

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5237449号
(P5237449)

(45) 発行日 平成25年7月17日(2013.7.17)

(24) 登録日 平成25年4月5日(2013.4.5)

(51) Int.Cl.		F I		
G06F 11/30	(2006.01)	G06F 11/30		A
G06F 11/34	(2006.01)	G06F 11/30	305A	
		G06F 11/34		S

請求項の数 24 (全 15 頁)

(21) 出願番号	特願2011-516741 (P2011-516741)	(73) 特許権者	510340791
(86) (22) 出願日	平成21年6月26日 (2009.6.26)		エクサフロップ・エルエルシー
(65) 公表番号	特表2011-526395 (P2011-526395A)		アメリカ合衆国・カリフォルニア・940
(43) 公表日	平成23年10月6日 (2011.10.6)		43・マウンテン・ビュー・アンフィシア
(86) 国際出願番号	PCT/US2009/048966		ター・パークウェイ・1600
(87) 国際公開番号	W02009/158683	(74) 代理人	100108453
(87) 国際公開日	平成21年12月30日 (2009.12.30)		弁理士 村山 靖彦
審査請求日	平成24年6月26日 (2012.6.26)	(74) 代理人	100064908
(31) 優先権主張番号	61/076,104		弁理士 志賀 正武
(32) 優先日	平成20年6月26日 (2008.6.26)	(74) 代理人	100089037
(33) 優先権主張国	米国 (US)		弁理士 渡邊 隆
早期審査対象出願		(74) 代理人	100110364
			弁理士 実広 信哉

最終頁に続く

(54) 【発明の名称】 データセンタの熱監視

(57) 【特許請求の範囲】

【請求項1】

複数のコンピュータのうちのあるコンピュータの温度をそれぞれ示す複数の測定値を、通信ネットワークを介して中央電子監視システムで受信するステップと、

前記測定値に基づいて前記複数のコンピュータのそれぞれについて熱的健全性値を計算するステップとを備え、

前記熱的健全性値を計算するステップは、前記コンピュータのプラットフォームタイプを標準化するステップを含む、コンピュータ実施の方法。

【請求項2】

前記熱的健全性値を計算するステップが前記測定値と格納されたしきい値とを比較するステップを含む、請求項1に記載の方法。

10

【請求項3】

前記測定値が温度、前記コンピュータの使用電力、前記コンピュータの利用率、または前記コンピュータのファンの回転速度のうちの一つまたは複数を含む、請求項1に記載の方法。

【請求項4】

異なるタイプの複数の測定値を受信するステップをさらに備え、前記熱的健全性値を計算するステップが前記複数の測定値を結合するステップを含む、請求項3に記載の方法。

【請求項5】

前記測定値が、前記コンピュータの中央処理装置、前記コンピュータのディスクドライ

20

ブ、前記コンピュータのメモリモジュール、前記コンピュータのマザーボード、または前記コンピュータ付近の空気のうちの一つまたは複数の温度測定値を備える、請求項1に記載の方法。

【請求項6】

前記熱的健全性値が0から1の範囲または0から10の範囲の値をとる無次元値である、請求項1に記載の方法。

【請求項7】

前記熱的健全性値についての記述子を、記述子の有限集合から選択するステップをさらに備える、請求項1に記載の方法。

【請求項8】

前記熱的健全性値がしきい値を超えるとアラームを生成するステップ、または、電力、冷却、または利用率のうちの一つまたは複数、前記熱的健全性値に基づいて調整するステップのうち少なくとも一つのステップをさらに備える、請求項1に記載の方法。

【請求項9】

前記複数のコンピュータのうち少なくともいくつか異なるプラットフォームであり、複数のコンピュータのそれぞれについて熱的健全性値を計算するステップがそれぞれの熱的健全性値が値の同じ範囲内に収まるように標準化するステップを含む、請求項1に記載の方法。

【請求項10】

値の前記範囲が、異なるプラットフォームにわたって一様であるコンピュータ挙動に対するマッピングをもたらす、請求項9に記載の方法。

【請求項11】

電力ドメイン、冷却ドメイン、論理ドメイン、および物理的位置ドメインのうちの一つまたは複数にわたって前記熱的健全性値を集約するステップをさらに備える、請求項1に記載の方法。

【請求項12】

プロセッサに、

複数のコンピュータのうちのあるコンピュータの温度をそれぞれ示す複数の測定値を、通信ネットワークを介して中央電子監視システムで受信させる手順と、

前記測定値に基づいて前記複数のコンピュータのそれぞれについて熱的健全性値を計算させる手順とを実行させ、

前記熱的健全性値を計算する手順は、前記コンピュータのプラットフォームタイプを標準化する手順を含む、コンピュータプログラム。

【請求項13】

前記熱的健全性値を計算する手順が前記測定値と格納されたしきい値とを比較する手順を含む、請求項12に記載のコンピュータプログラム。

【請求項14】

前記測定値が温度、前記コンピュータの使用電力、前記コンピュータの利用率、または前記コンピュータのファンの回転速度のうちの一つまたは複数、請求項12に記載のコンピュータプログラム。

【請求項15】

プロセッサに、異なるタイプの複数の測定値を受信する手順をさらに実行させ、前記熱的健全性値を計算する手順が前記複数の測定値を結合する手順を含む、請求項14に記載のコンピュータプログラム。

【請求項16】

前記測定値が、前記コンピュータの中央処理装置、前記コンピュータのディスクドライブ、前記コンピュータのメモリモジュール、前記コンピュータのマザーボード、または前記コンピュータ付近の空気のうちの一つまたは複数の温度測定値を備える、請求項14に記載のコンピュータプログラム。

