

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5006348号
(P5006348)

(45) 発行日 平成24年8月22日 (2012. 8. 22)

(24) 登録日 平成24年6月1日 (2012. 6. 1)

(51) Int. Cl.

F I

G 0 6 F 12/08 (2006. 01)

G 0 6 F 12/00 (2006. 01)

G 0 6 F 13/00 (2006. 01)

G 0 6 F 12/08 5 1 7 B

G 0 6 F 12/00 5 4 6 K

G 0 6 F 12/08 5 0 9 Z

G 0 6 F 12/08 5 5 1 H

G 0 6 F 12/08 5 5 9 Z

請求項の数 11 (全 13 頁) 最終頁に続く

(21) 出願番号 特願2008-558274 (P2008-558274)
 (86) (22) 出願日 平成19年2月6日 (2007. 2. 6)
 (65) 公表番号 特表2009-529183 (P2009-529183A)
 (43) 公表日 平成21年8月13日 (2009. 8. 13)
 (86) 国際出願番号 PCT/US2007/003345
 (87) 国際公開番号 W02007/102969
 (87) 国際公開日 平成19年9月13日 (2007. 9. 13)
 審査請求日 平成22年2月5日 (2010. 2. 5)
 (31) 優先権主張番号 11/370, 585
 (32) 優先日 平成18年3月8日 (2006. 3. 8)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 500046438
 マイクロソフト コーポレーション
 アメリカ合衆国 ワシントン州 9805
 2-6399 レッドモンド ワン マイ
 クロソフト ウェイ
 (74) 代理人 100077481
 弁理士 谷 義一
 (74) 代理人 100088915
 弁理士 阿部 和夫
 (72) 発明者 アニル ケー. ルイア
 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 パテンツ内

最終頁に続く

(54) 【発明の名称】 応答出力キャッシュに対するマルチキャッシュ協調

(57) 【特許請求の範囲】

【請求項 1】

データをクライアントシステムに提供するサーバシステムを含むコンピューティング環境において、前記サーバシステムはデータを格納するための複数のキャッシュを備え、最初に最小コストのキャッシュでクライアントによる要求データをキャッシュする方法であって、前記方法は、

キャッシュでキャッシュすべきデータを受信するステップと、

前記データが以前に先行キャッシュでキャッシュされていなかったことを判定するステップと、

前記データが以前に先行キャッシュでキャッシュされていなかった場合に、

前記キャッシュが前記データをキャッシュする機能的な能力を含むかを、イベントによる生成時にどの特定のタイプのデータがキャッシュされるべきかを定義する設定に基づいて判定するステップであって、(i) 前記キャッシュの前記機能的な能力を、前記複数のキャッシュに適用可能なオブジェクトに格納されるような共通な設定集合であって前記複数のキャッシュの全てに適用する共通な設定集合と比較することによって、かつ、(i i) 前記キャッシュの前記機能的な能力を、前記複数のキャッシュに適用可能な協調的な設定集合と比較することによって、判定するステップであって、前記協調的な設定集合は、イベントによる生成時にどの特定のタイプのデータがキャッシュされるべきかを定義し、該協調的な設定集合は前記複数のキャッシュ内のキャッシュの各々に対する個々の設定からなる集合であり、前記機能的な能力を前記共通な設定集合および前記協調的な設定集

10

20

合と比較することは、少なくとも前記キャッシュの速度及び／又はコストを考慮することを含む、判定するステップと、

前記データをキャッシュする機能的な能力が判定されると、前記キャッシュで前記データをキャッシュするステップと
を備えることを特徴とする方法。

【請求項 2】

前記データが格納されるキャッシュは後続キャッシュであり、前記先行キャッシュは、前記後続キャッシュよりコストが低いキャッシュであることを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記キャッシュは、先行キャッシュによって階層的に先行される後続キャッシュであり、キャッシュでキャッシュすべきデータを受信する前記ステップは、前記データを前記後続キャッシュで受信することを含み、該データは、前記先行キャッシュで前記データがキャッシュされないと判定された後に前記後続キャッシュに対して該データを転送する前記先行キャッシュから受信されることを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記共通な設定集合および前記協調的な設定集合は、前記オブジェクトを参照することによってアクセスされ、該オブジェクトはキャッシュポリシーオブジェクトであり、

前記キャッシュで前記データをキャッシュする前記ステップは、前記キャッシュポリシーオブジェクトに格納される設定が前記データがキャッシュされるべきであることを示す場合に実施されることを特徴とする請求項 1 に記載の方法。

【請求項 5】

キャッシュでキャッシュすべきデータを受信するステップは、前記データを先行キャッシュから受信するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 6】

前記先行キャッシュはウェブサーバキャッシュであり、前記後続キャッシュはアプリケーションレベルキャッシュであることを特徴とする請求項 2 または 5 に記載の方法。

