】

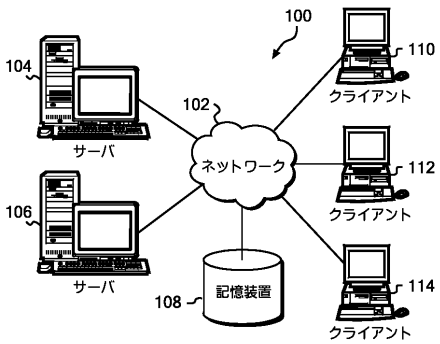
データ処理システムが、介在する専用ネットワーク又は公衆ネットワークを介して他のデータ処理システム或いは遠隔のプリンタ又は記憶装置に接続されることを可能にするために、ネットワーク・アダプタがシステムに接続されてもよい。モデム、ケーブル・モデム、及びイーサネット・カードは数少ない現在利用可能なタイプのネットワーク・アダプタである。

【００４８】

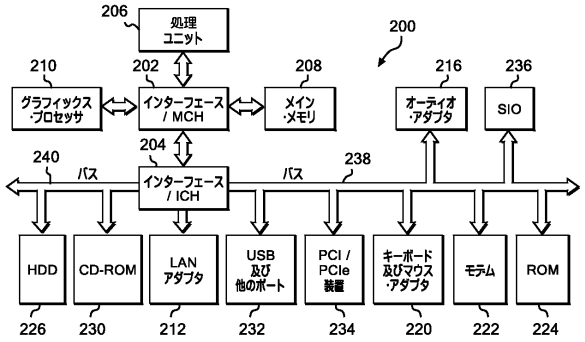
本発明に関する記述は説明を目的として示され、網羅的であること又は開示された形式の発明に限定されることを意図するものではない。当業者には多くの修正および変更が明らかであろう。実施例は、本発明の原理及び実用的応用例を最もよく説明するために、また、当業者が、意図した特定の用途に適するように種々の修正を備えた種々の実施例のた

めに本発明を理解することを可能にするために、選択及び説明されたものである。

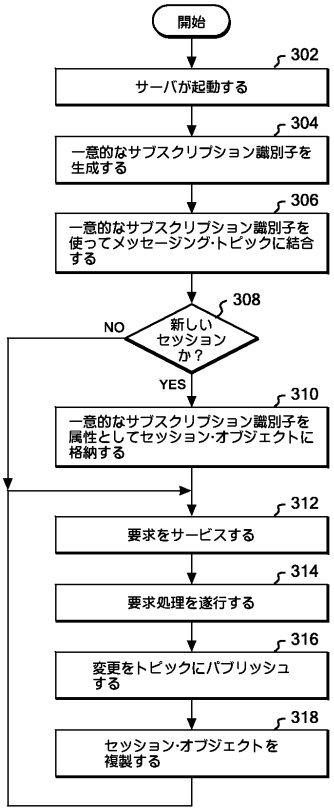
【図１】



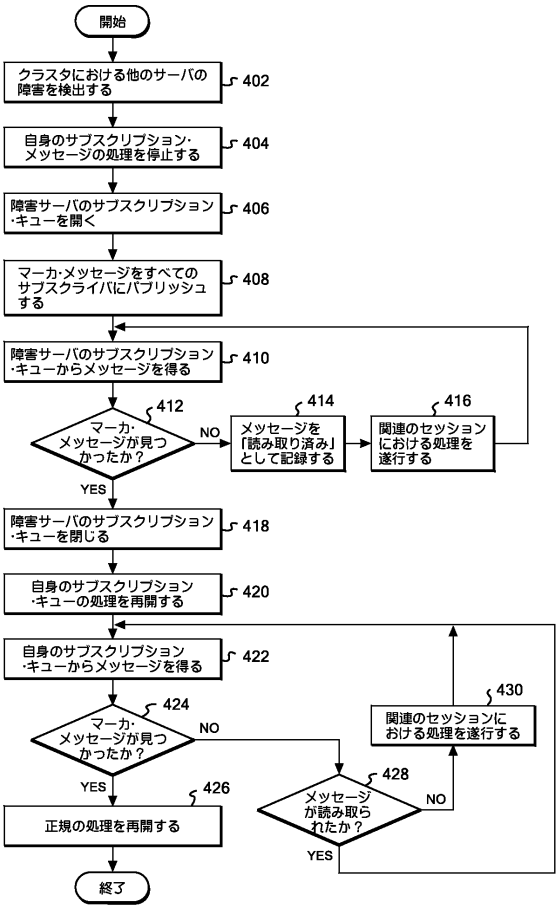
【図２】



【図３】



【図４】


開始
クラスタにおける他のサーバの障害を検出する
402
自身のサブスクリプション・メッセージの処理を停止する
404
障害サーバのサブスクリプション・キューを開く
406
マーカ・メッセージをすべてのサブスクライバにパブリッシュする
408
障害サーバのサブスクリプション・キューからメッセージを得る
410
412
マーカ・メッセージが見つかったか？
NO
414
メッセージを「読み取り済み」として記録する
416
関連のセッションにおける処理を遂行する
YES
障害サーバのサブスクリプション・キューを閉じる
418
自身のサブスクリプション・キューの処理を再開する
420
自身のサブスクリプション・キューからメッセージを得る
422
424
マーカ・メッセージが見つかったか？
NO
430
関連のセッションにおける処理を遂行する
YES
426
正規の処理を再開する
428
メッセージが読み取られたか？
NO
YES
終了

フロントページの続き

(72)発明者　ギルフィックス、マイケル
アメリカ合衆国７８７３１、テキサス州、オースチン、キャニオンサイド・トレール　４３０１
(72)発明者　ムーア、ビクター
アメリカ合衆国３２０２４、フロリダ州、レイク・シティ、サウスウエスト・ダイアル・アベニュー　７７６
(72)発明者　ロウベル　ジュニア、アンソニー、ウィリアム
アメリカ合衆国２７６１５、ノース・カロライナ州、ローリー、レスリー・ドライブ　１０６１２
(72)発明者　チェン、ベンソン、クアン、イー
アメリカ合衆国２７７１３、ノース・カロライナ州、ダーハム、チャンセラーズ・リッジ・ドライブ　１０１４
(72)発明者　ギルモア、マーク、デイビッド
アメリカ合衆国２７７１３、ノース・カロライナ州、ダーハム、シルバーウッド・コート　１
(72)発明者　タルアビブ、オフィラ
イスラエル国６０９４６、ビザロン、モシャブ、ピー．オー．ボックス　１０９

審査官　▲高▼橋　正▲徳▼

(56)参考文献　特開２００７－１７９３１０（ＪＰ，Ａ）
特開２００５－３０１４４２（ＪＰ，Ａ）
国際公開第２００５／０９１１３４（WO，Ａ１）
特開２００５－２０９１９１（ＪＰ，Ａ）
特表２００５－５０５８３３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F　11/16-11/20,
G06F　9/50