【請求項17】

10

20

30

40

50

前記熱的健全性値が0から1の範囲または0から10の範囲の値をとる無次元値である、請求項12に記載のコンピュータプログラム。

【請求項18】

プロセッサに、前記熱的健全性値についての記述子を、記述子の有限集合から選択する手順をさらに実行させる、請求項12に記載のコンピュータプログラム。

【請求項19】

プロセッサに、前記熱的健全性値がしきい値を超えるとアラームを生成する手順、または、電力、冷却、および利用率のうちの1つまたは複数、前記熱的健全性値に基づいて調整する手順のうちの少なくとも1つの手順をさらに実行させる、請求項12に記載のコンピュータプログラム。

10

【請求項20】

前記複数のコンピュータのうちの少なくともいくつか異なるプラットフォームであり、複数のコンピュータのそれぞれについて熱的健全性値を計算する前記手順がそれぞれの熱的健全性値が値の同じ範囲に収まるように標準化する手順を含む、請求項12に記載のコンピュータプログラム。

【請求項21】

値の前記範囲が、異なるプラットフォームにわたって一様であるコンピュータ挙動に対するマッピングをもたらす、請求項20に記載のコンピュータプログラム。

【請求項22】

プロセッサに、電力ドメイン、冷却ドメイン、論理ドメイン、および物理的位置ドメインのうちの1つまたは複数にわたって前記熱的健全性値を集約する手順をさらに実行させる、請求項20に記載のコンピュータプログラム。

20

【請求項23】

複数のコンピュータのうちのあるコンピュータの温度をそれぞれ示す複数の測定値を、通信ネットワークを介して中央電子監視システムで受信するステップと、

前記測定値に基づいて前記複数のコンピュータのそれぞれについて熱的健全性値を計算するステップとを備え、

前記熱的健全性値を計算するステップは、前記コンピュータのプラットフォームタイプを標準化するステップを含み、

前記複数のコンピュータのうちの少なくともいくつか異なるプラットフォームであり、複数のコンピュータのそれぞれについて熱的健全性値を計算するステップがそれぞれの熱的健全性値が値の同じ範囲内に収まるように標準化するステップを含む、コンピュータ実施の方法。

30

【請求項24】

プロセッサに、

複数のコンピュータのうちのあるコンピュータの温度をそれぞれ示す複数の測定値を、通信ネットワークを介して中央電子監視システムで受信させる手順と、

前記測定値に基づいて前記複数のコンピュータのそれぞれについて熱的健全性値を計算させる手順とを実行させ、

前記熱的健全性値を計算する手順は、前記コンピュータのプラットフォームタイプを標準化する手順を含み、

40

前記複数のコンピュータのうちの少なくともいくつか異なるプラットフォームであり、複数のコンピュータのそれぞれについて熱的健全性値を計算する前記手順がそれぞれの熱的健全性値が値の同じ範囲に収まるように標準化する手順を含む、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本明細書は、コンピュータの熱監視(thermal monitoring)に関する。

【背景技術】

50

【0002】

マイクロプロセッサによって消費される電力は熱に変換される。単一のマザーボード上に搭載された1対のマイクロプロセッサは、200~400ワット、またはそれ以上の電力を使うことがある。データセンタ内のコンピュータについて説明するために、使用電力(power draw)を数千(または数万)倍すると、発熱の可能性を理解することができる。データセンタオペレータは、コンピュータを動作させるための電力に対してだけでなく、コンピュータを冷却するためにも支出しなければならない。熱を除去するためのコストが、大型データセンタの運営の主要コストとなる場合がある。

【0003】

さらに、マイクロプロセッサ、メモリ、またはディスクドライブなどのコンピュータのコンポーネントは、高熱にさらされると故障しやすい。熱を逃がすために、ファンがコンピュータのコンポーネントにわたって一定の空気流を送ることができ、周囲の空気への熱伝達を改善するために、いくつかのコンポーネント、特にマイクロプロセッサにヒートシンクを取り付けることができる。これらの予防措置にもかかわらず、特に高い利用率で動作しているとき、コンピュータコンポーネントは依然として過熱し、障害を起こすことがある。