【請求項 7】

前記先行キャッシュはネイティブキャッシュであり、前記後続キャッシュはマネージキャッシュであることを特徴とする請求項 2 または 5 に記載の方法。

【請求項 8】

複数のデータ構造を格納するコンピュータ読み取り可能な記録媒体であって、

前記複数のデータ構造は、

階層的に関連する複数のキャッシュであって、

第 1 の速度を有する第 1 のキャッシュと、

前記第 1 のキャッシュよりも遅い第 2 の速度を有する第 2 のキャッシュであって、前記第 2 のキャッシュは、その低速度のために、前記第 1 のキャッシュよりも階層内で低く構成され、その結果、前記第 1 のキャッシュがデータを受信し前記第 2 のキャッシュよりも先にデータを要求する、第 2 のキャッシュとを含む、複数のキャッシュと、

前記複数のキャッシュに対する設定であってイベントによる生成時にどの特定のタイプのデータがキャッシュされるべきかに関する設定を定義するデータを保持するポリシーオブジェクトであって、データがキャッシュに送信される場合、該キャッシュに対する前記ポリシーオブジェクトを参照することによって該キャッシュがデータをキャッシュできるかまたはできない、ポリシーオブジェクトと

を備え、

前記第 1 のキャッシュ及び前記第 2 のキャッシュは、前記ポリシーオブジェクトを使用することによって、キャッシュされるべきデータが、前記階層内の前記第 1 のキャッシュが該第 1 のキャッシュでキャッシュできないタイプであると判定する場合に、前記第 2 のキャッシュが前記ポリシーオブジェクトを使用して該第 2 のキャッシュでキャッシュできるかを判定するために、該データが自動的に前記第 2 の、階層的に低いキャッシュに渡さ

10

20

30

40

50

れるように構成されることを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項 9】

データを生成および／または取り出すコンピュータ実行可能命令をさらに備えることを特徴とする請求項 8 に記載のコンピュータ読み取り可能な記録媒体。

【請求項 10】

前記ポリシーオブジェクトに格納すべきポリシー情報を生成するためのコンピュータ実行可能命令をさらに備えることを特徴とする請求項 8 に記載のコンピュータ読み取り可能な記録媒体。

【請求項 11】

データをクライアントシステムに提供するサーバシステムを含むコンピューティング環境において、前記サーバシステムはデータを格納するための複数のキャッシュを備え、最初に最小のコストのキャッシュでクライアントによって要求されるデータをキャッシュする方法を実装するための実行命令を含むコンピュータプログラムであって、

前記方法は、

キャッシュでキャッシュすべきデータを受信するステップと、

前記データが以前に先行キャッシュでキャッシュされていなかったことを判定するステップと、

前記データが以前に先行キャッシュでキャッシュされていなかった場合に、

前記キャッシュが前記データをキャッシュする機能的な能力を含むかを、イベントによる生成時にどの特定のタイプのデータがキャッシュされるべきかを定義する設定に基づいて判定するステップであって、(i) 前記キャッシュの前記機能的な能力を、前記複数のキャッシュに適用可能なオブジェクトに格納されるような共通な設定集合であって前記複数のキャッシュの全てに適用する共通な設定集合と比較することによって、かつ、(i i) 前記キャッシュの前記機能的な能力を、前記複数のキャッシュに適用可能な協調的な設定集合と比較することによって、判定するステップであって、前記協調的な設定集合は、イベントによる生成時にどの特定のタイプのデータがキャッシュされるべきかを定義し、該協調的な設定集合は前記複数のキャッシュ内のキャッシュの各々に対する個々の設定からなる集合であり、前記機能的な能力を前記共通な設定集合および前記協調的な設定集合と比較することは、少なくとも前記キャッシュの速度及び／又はコストを考慮することを含む、判定するステップと、

前記データをキャッシュする機能的な能力が判定されると、前記キャッシュで前記データをキャッシュするステップと

を行なうように構成される、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、応答出力キャッシュに対するマルチキャッシュ協調を実施する方法およびシステムに関する。

【背景技術】

【0002】

コンピュータおよびコンピューティングシステムは現代の生活のほぼ全ての側面に影響を及ぼしている。コンピュータは一般に仕事、レクリエーション、ヘルスケア、運輸、娯楽、家庭管理、等に関与している。コンピュータの機能性はまた、その能力が強化され、様々なネットワーク接続を通して相互接続されるようになってきている。

【0003】

コンピューティングシステムは他システムからデータを要求することができる。このデータを多数の様々な方法で取得することができる。例えば、データベースまたは他のデータストアのような収納場所からデータを取り出すことで、データを取得することができる。また、データを提供するシステムにより、データを動的に生成することもできる。例えば、ワシントン州レッドモンドの本件特許出願人から入手可能な a s p (active server pa

10

20

30

40

50

ge) . N E T は、ウェブページデータを動的に作成するためのツールを含む。通常のウェブサーバでさえも、ウェブページ要求内の引数に基づいて、カスタマイズされたウェブページデータを提供するウェブページ要求を用いて引数を受信する機能性を含む。

【 0 0 0 4 】

コンピューティングシステムの性能を増大させる 1 つの方法は、要求データを出力キャッシュすることである。キャッシュは、最近の要求データを、同一データに対する後続の要求があるときに迅速に取り出せる位置に格納することを含む。例えば、データを非常に迅速に取り出せるように、データをコンピュータシステムの物理メモリにキャッシュすることができる。

【 0 0 0 5 】

キャッシュデータを格納および保持する方法を決定するために様々なキャッシュアルゴリズムが作成されている。例えば、F I F O (first in first out) は先入れ先出しの頭字語である。この方式では、データはキャッシュ位置にキャッシュされ、追加のキャッシュ項目に対する場所を作成する必要がある場合、最も古い項目がキャッシュから削除される。他の方式は予測フィルタを使用して、キャッシュ対象およびキャッシュ時間を決定する。

【 0 0 0 6 】

コンピュータシステムの個々のコンポーネントはそれぞれ自身のキャッシュコンポーネントを有することがよくある。例えば、ウェブサーバのコンテキストでは、そのスタックの底にオペレーティングシステムを有するスタックが存在し、ウェブサーバがオペレーティングシステムの上にあり、アプリケーションがウェブサーバの上にある。この 1 つの例は、a s p . N E T フレームワークをホストする I I S (internet information server) サーバをホストする W i n d o w s (登録商標) オペレーティングシステムであり、その全てはワシントン州レッドモンドの本件特許出願人から入手可能である。オペレーティングシステム、ウェブサーバ、およびアプリケーションの各々はそれぞれ自身のキャッシュを含む。

【 発明の開示 】

【 発明が解決しようとする課題 】

【 0 0 0 7 】

データをクライアントに提供しよう協働するときに、これらの個々のコンポーネントの各々がデータをキャッシュする場合、冗長なキャッシュが実施される可能性がある。この結果、リソースが浪費される恐れがある。例えば、冗長なキャッシュ項目を物理メモリに保持することで、他の項目のキャッシュに利用可能な物理メモリの量が大幅に減少する。

【 0 0 0 8 】

本明細書で主張する主題は、任意の欠点を解決する実施形態または上述したもののような環境でのみ動作する実施形態に限定されない。むしろ、本背景は、本明細書で説明した幾つかの実施形態を実施可能な 1 つの例示的な技術を示すために与えるに過ぎない。

【 課題を解決するための手段 】

【 0 0 0 9 】

一実施形態はデータのキャッシュ方法に関する。本方法は、例えば、データをクライアントシステムに提供するサーバシステムを含むコンピューティング環境で実施することができる。サーバシステムは、データを格納する多数のキャッシュを含む。本方法により、クライアントが要求したデータを最小コストのキャッシュが最初にキャッシュすることができる。本方法は、キャッシュにキャッシュすべきデータを受信することを含む。本方法はさらに、データが以前に先行キャッシュでキャッシュされていなかったことを判定することを含む。データが以前に先行キャッシュでキャッシュされていなかった場合、データをそのキャッシュにキャッシュする。

【 0 0 1 0 】

本明細書で説明する別の実施形態は、キャッシュからデータを取り出すことに関する。

10

20

30

40

50

本方法を、例えば、データをクライアントシステムに提供するサーバシステムを含むコンピュータ環境で実施することができる。サーバシステムは、データを格納するための多数のキャッシュを含む。本方法により、データをコストの低いキャッシュから最初に取り出すことができる。本方法は、キャッシュにおいてデータに対する要求を受信することを含む。本方法はさらにキャッシュでデータが利用可能であるかどうかを判定することを含む。キャッシュでデータが利用可能である場合、その要求されたデータが返却される。キャッシュでデータが利用可能でない場合、その要求を後続キャッシュに転送する。

【 0 0 1 1 】

本明細書で説明する別の実施形態は、多数のデータ構造を格納したコンピュータ可読媒体に関する。その多数のデータ構造は多数のキャッシュを含む。その多数のキャッシュは、第1の速度を有する第1のキャッシュと、第1のキャッシュに続いて結合された第2のキャッシュとを含む。第2のキャッシュは、第1のキャッシュより低速な第2の速度を有する。第1のキャッシュに続くように接続することで、第1のキャッシュはデータとデータ要求とを第2のキャッシュの前に受信する。データ構造はさらにポリシー・オブジェクトを含む。ポリシー・オブジェクトは複数のキャッシュに対する設定を定義する。

【 0 0 1 2 】

本要約は選択した概念を簡潔な形で導入するために与える。その概念を「発明を実施するための最良の形態」でさらに後述する。本要約は特許請求主題の主要な特徴または本質的な特徴を特定するようには意図しておらず、特許請求主題の範囲の決定を支援するものとしても意図していない。

【 0 0 1 3 】

さらなる特徴および利点を以下の記述で説明し、それらはその記述から部分的に明らかになるであろう。または、それらを本明細書の教示を実施することで習得することができる。本発明の特徴および利点を、特に添付請求項で指摘した手段およびそれらの組み合わせにより理解および取得することができる。本発明の特徴は、以下の記述および添付請求項からより完全に明らかとなり、または以降説明する本発明の実施により習得することができる。

【 0 0 1 4 】

上述並びに他の利点および特徴を取得可能な方法を説明するため、簡潔に上述した主題のより詳細な説明を添付図面で示した具体的な実施形態を参照して行う。これらの図面は一般的な実施形態を示すに過ぎず、従って範囲を限定すると考えるべきではないことを理解し、添付図面の使用を通して実施形態をさらに具体的および詳細に記述および説明する。