【発明の概要】

【課題を解決するための手段】

【0004】

一態様では、本方法はコンピュータの温度を示す測定値を受信するステップ、およびその測定値に基づいてコンピュータの熱的健全性値を計算するステップを含む。

【0005】

本発明の実施形態は、以下の特徴のうちの1つまたは複数を含むことができる。熱的健全性値を計算するステップは、測定値と格納されたしきい値とを比較するステップを含む。測定値は温度、コンピュータの使用電力、コンピュータの利用率、またはコンピュータのファンの回転速度のうちの1つまたは複数であってもよい。異なるタイプの複数の測定値を受信してもよく、熱的健全性値を計算するステップは複数の測定値を結合してもよい。測定値は、コンピュータの中央処理装置、コンピュータのディスクドライブ、コンピュータのメモリモジュール、コンピュータのマザーボード、またはコンピュータ付近の空気の中の1つまたは複数の温度測定値であってもよい。熱的健全性値は、たとえば0から1の範囲または0から10の範囲の数などの無次元値であってもよい。熱的健全性値についての記述子は、記述子の有限集合から選択されてもよい。熱的健全性値を計算するステップは、コンピュータのプラットフォームタイプを標準化するステップを含むことができる。熱的健全性値がしきい値を超えると、アラームが生成されてもよい。電力、冷却、および利用率のうちの1つまたは複数が、熱的健全性値に基づいて調整されてもよい。複数のコンピュータのそれぞれからコンピュータの温度を示す測定値が受信されてもよく、複数のコンピュータのそれぞれについて熱的健全性値が計算されてもよい。複数のコンピュータのうちの少なくともいくつかは異なるプラットフォームでよく、複数のコンピュータのそれぞれについて熱的健全性値を計算するステップは、それぞれの熱的健全性値が値の同じ範囲内に収まるように標準化するステップを含むことができる。値の範囲は、異なるプラットフォームにわたって一様であるコンピュータ挙動にマッピングしてもよい。熱的健全性値は集約されてもよく、たとえば、複数のコンピュータについての熱的健全性値を生成するために複数のコンピュータの熱的健全性値が平均化されてもよい。電力ドメイン、冷却ドメイン、論理ドメイン、および物理的位置ドメインのうちの1つまたは複数にわたって熱的健全性値が集約されてもよい。

【0006】

他の態様では、マシン可読記憶メディア内に有形に実施されたコンピュータプログラム製品が、プロセッサにこれらの方法を実行させるための命令を備える。

【0007】

諸利点には以下のうちの1つまたは複数が含まれる。データセンタ内のコンピュータ

10

20

30

40

50

の熱的健全性を、オペレータが異なる集合体レベルで容易に評価できる。熱的健全性値を経時的に監視することができ、過去の挙動は将来の熱的健全性を予測するために使用できる。コンピュータが故障の危機にあることを熱的健全性変数が示すと是正措置をとることができ、したがって信頼性を改善し、データセンタのダウンタイムを減らす。

【0008】

1つまたは複数の実施形態の詳細は、添付の図面および以下の記述において説明する。他の諸特徴、諸目的、および諸利点は、記述および図面、ならびに特許請求の範囲から明らかになるであろう。

【図面の簡単な説明】

【0009】

【図1A】データセンタとして動作しているファシリティの側断面図である。

【図1B】データセンタとして動作しているファシリティの平面図である。

【図1C】データセンタ内のコンピュータのコンポーネントの概略図である。

【図2】監視コンピュータに温度情報を提供するように構成されたデータセンタのコンピュータの概略図である。

【図3】コンピュータコンポーネントについての故障温度のデータを収集する方法を示す流れ図である。

【図4】データパワーセンタを熱的に監視する方法を示す流れ図である。

【発明を実施するための形態】

【0010】

大規模インターネットサービス、およびそれをサポートすることが必要とされる大規模並列計算インフラストラクチャは、数千または数万の計算ノードで構成されているデータセンタファシリティ、それらの関連記憶階層、および相互接続インフラストラクチャなどの、ウェアハウスサイズの計算システムの設計を必要とすることがある。データセンタは異なるタイプのコンピュータなど、多くの異なるタイプのコンピュータ機器を含むことができ、一般的にそれぞれの機器は複数のより小さい電子コンポーネントで構成される。

【0011】

上述のように、コンピュータコンポーネントは熱を発生させる。たとえば、コンピュータコンポーネントの温度が短期的な性能低下または長期的な信頼性低下をもたらすのに十分なほど上昇する状態である熱イベントは、冷却プラントの故障、非常に多数のコンピュータを近接して展開すること(その結果、冷却システムの局所的な冷却能力よりも局所熱の方が大きくなる)、過剰利用、および季節変動などの様々な要因によって引き起こされうる。

【0012】

異なるコンピュータは高温に対する敏感性が異なる可能性があり、たとえばコンポーネントが故障し、低下した性能または長期的信頼性損失を被る温度、および故障の性質は、異なる構成のプラットフォームによってそれぞれ異なる。さらに、未処理の温度測定値によって表されるコンピュータへの脅威を理解できるのは専門家だけである可能性が高い。温度データを非専門家であるオペレータにとってより有益にするために、たとえば人間がリアルタイムで意思決定するために、または事前にプログラムされた応答を形成するために、コンピュータまたはコンピュータのグループの温度データを熱的健全性値(a thermal health value)に変換することができる。

【0013】

熱的健全性値は単一のマシンについて生成できるが、熱的健全性は多数のマシン、たとえばクラスタ全体を評価するために特に有益である。

【0014】

熱的健全性値はマージン、たとえばコンピュータまたはコンピュータのグループが低下した性能にどれだけ「近い」が示すことができる。データセンタの全体的な熱的健全性は、単一の最悪の場合または最大値よりもむしろ分散によって決定されうる。

【0015】

10

20

30

40

50

熱監視問題への序論として、例示的なデータセンタファシリティを説明する。

【0016】

図1Aおよび1Bは、データセンタとして機能する例示的ファシリティ10を示すための側面図および平面図である。ファシリティ10は密閉空間12を含み、基本的に建物全体、あるいは建物内の1つまたは複数の部屋を占めることができる。密閉空間12は、コンピュータ機器の多数(数十、数百、または数千)のラックを設置するのに十分に大きく、したがって数百、数千、または数万のコンピュータを収容できる。

【0017】

ラックマウントコンピュータのモジュール、たとえばケージ20は、アクセス通路24によって分離された列22内のスペースに配置される。それぞれのケージ20は、複数のラック26、たとえば4から8個のラックを含むことができ、それぞれのラックは複数のコンピュータ28、たとえばトレイを含む。