【 発明を実施するための最良の形態 】

【 0 0 1 5 】

本明細書の実施形態は、以下でさらに詳細に説明するように様々なコンピュータハードウェアを含む専用または汎用のコンピュータを備えることができる。

【 0 0 1 6 】

本明細書で説明する一実施形態は、キャッシュを互いに階層方式で接続したキャッシュ・トポロジを含む。そのトポロジでは、最高速および/または最小コストのキャッシュが最初にアクセスされて、データのキャッシュまたはキャッシュデータの取り出しのいずれかを行う。特定のキャッシュが所与のデータ集合をキャッシュできない場合、またはその所与のデータ集合がキャッシュで利用可能でない場合には、データまたはデータ要求を後続キャッシュに渡す。後続キャッシュはより低速またはよりコストが高くてよいが、所与のデータ集合をキャッシュするための機能性を含むことができる。本キャッシュ・トポロジは、階層方式で接続した任意の適切な数のキャッシュを含むことができ、より低速および/またはよりコストが高いキャッシュの前により高速および/またはよりコストが低いキャッシュにアクセスすることができる。

【 0 0 1 7 】

一実施形態では、キャッシュすべきデータを定義するポリシー情報が、階層内のキャッ

シュに対して利用可能である。従って、階層内のキャッシュはポリシー情報を参照することができ、それとともに、所与のデータ集合をキャッシュにキャッシュすべきかどうかを判定する際にキャッシュの個々の能力にアクセスすることができる。

【0018】

図1を参照すると、キャッシュ・トポロジ100が多数のキャッシュ102-1、102-2、および102-Nを階層的配置で含む例を示してある。図示した例では、キャッシュデータ104またはキャッシュすべきデータ106に対する要求を第1のキャッシュ102-1に送信することができる。説明の簡潔さのため、第1の例を、データ106が第1のキャッシュ102-1に送信される場合に対して与える。データ106が第1のキャッシュ102-1に送信されるとき、第1のキャッシュ102-1は、送信されているデータタイプを判定することができる。第1のキャッシュ102-1は次いでポリシー情報108を参照して、データ106がキャッシュされるべきかどうかを判定することができる。データ106がキャッシュされるべきである場合、第1のキャッシュ102-2は次いで、キャッシュするデータに対する自身の能力を評価することができる。データ106がキャッシュ102-1でキャッシュ可能なタイプである場合、データ106は第1のキャッシュ102-1でキャッシュされる。

10

【0019】

他方、データがキャッシュ102-1によってキャッシュされることが可能なタイプでない場合、そのデータを第2のキャッシュ102-2に転送する。第2のキャッシュ102-2は、ポリシー情報108を参照してデータ106をキャッシュすべきかどうかを判定すること、および自身の能力を評価してデータ106をキャッシュできるかどうかを判定することのように、第1のキャッシュ102-1が実施するものと同様なアクションを実施する。本プロセスを、キャッシュがデータ106をキャッシュするかまたはデータ106を与えた後にデータ106をキャッシュするキャッシュがなくなるまで、省略記号および102-Nで表した他のキャッシュに対して継続することができる。

20

【0020】

同様なプロセスを、データの取り出しに対して実施することができる。例えば、キャッシュデータ104に対する要求を第1のキャッシュ102-1に送信することができる。第1のキャッシュ102-1がキャッシュデータを有する場合、そのキャッシュデータは第1のキャッシュ102-1から返却される。第1のキャッシュ102-1がキャッシュデータを有さない場合、第1のキャッシュ102-1はキャッシュデータ104に対する要求を第2のキャッシュ102-2に渡す。本プロセスを、省略記号および102-Nにより表したキャッシュを通して継続することができる。

30

【0021】

前述のように、最高速および/または最小コストのキャッシュは、後続キャッシュがより低速および/またはコストがより高いキャッシュである第1のキャッシュ102-1であることができる。例えば、一実施形態では、第1のキャッシュ102-1は、ウェブサーバのような低レベルのコンポーネントに対するキャッシュで可能である。第2のキャッシュ102-2は、例えば、ワシントン州レッドモンドの本件特許出願人から入手可能なASP.NETのようなアプリケーションレベルコンポーネントのような高レベルのコンポーネントに対するキャッシュで可能である。本例では、第1のキャッシュ102-1はネイティブキャッシュ(native cache)で可能であり、第2のキャッシュ102-2はマネージキャッシュ(managed cache)で可能である。ネイティブキャッシュは一般にマネージキャッシュより高速であり、よりコストが低い。

40

【0022】

前述のように、図1はキャッシュポリシー108を示している。幾つかの実施形態では、キャッシュポリシーはキャッシュの各々に対するポリシー設定を含むことができる。例えば、キャッシュポリシーは、キャッシュ102-1、102-2、および102-Nに対応するキャッシュポリシー108-1、108-2、および108-Nを含むことができる。全てのキャッシュ設定がキャッシュポリシー108内部で知られるように、キャッ

50

シュポリシー 108 - 1、108 - 2、および 108 - N を相互接続できる。これにより、キャッシュデータ 106 に対して可能な任意のキャッシュが許可される。従って、最高速および / または最小コストのキャッシュを使用して、キャッシュ能力と、キャッシュの各々に対するキャッシュポリシーを含むキャッシュポリシー 108 と、に基づいてキャッシュできる場合にデータをキャッシュすることができる。

【0023】

代替的な実施形態では、キャッシュポリシー 108 は、キャッシュ 102 - 1、102 - 2、および 102 - N の各々が参照可能な汎用的な共有構成が可能である。従って、キャッシュポリシー 108 が規定するようにデータ 106 を格納する能力を有する第 1 のキャッシュは、データ 106 を格納することになる。キャッシュがデータ 106 を格納しているとき、後続キャッシュがデータ 106 をキャッシュすることを試行しないように、データ 106 を印付けすることができる。一実施形態では、キャッシュすべきでないデータとしてデータ 106 を単純に印付けすることで、これを実現できる。この方法では、キャッシュ 102 - 1、102 - 2、および 102 - N を、特に互いに通信するよう設計する必要はない。むしろ、先行キャッシュはデータ 106 をキャッシュすべきでないデータとして印付けすることができ、従って後続キャッシュがデータをキャッシュすることを防ぐことができる。さらに、図 1 に示すモデルは、キャッシュが互いに通信するよう特に設計する必要なしに任意数の後続キャッシュを追加できるという点で、高度に拡張可能である。

【0024】

次に、図 2 を参照して、ワークフローパイプライン 200 を示す。ワークフローパイプライン 200 は、受信されている要求 202 を示す。要求 202 は、例えば、サーバシステムに接続したクライアントシステムからの要求である可能性がある。要求 202 を出力キャッシュにより処理可能である場合、よりコストが高いコンピューティングおよび / またはネットワーク操作を通して応答を再生成または再フェッチするのではなく、要求 202 に対する応答をキャッシュから提供することで、コンピューティングリソースを温存することができる。さらに、要求 202 に対する応答がキャッシュで利用可能でない場合でも、或る特定の状況では、要求に対する応答をキャッシュに格納してその応答内のデータを要求するクライアントが後で利用できるという具合に、要求に対する応答を本明細書で以下でより詳細に説明する様々なプロセスおよびイベントを通して生成することができる。

【0025】

キャッシュ解決機能 (resolve cache function) 204 は、要求 202 を受信すると実施することができる。要求 202 に対する応答がキャッシュ 206 で利用可能であるときに、キャッシュ解決機能 202 によりキャッシュデータをキャッシュ 206 から返却することができる。要求 202 に対するキャッシュ応答を取得するため、キャッシュデータ 104 に対する要求が送信される。図 2 に示す実施形態では、キャッシュデータ 104 に対する要求を、上述のように、より高速および / またはコストがより低いキャッシュであり得る第 1 のキャッシュ 102 - 1 に送信することができる。第 1 のキャッシュ 102 - 1 が要求 202 に対する応答を含む場合、第 1 のキャッシュ 102 - 1 は、その応答を返却し、キャッシュデータ 104 に対する要求がキャッシュ 102 - 2 および 102 - N のような後続キャッシュに続くことを防ぐことができる。要求 202 に対する応答が第 1 のキャッシュ 102 - 1 で利用可能でない場合、キャッシュデータ 104 に対する要求を、図 1 の説明に関して上述したように後続キャッシュに渡すことができる。前述のように、後続キャッシュは、応答が見つかるまで、またはキャッシュの各々が応答をチェックされるまで、要求 202 に対する応答をチェックし続けることができる。

【0026】

図 2 は決定ブロック 208 を示す。決定ブロック 208 は、応答がキャッシュの 1 つに格納されている場合、応答を、210 で示すようにクライアントに送信することを示す。応答がキャッシュの 1 つに格納されない場合、イベント 212 で示すように様々なアクシ

ョンを実施して応答を別の位置から生成または取り出すことができる。

【 0 0 2 7 】

前述のように、図 2 はイベント 2 1 2 を示す。イベントは、要求 2 0 2 への応答としてのデータを生成するため、および / または要求 2 0 2 に応じてデータを取得するために実施する様々なアクティビティが可能である。例えば、イベント 2 1 2 は a s p . N E T のようなアプリケーションレベルコンポーネントにより実施するアクティビティを含むことができる。イベント 2 1 2 は、要求 2 0 2 を満たすデータとして使用すべきデータベース内に格納したデータを収集するための、データベース抽出機能を含むことができる。幾つかの実施形態により、イベントを任意の適切なレベルで実施することができる。例えば、イベント 2 1 2 をカーネルレベル、システムレベル、および / またはアプリケーションレ