10

【0018】

図1Cはラックからのコンピュータの概略図である。一般的に、それぞれのコンピュータ28はマザーボードなどの回路基板50を含むことができ、その上に中央処理装置(CPU)52、メモリ54、1つまたは複数のディスクドライブ56、およびコンピュータをデータセンタ内の他のコンピュータのネットワークに接続するための入力/出力ポート58などの様々なコンピュータ関連コンポーネントが搭載されている。

【0019】

図1Aおよび1Bに戻ると、ファシリティはもう1つの熱交換処理を介してデータセンタから熱を除去するための冷却システム、たとえば部屋に冷気を入れるための空気調節システム、またはラックを越えて冷却液を運ぶ冷却コイル、およびラックマウントコンピュータに接続して、コンピュータと外部ネットワーク、たとえばインターネットとの間でデータを搬送するためのデータグリッド40(図1C参照)も含む。

20

【0020】

上述のように、マイクロプロセッサ、メモリ、またはディスクドライブなどのコンピュータのコンポーネントは高温にさらされると故障することがある。一般的にコンポーネントの故障は、コンポーネントの性能が変わる(一般的に減少する)第1温度ゾーン、コンポーネントのエラー率、たとえばディスクドライブの読み出し/書き込みエラーが著しく増加する、より高い第2温度ゾーン、および、コンポーネントが完全に故障する、さらに高い第3温度ゾーンなどの複数のモードを含む。コンポーネントが異なるモードの故障を示す温度は、実験的手段を通じて独立に特徴づけることができる。しかし、コンポーネントデューティサイクル、電圧、および製造上のばらつきを含む、性能に影響を与える他の問題点がありうる。

30

【0021】

温度に応じてコンポーネントが故障する順番は、異なるプラットフォームにわたって同様であることも異なることもあり、たとえばいくつかのプラットフォームでは温度が上昇すると最初にメモリが故障し、次にディスクドライブ、続いてCPUが故障することがある。他の順番も可能である。たとえばいくつかのプラットフォームでは、CPU性能が熱暴走の影響を最も受けやすく、次にディスクドライブ、続いてメモリが影響を受けやすい場合がある。しかし、故障が発生する正確な温度は、異なる構成のプラットフォームによって様々である。たとえば、異なるモデルのCPUは異なる温度で故障し始める。したがって、たとえ温度データが利用可能でも、温度データ自体は必ずしもマシンの熱的健全性に関する有益な情報を提供するわけではなく、ましてデータセンタ内で一般的に生じるような異なる構成を有するマシンの集合の熱的健全性に関する有益な情報を提供するわけではない。

40

【0022】

したがって、上述のように、データセンタの人間のオペレータがコンピュータまたはコンピュータのグループの熱状態の表示を有することは有益であり、コンピュータが故障しているか、または故障の危機にあるかを評価するために特定のプラットフォームの知識を

50

必要としない。故障には性能の低下、または回復不能な障害、あるいは両方を含むことができ、異なるタイプの故障は熱的健全性を決定する際に異なる重み付けを受けることができる。

【0023】

温度が上昇するにつれてコンポーネントが同じ順番の故障モードをたどる傾向があるという事実は、温度が上昇するにつれてエラーモードの一定の順序をもたらす傾向がある。したがって、温度が上昇するにつれて、コンピュータは熱的「健全性」のいくつかのゾーンをたどる傾向があり、たとえばより低温ではコンピュータは正常に動作するので「健全」と見なすことができ、より高温では一定のエラーが発生するのでコンピュータは「不健全」と見なすことができ、ひとたびコンピュータが完全に故障すると「故障した」と見なすことができる。さらに、エラーがより頻繁になるか、より多数のコンポーネントでエラーが発生し始めると、「不健全」段階はいくつかのレベルを含むことができる。さらに、温度がさらに上昇するとコンピュータが「不健全」段階に入ることを示す「危険」段階があってもよく、さらに高温では、コンピュータが完全な故障寸前であることを示す「限界」段階があってもよい。定義された独自の健全性レベル段階が、ほとんどいくつでもあってもよい。

10

【0024】

コンピュータの熱状態のこの一般的表示を「熱的健全性変数(thermal health variable)」と呼ぶことができ、それは「熱的健全性値」を有することができる。したがって、熱的健全性変数は、単一コンピュータまたはデータセンタ内のコンピュータの集合体の熱状態のメトリック、一般的に無次元メトリックであるが、コンピュータの特定の構成に依存しない。

20

【0025】

熱的健全性変数はコンピュータの構成の違いを超えて標準化できる。すなわち熱的健全性値は、複数のコンピュータ、たとえばデータセンタのすべてのコンピュータで共有されるスケールの範囲に収まるように個々のコンピュータについて計算され、その共有スケールの少なくとも端点(および場合によってはスケール内の少なくともいくつかの midpoint)がコンピュータの同様の挙動を表す。このような熱的健全性値は、マシンの温度を示す最近の測定と、その挙動が発生する温度の過去の実験的測定との組合せから計算されうる。データセンタは異種機器の集合なので、これらの違いを超えて標準化することはクラスタ全体の熱的健全性を評価するための有効なパラメータの提供に役立つ。基本的に、熱的健全性変数はコンピュータまたはコンピュータのグループの健全性を表すが、個々のマシンの仕様とは無関係である。

30

【0026】

場合によっては、システムは最も不健全なコンポーネントと同程度の健全性でしかない。たとえば、CPUが「故障した」場合、コンピュータ全体が故障していることになる。場合によっては、熱的健全性はコンポーネント熱的健全性の集合体である。たとえばデータセンタレベルでの熱的健全性は、個々のマシンについての熱的健全性の集合体である。いくつかの実施形態では、コンピュータの集合体の熱的健全性を提供するために、集合体内の個々のコンピュータの熱的健全性を単に平均化することによって熱的健全性を集約(aggregate)することができる。

40

【0027】

図2は熱的健全性値を決定するためのシステムの概略図である。データセンタ内のコンピュータ28のうちの少なくともいくつかは、ネットワーク、たとえばデータグリッド40を通じて汎用コンピュータ、たとえばワークステーションなどの監視コンピュータ60に接続されている。

【0028】

これらのコンピュータ28はそれぞれコンポーネントのうちの少なくとも1つの温度を測定する少なくとも1つのセンサを含む。たとえばコンピュータ28はCPU、メモリ、ハードドライブ、またはマザーボードの温度を測定するためにセンサを使用できる。さらにコン

50

コンピュータ28は、マシン自体を冷却するためにマシンがどれほど忙しく働いているかを示すファン速度などの、熱的健全性を示す他のデータを累算できる。たとえば2つのマシンの温度は同じだが、第1マシンのファン速度が最大可能値で第2マシンのファン速度が標準値の場合、第1マシンは悪条件において熱的健全性を維持する能力がより少ないため、このマシンについての熱的健全性の方が少々悪いと報告される。さらに、コンピュータ28付近の空気などの環境の温度を測定するためにラックまたはケージに設置されたセンサがあってもよく、このデータは熱的健全性を決定する際に使用できる。さらに、温度は電力および利用率に対応すると仮定できるので、コンピュータは使用電力および/または利用率についてのデータを収集できる。コンピュータ28は他のトレイレベルの熱センサから、および他のラックまたはケージレベルのセンサ、たとえば吸気温(EAT)、吐気温(LAT)、吸水温(EWT)、および吐水温(LWT)を測定するラックまたはケージのための冷却インフラストラクチャを監視するセンサなどから熱データを受信できる。

10

【0029】

いくつかの実施形態では、それぞれのコンピュータが、個々のコンポーネントの仕様およびセンサからの利用可能データに基づいてそれ自体の熱的健全性を計算する。いくつかの実施形態では、いくつかまたはすべてのコンピュータが測定値を監視コンピュータ60に送信し、監視コンピュータ60は受信したデータに基づいて熱的健全性を計算する。監視コンピュータ60は、他のトレイレベル熱センサから、およびハードウェア健全情報の他のソースから熱データを受信できる。いくつかの実施形態では、コンピュータ28によって計算された熱的健全性値は監視コンピュータ60に送信され、監視コンピュータはコンピュータ28の集合体についての熱的健全性を計算できる。

20

【0030】

図3はコンピュータの故障温度を決定する方法を示す流れ図である。初めに、コンピュータコンポーネントの故障温度を含む温度への1つまたは複数のコンピュータの反応が実験的に決定される(ステップ102)。たとえば、知られているコンポーネントを有するテストコンピュータが温度制御された環境、たとえばオープン内に設置されうる。環境の温度が上昇する間、テストコンピュータを動作させる。テストコンピュータの性能が監視され、たとえばテストコンピュータはテストルーチンを実行してエラーを報告し、その間熱電対または他のセンサがコンポーネントの温度を測定できる。このデータから、コンポーネントが故障する温度が決定できる。コンポーネントおよびコンポーネントの故障温度を示すデータはデータベースに格納できる(ステップ104)。

30

【0031】

プラットフォームのタイプごとに、そのプラットフォームに利用可能なセンサ測定値と熱的健全性値とを関連付けるために機能を生成できる。これらの機能もプラットフォームを機能に関連付けるデータベースに格納できる。データセンタに設置されたコンピュータは、そのプラットフォームに対応する機能を含むように(設置前または設置後に)構成されうる。関連するしきい値を設定するための、何が特定のコンポーネントまたはコンピュータの「故障」を構成するかについての最初の決定は、熱的健全性監視システムの設計者によるいくつかの判断を含むことができるが、マッピングに一貫性があり、熱的健全性値の範囲の一方が良好な性能を表し、他方がほぼ完全な故障を表す限り、熱的健全性変数は有益な情報を提供する。さらに、異なるプラットフォームが同じ熱的健全性値を有する場合は同じ性能劣化を受けるように、同様のコンポーネントについてある程度一貫性のある基準を提供することが可能である。たとえば、メモリなどの特定のタイプのコンポーネントの一定のエラー率(たとえば、エラー/秒に関して)は、一定の熱的健全性値、たとえば「健全(healthy)」および「不健全(unhealthy)」などの異なる所望の記述子間のしきい値に一貫してマッピングできるように設計者によって選択できる。

40

【0032】

図4はデータセンタを運営する間に熱的健全性値を決定する方法を示す流れ図である。コンピュータのうちのいくつかまたはすべてが、コンポーネントの温度を示すデータを監視コンピュータに送信する(ステップ110)。コンピュータごとに、監視コンピュータは受

50

信した温度データおよび以前に測定したコンピュータ内のコンポーネントの故障温度に基づいて熱的健全性変数を計算する(ステップ112)。いくつかの実施形態では、コンピュータは機能を格納してそれ自体の熱的健全性を個々のコンポーネントの仕様および利用可能センサに基づいて計算し、次いで計算した熱的健全性を監視コンピュータに送信する。