10

【 0 0 2 8 】

図 2 はさらにキャッシュ更新アクション 2 1 4 を示す。データがイベント 2 1 2 によって生成されまたは取り出されるとき、データをキャッシュ 2 0 6 の 1 つに格納することができる。これにより、そのデータを要求する後続の要求に対する応答としてそのデータを使用することができる。前述のように、データ 1 0 6 を第 1 のキャッシュ 1 0 2 - 1 に送信することができる。第 1 のキャッシュ 1 0 2 - 1 はキャッシュポリシー 1 0 8 を参照する。キャッシュポリシー 1 0 8 が、データ 1 0 6 はキャッシュされるべきであり、第 1 のキャッシュ 1 0 2 - 1 はデータ 1 0 6 をキャッシュ可能であると指示する場合、第 1 のキャッシュ 1 0 2 - 1 はデータ 1 0 6 をキャッシュする。次いで、第 1 のキャッシュ 1 0 2 - 1 は後続キャッシュに対し、データ 1 0 6 をキャッシュすべきでないことを示す。一実施形態では、これを、キャッシュすべきでないデータとしてデータ 1 0 6 を印付けすることで実現することができる。あるいは、第 1 のキャッシュ 1 0 2 - 1 は、データ 1 0 6 をキャッシュしないために、別個のメッセージをキャッシュ 1 0 2 - 2 および 1 0 2 - N のような後続キャッシュに送信することができる。

20

【 0 0 2 9 】

図 2 はさらにハンドラ 2 1 6 を示す。ハンドラ 2 1 6 を、イベント 2 1 2 で表すようにコンピュータハードウェアおよびソフトウェアと関連付けることができる。ハンドラ 2 1 6 はポリシー情報をキャッシュポリシー 1 0 8 に提供することもできる。特に、ハンドラ 2 1 6 は、どの特定のタイプのデータをイベント 2 1 2 による生成時にキャッシュすべきかを定義する情報を利用可能とすることができる。本情報をハンドラ 2 1 6 によりキャッシュポリシー 1 0 8 に送ることができる。

30

【 0 0 3 0 】

次に、図 3 を参照して、データを格納する方法 3 0 0 を示す。方法 3 0 0 は、例えば、データをクライアントシステムに提供するサーバシステムを含むコンピューティング環境で実施することができる。そのサーバシステムはデータを格納するための多数のキャッシュを含む。方法 3 0 0 は多数の動作を含む。例えば、方法 3 0 0 は、キャッシュにキャッシュすべきデータを受信することを含む（動作 3 0 2）。方法 3 0 0 はさらに、以前にデータが先行キャッシュでキャッシュされていなかったことを判定することを含む（動作 3 0 4）。データが以前に先行キャッシュでキャッシュされていなかった場合、方法 3 0 0 は、キャッシュポリシーによりデータをこのキャッシュでキャッシュできるかどうかを判定することを示す（動作 3 0 5）。キャッシュポリシーによりデータをこのキャッシュにキャッシュできる場合、方法 3 0 0 はデータをそのキャッシュにキャッシュすることを示す（動作 3 0 6）。そうでない場合、データをそのキャッシュにキャッシュしない（3 0 8）。

40

【 0 0 3 1 】

幾つかの実施形態では、データを後続キャッシュに転送できる。後続キャッシュがデータをキャッシュしないように、データを印付けすることができる。例えば、キャッシュすべきでないデータとしてデータを印付けすることができ、その結果、後続キャッシュはそのデータをキャッシュしない。あるいはそのキャッシュは、そのデータをキャッシュすべ

50

きでないことを後続キャッシュに伝えることができる。

【 0 0 3 2 】

本明細書で前述したように、先行キャッシュは後続のキャッシュよりコストが低いキャッシュで可能である。例えば図 1 に示すように、第 1 のキャッシュ 1 0 2 - 1 は、それが第 2 のキャッシュより高速なキャッシュであるか、またはそれが第 2 のキャッシュ 1 0 2 - 2 ほどリソースを必要としない意味で、低いコストで可能である。同様に、後続キャッシュは上記のキャッシュよりコストが高いキャッシュで可能である。従って、よりコストが低い、および / またはより高速なキャッシュは、よりコストが高く、および / またはより低速なキャッシュ前にデータを格納する機会を有する。

【 0 0 3 3 】

方法 3 0 0 はさらに、キャッシュがデータをキャッシュする機能的な能力を含むかどうかを判定することを含むことができる。これを、1 つの例では動作 3 0 5 で実施することができる。キャッシュがデータをキャッシュするための機能的な能力を含む場合、キャッシュにおいてデータのキャッシュを実施する (動作 3 0 6)。例えば、図 1 を再度参照すると、第 1 のキャッシュ 1 0 2 - 1 はキャッシュポリシー 1 0 8 を参照して、どのタイプのデータをキャッシュすべきかを判定することができる。第 1 のキャッシュ 1 0 2 - 1 は、次いで自身のキャッシュ能力を参照して、自身がデータをキャッシュする能力を有するかどうかを判定することができる。

【 0 0 3 4 】

従って、キャッシュがデータをキャッシュする機能的な能力を備えているかどうかを判定することは、そのキャッシュの機能的な能力を、キャッシュポリシー 1 0 8 のような、キャッシュの数に適用可能なオブジェクトの共通な設定集合と比較することを含むことができる。

【 0 0 3 5 】

前述のように、キャッシュがデータをキャッシュする機能的な能力を備えるかどうかを判定することは、そのキャッシュの機能的な能力を、複数のキャッシュに適用可能な協調的な設定集合と比較することが上げられる。協調的な設定は、情報を共有可能な複数のキャッシュ内のキャッシュの各々に対する個々の設定から成る集合が可能である。従って、先行キャッシュがデータをキャッシュできるかどうかの判定において、先行キャッシュは後続キャッシュからのポリシー設定を使用できる。

【 0 0 3 6 】

方法 3 0 0 は、キャッシュポリシーを参照することでデータがキャッシュされるべきかを判定すること、データをキャッシュすべきとキャッシュポリシーが示す場合にキャッシュにおいてデータをキャッシュすること (動作 3 0 6) を含むことができる。これを、例えば動作 3 0 5 で実施することができる。上述のように、キャッシュポリシー 1 0 8 は、どのタイプのデータを複数のキャッシュ内のキャッシュでキャッシュすべきかを述べる情報を含む。

【 0 0 3 7 】

方法 3 0 0 は、キャッシュでキャッシュすべきデータを受信することが先行キャッシュからデータを受信することを含むようにしても良い。例えば、機能的な制限のため先行キャッシュがデータをキャッシュできない場合に、キャッシュは先行キャッシュからデータを受信できる。

【 0 0 3 8 】

前述のように、キャッシュはより低速なキャッシュより前により高速なキャッシュにアクセスする様々な機能性を有することができる。従って、一実施形態では、キャッシュはウェブサーバキャッシュとすることが可能であり、後続キャッシュはアプリケーションレベルキャッシュとすることが可能である。同様に、キャッシュはネイティブキャッシュとすることが可能であり、後続キャッシュはマネージキャッシュとすることが可能である。

【 0 0 3 9 】

次に、図 4 を参照して、方法 4 0 0 を示す。その方法を、例えば、データをクライアン

10

20

30

40

50

トシステムに提供するサーバシステムを含むコンピューティング環境で実施することができる。そのサーバシステムはデータを格納するための多数のキャッシュを含む。本方法は、クライアントが要求するキャッシュデータを提供するための様々な動作を含む。本方法はキャッシュにおいてデータに対する要求を受信することを含む（動作402）。例えば、図2に示すように、要求104を受信することができる。

【0040】

本方法はさらに、データがキャッシュで利用可能であるかどうかを判定することを含む（動作404）。例えば、第1のキャッシュ102-1に対する要求によりデータを要求する場合、データ104が第1のキャッシュ102-1で利用可能である。データがキャッシュで利用可能である場合、要求データが返却される（動作406）。例えば、図2は10
応答をクライアント210に送信することを示す。データがキャッシュで利用可能でない場合、要求を後続キャッシュに転送する（動作408）。例えば、要求104を第2のキャッシュ102-2に転送することができる。

【0041】

本明細書の他の実施形態で説明するように、キャッシュは後続キャッシュより高速が可能である。

【0042】

特許出願である方法400は、さらに、データが複数のキャッシュで利用可能でない場合に、その複数のキャッシュ以外の位置からデータを生成または取り出すことを含むことができる。例えば、図2に示すように、様々なアクションをイベント212が示すように20
実施して、データを生成、または例えばデータベースもしくは他のデータ格納からデータを取り出すことができる。

【0043】

データを取り出しまたは生成すると、方法400はさらに、データをキャッシュに提供するための動作、データが以前にキャッシュの数の先行キャッシュでキャッシュされていなかったことを判定するための動作、およびデータが以前にキャッシュされていなかったことを判定するステップで判定したようにデータが以前にキャッシュの数のうちの先行キャッシュでキャッシュされていなかった場合に、データをキャッシュでキャッシュするための動作、を含むことができる。

【0044】

さらに、データを生成または取り出すことは、複数のキャッシュにおいてキャッシュすべきデータを指定するキャッシュポリシー情報を複数のキャッシュに提供することを含める。

【0045】

実施形態は、自身に格納したコンピュータ実行可能命令またはデータ構造を携行または有するコンピュータ可読媒体を含むこともできる。上記のコンピュータ可読媒体は、汎用または専用のコンピュータによりアクセス可能な任意の利用可能な媒体が可能である。限定ではなく例として、上記のコンピュータ可読媒体はRAM、ROM、EEPROM、CD-ROMもしくは他の光ディスク格納、磁気ディスク格納もしくは他の磁気格納装置、40
または所望のプログラムコード手段をコンピュータ実行可能命令またはデータ構造の形式で携行または格納するために使用可能であり、および汎用または専用のコンピュータによりアクセス可能な任意の他の媒体を備えることができる。情報をネットワークまたは別の通信接続（ハードワイヤード、無線、もしくはハードワイヤードまたは無線の組合せのいずれか）上でコンピュータに転送または提供する場合、コンピュータはその接続をコンピュータ可読媒体として正しいものと見なす。従って、任意の上記の接続はコンピュータ可読媒体であると正しく称せられる。上の組合せも、コンピュータ可読媒体の範囲に含まれるべきである。

【0046】

コンピュータ実行可能命令は、例えば、汎用コンピュータ、専用コンピュータ、または専用処理装置に或る特定の機能または機能グループを実施させる命令およびデータを備え50

る。特許請求主題を構造的特徴および／または方法論的動作に固有な言葉で説明したが、添付請求項で定義したクレーム主題は必ずしも上述の特定の特徴または動作に限定されないことは理解されるべきである。むしろ、上述の特定の特徴および動作は請求項を実行する形態の例として開示してある。