【0033】

熱的健全性変数は、温度センサ、ファン速度、電力、および利用率を含む、それぞれのコンピュータの様々なセンサからのデータから計算できる。異なるタイプのプラットフォームにわたって比較できるようにするために、同様の熱的健全性値での同様のマシンの挙動を示すために熱的健全性変数が標準化される。たとえば、最も低い熱的健全性値、たとえば0では、すべてのマシンが危険性なしに温度を上昇させるための相当のマージンがあるというカテゴリに入り、最も高い熱的健全性値、たとえば10では、すべてのマシンが回復不能な障害というカテゴリに入ることになる。以前示したように、マシンの挙動は以前に測定された故障温度と相関関係がある(ステップ112)。

10

【0034】

計算された熱的健全性変数は無次元メトリック、たとえば数であってもよい。いくつかの実施形態では、熱的健全性変数は可能値の有限小数を有する。たとえば、熱的健全性変数は0から10または1から5の範囲の整数でよく、たとえば数が低いほどより健全であることを示す。いくつかの実施形態では、熱的健全性変数は非常に多数の可能値を有する数、たとえば、コンピュータ内のデータの言葉によって表される数でよく、このような実施形態では熱的健全性値はたとえば0から1または0から10の範囲の実数であってもよい。

20

【0035】

熱的健全性変数はオペレータに表示することができる。いくつかの実施形態では、熱的健全性変数は関連テキスト記述、たとえば「良い」、「OK」、または「悪い」で表示できる。他の記述子「相当のマージンがある」、「標準操作状態」、「標準操作状態を超える」、「低下した性能」、「エラー率上昇」および「回復不能な故障」を含むことができる。いくつかの実施形態では、数値は表示されずにテキスト記述だけが表示される。

【0036】

熱的健全性変数が可能値の有限小数を有する場合、値との1対1ベースでテキスト記述がマッピングされてもよく、熱的健全性値それ自体がテキスト記述(数ではなく)でもよい。熱的健全性変数が事実上連続的である場合、値の様々な範囲をテキスト記述子に関連付けることができる。たとえば、熱的健全性変数が0と1との間のスケールの場合、0から0.75の値はマシンが予測された温度範囲内で動作していることを示すことができ、0.75から0.9の値はマシンが予測された制御可能な温度範囲外で動作していることを示すことができ、0.9から1の値はマシンが抑制または最大許容可能温度に達した可能性があることを示すことができ、および1の値はマシンが最大許容可能温度に達したことを示す。

30

【0037】

一般的に、熱的健全性変数を計算するために、受信された温度データが、以前に測定されたテストコンピュータの故障温度がしきい値を決定する機能に供給される。熱的健全性変数を決定するために、受信された温度はしきい値と比較される。たとえば、しきい値の第1セットのうちのいずれかを超えることは熱変数を「OK」と設定し、しきい値の第2セットのうちのいずれかを超えることは熱変数を「悪い」と設定する。機能はかなり複雑な場合があり、温度測定としきい値との間の違いからの値の計算、およびコンポーネントのタイプまたはしきい値に基づく違いの重み付けを含む。

40

【0038】

計算された値が複数のコンピュータで共有されるスケール内に入る限り、特定のマシンのコンポーネントの許容差によってマシンごとに異なる機能内の定数(たとえば実験的に測定されたしきい値)に加えて、機能内で使用される変数(たとえば、温度または利用率などの測定のタイプ)、および機能内の変数の関係はマシンごとに異なる。たとえば、あるコンピュータは熱的健全性値を計算するために温度測定を使用することがあり、他のコンピュータは熱的健全性値を計算するために利用率を使用することがある。値が共有された

50

スケールを使用する限り、コンピュータの集合体の熱的健全性を決定するために、値を集約、たとえば単に平均化することができる。いくつかの実施形態では、コンピュータの集合体の熱的健全性について、あるコンピュータが他のコンピュータよりも大きく重み付けされる。

【0039】

いくつかの実施形態では、熱的健全性値は他の数値測定および実験的に測定された挙動のしきい値から計算された数値である。1例を挙げると、0が最大可能マージンを示し、1がマシンは最高/最悪挙動ゾーンにある、またはそれ以上であることを示すものとして、0と1との間の標準化された無次元熱的健全性値を導くためのアルゴリズムは以下の通りである。すなわち、

$$THV=1-[T_{allowable}-T]/M_{max}$$

上式で、THVは熱的健全性値であり、 $T_{allowable}$ は最高/最悪挙動変更前の最大許容可能温度であり、Tはセンサから報告された温度であり、 M_{max} は0と1との間の結果を標準化するために使用される予測された最大許容可能マージンである。

【0040】

監視コンピュータはデータセンタを通じて個々のコンピュータと通信する。任意選択で、監視コンピュータは熱的健全性データをさらなる処理のために集約でき、たとえば監視コンピュータは任意選択で1つまたは複数のコンピュータの集合体についての熱的健全性変数を計算できる(ステップ114)。集合体の熱的健全性値は、集合体内の個々のコンピュータの熱的健全性変数の平均として計算できるが、累積分布関数または平均二乗などの他の技法が適用されてもよい。

【0041】

集合体は位置ベースの集合体でよく、すなわちコンピュータの物理的位置に基づいてもよい。たとえば、熱的健全性値は所与のラック、モジュール、列、クラスタ、またはデータセンタ内のコンピュータのグループについて決定されてもよい。

【0042】

集合体はインフラストラクチャベースの集合体でよく、すなわち共通インフラストラクチャコンポーネントを共有するコンピュータに基づいてもよい。たとえば、熱的健全性値は特定の電力または冷却ドメイン内のコンピュータのグループについて決定されてもよい。