【0047】

本発明は、その精神または本質的な特徴から逸脱せずに他の特定の形態で具現化することができる。説明した実施形態は全ての点で例に過ぎず限定的ではないと考えるべきである。本発明の範囲は従って、前述の説明によってではなく添付請求項によって示される。請求項の等価物の意味および範囲内にある全ての変更は、その範囲内に包含されるべきである。

10

【図面の簡単な説明】

【0048】

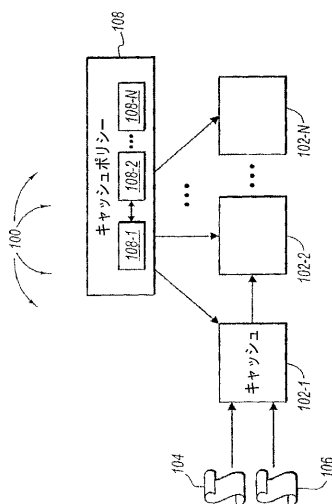
【図1】キャッシュの階層的編成を示す図である。

【図2】データのキャッシュおよび要求に対するアクティビティのパイプラインを示す図である。

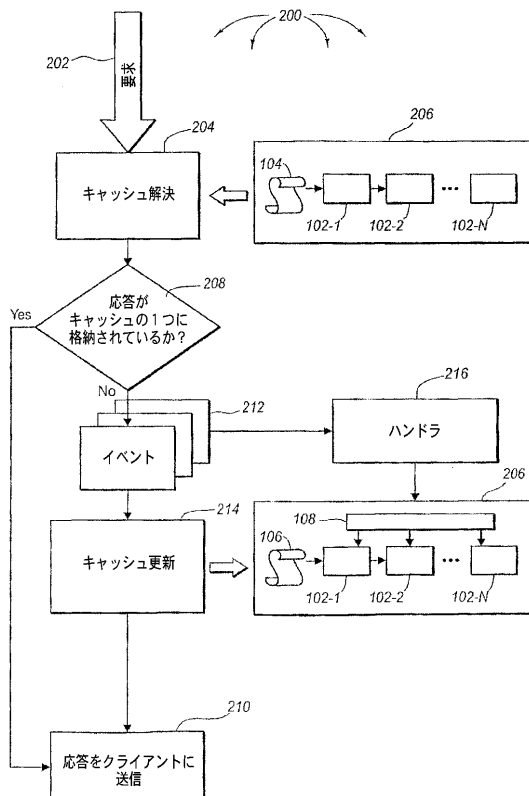
【図3】データのキャッシュ方法を示す図である。

【図4】キャッシュデータの取り出し方法を示す図である。

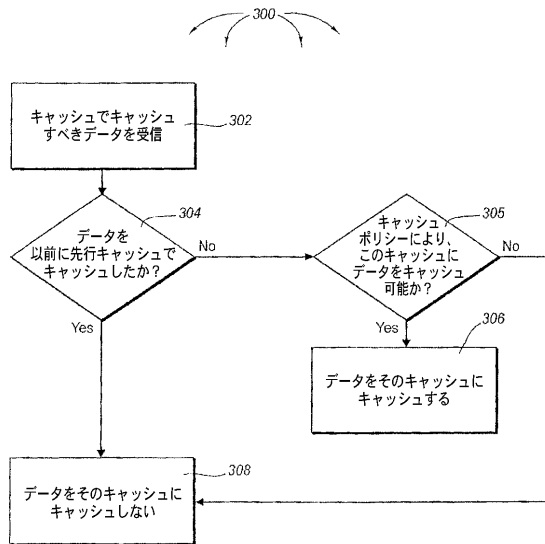
【図1】



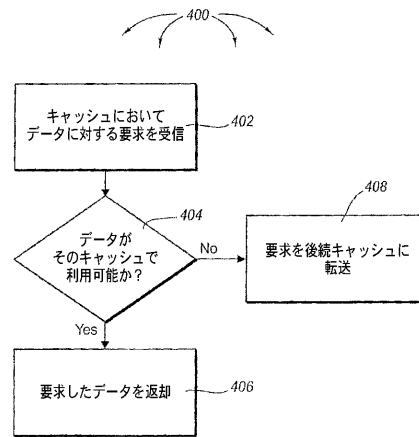
【図2】



【図 3】



【図 4】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 13/00 5 4 0 B

(72)発明者 エリック ビー．オルソン
アメリカ合衆国 9 8 0 5 2 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 パテンツ内

(72)発明者 マイケル ボロダルスキー
アメリカ合衆国 9 8 0 5 2 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 パテンツ内

審査官 桜井 茂行

(56)参考文献 特表 2 0 0 4 - 5 1 6 5 3 2 (J P , A)
特開 2 0 0 1 - 1 2 5 8 3 0 (J P , A)
米国特許第 6 4 3 0 6 5 4 (U S , B 1)
米国特許第 6 8 5 1 0 2 4 (U S , B 1)
米国特許第 6 4 2 5 0 5 8 (U S , B 1)
米国特許第 5 7 5 8 1 1 9 (U S , A)

(58)調査した分野(Int.Cl. , D B 名)

G06F 12/00

G06F 12/08

G06F 13/00