【0043】

集合体はサービスベースの集合体でよく、すなわちコンピュータがデータセンタ内で実行する機能に基づいてもよく、コンピュータがサービスする顧客に基づいてもよい。たとえば、熱的健全性値は検索結果、電子メール、または地図作成をサービスするデータセンタ内のコンピュータのグループについて決定されてもよい。集合体は論理ネットワークグルーピングでよく、たとえば特定のネットワーキングドメインまたはサブドメイン内のコンピュータに基づいてもよい。

【0044】

集合体の熱的健全性を累積分布関数、すなわち所与の熱的健全性値またはそれ以下で動作しているコンピュータの割合として提示することが有益である場合がある。

【0045】

熱的健全性変数が決定された後で、たとえば制御ソフトウェアによって自動的に、熱故障を防止するためにコンピュータの性能を調整するために使用されうる。たとえば熱的健全性変数が一定のしきい値を超えるか警告を示すと、たとえば集合体内のコンピュータの利用率を減らして電力使用および温度を下げるために照会を他のコンピュータに転送することによって、集合体内のコンピュータを抑制することができる。別の例を挙げると、熱的健全性変数は冷却システムまたは他の冷却リソースを制御するために使用でき、たとえば熱的健全性値が一定のしきい値を超えた場合、コンピュータの温度を下げて故障の危険性を減らすために熱除去率を上げることができる。逆に、熱的健全性値が一定のしきい値を下回る場合、これはコンピュータがより高温で動作するための多少のマージンを有する

10

20

30

40

50

ことを示すことができ、冷却コストを下げるために熱除去率を下げるができる。極限状況では、臨界しきい値を超える1つまたは複数のコンピュータをコンピュータコンポーネントへの回復不能な故障を回避するために自動的にシャットダウンできる。

【0046】

監視コンピュータは、熱的健全性変数がしきい値を超えると、たとえばある事前設定値を超えると、データセンタのオペレータに警告するためのアラームを自動的に生成できる。たとえば、ラックレベルの熱的健全性値は、所与のラック内のすべてのコンピュータの平均熱的健全性値として生成されうる。アラーム状況の例を挙げると、事前設定パーセンテージ、たとえばクラス内の4%から10%のラックが事前設定値、たとえば0から1のスケールで0.75を超える熱的健全性値を有する場合、アラームが生成されうる。機能は、いくつかの温度がしきい値に近い場合に熱的健全性変数が警告を示すようにセーフティマージンを含むことができる。

10

【0047】

熱的健全性変数は、マシン配置について決定を下すためにも使用されうる。たとえば、ラックレベルで集約された熱的健全性値を比較することによって、オペレータは追加の熱ロードを処理するためにより多い容量を有するのはどのラックが決定できる(たとえば、熱的健全性値が最も低いラックが追加の熱ロードを最もよく処理できるという仮定の下に)。したがって、追加の熱ロードを処理するためにより多い容量を有するラックへの新しいマシンの配置が優先されうる。

【0048】

一部の顧客はあるタイプの故障が別のタイプの故障よりも悪いと見なす場合があるので、故障のクラスの関連重み付け(たとえば低下した性能対回復不能な故障)が顧客のために、または顧客によって設定されうる。

20

【0049】

熱変化率、たとえば熱変数の時間導関数をコンピュータまたはコンピュータの集合体について計算できる。高い変化率は、マシンまたはマシンのグループに問題が発生していることを示すことができる。この熱変化率は、上述のように熱的健全性変数について標準化されうる。

【0050】

熱的健全性値は、たとえ特定のプラットフォーム上で実験的測定が行われていない場合でさえ計算してもよい。プラットフォームの特性、たとえばCPUのタイプ、メモリ、およびディスクドライブの数およびタイプが格納されていると仮定すると、ならびに同様の装置、たとえば同じまたは同様のコンポーネント上で実験的測定が行われたと仮定すると、おおよその故障温度の計算が可能であってもよい。たとえば、故障温度は他のテストコンポーネント、たとえば同じメーカーのコンポーネントの測定された故障温度から補間されうる。

30

【0051】

上記で説明した処理の様々なステップ、特に熱的健全性値の計算は、コンピュータによって実行される。熱的健全性変数の危険レベルの上昇に応じてコンピュータをシャットダウンすることなどのいくつかのステップは、熱的健全性変数の受信に応じて、たとえば表示されたときに人間のオペレータによって実行されてもよく、たとえば監視コンピュータの制御ソフトウェアによって自動的に実行されてもよい。

40

【0052】

マシン熱的健全性の利点は、標準化された値が生成されるので、オペレータまたは熱的健全性変数を使用する制御ソフトウェアを設計しているプログラマはプラットフォームのタイプについて知る必要がない点である。多くの次元の違いを超えて熱的健全性を集約することができ、コンピュータの健全性についての情報を収集するために多くの異なる統計方法を適用できる。熱的健全性変数は、コンピュータに加えてネットワークスイッチ、スマートラック、または電力装置などの、高温の結果故障する可能性がある他の形式のコンピューティング機器に適用されてもよい。

50

【 0 0 5 3 】

本明細書で説明した実施形態およびすべての機能操作は、デジタル電子回路、あるいは本明細書で開示した構造的な手段およびその構造的同等物またはそれらの組合せを含む、コンピュータソフトウェア、ファームウェア、またはハードウェアで実施できる。本発明の実施形態は、プログラム可能プロセッサ、コンピュータ、あるいは複数のプロセッサまたはコンピュータなどのデータ処理装置によって実行されるために、またはその動作を制御するために、1つまたは複数のコンピュータプログラム製品、すなわち情報媒体、たとえばメモリまたはディスクなどのマシン可読記憶メディアに有形に実施され、あるいは伝搬信号に実施された1つまたは複数のコンピュータプログラムとして実施できる。

【 0 0 5 4 】

本発明のいくつかの実施形態を説明してきた。しかし、本発明の趣旨および範囲から逸脱することなしに様々な修正が行われてもよいことが理解されよう。したがって、他の実施形態は添付の特許請求の範囲内である。

【 符号の説明 】

【 0 0 5 5 】

- 10 ファシリティ
- 12 密閉空間
- 20 ケージ
- 22 列
- 24 アクセス通路
- 26 ラック
- 28 コンピュータ
- 40 データグリッド
- 50 回路基板
- 52 中央処理装置 (CPU)
- 54 メモリ
- 56 ディスクドライブ
- 58 入力/出力ポート
- 60 監視コンピュータ

10

20

【図 1 A】

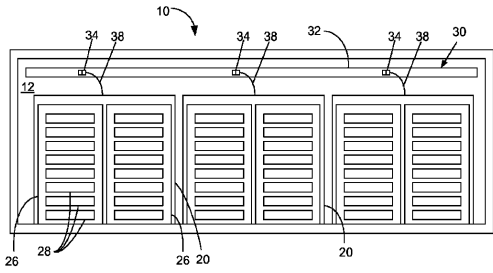


FIG. 1A

【図 1 B】

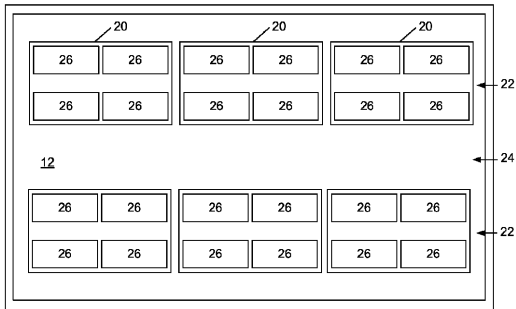


FIG. 1B

【図 1 C】

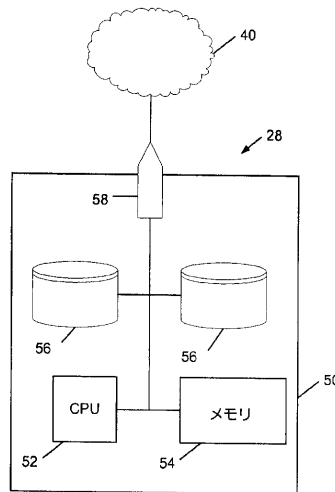


FIG. 1C

【図 2】

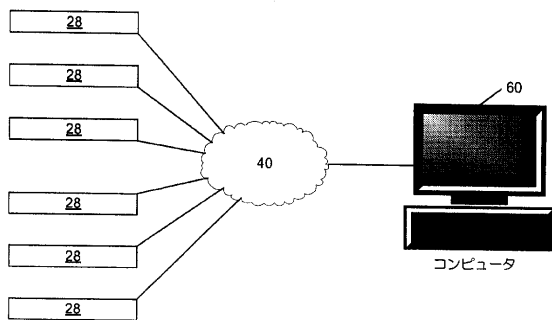


FIG. 2

【図 3】

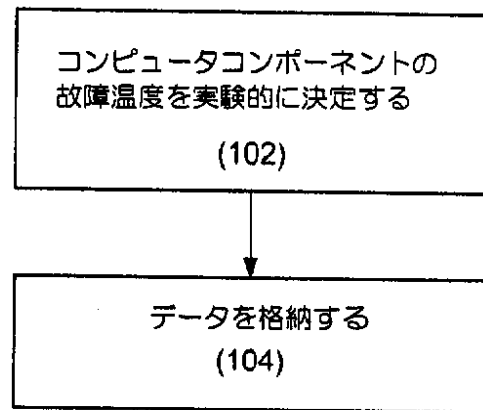


FIG. 3

【図4】

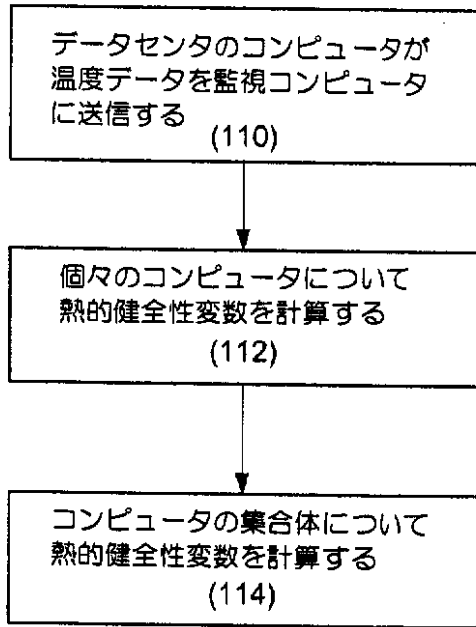


FIG. 4

フロントページの続き

- (72)発明者 デーヴィッド・ダブリュー・スタイヴァー
アメリカ合衆国・カリフォルニア・95050・サンタ・クララ・チャペル・ドライブ・1255
- (72)発明者 ショーン・ルイス
アメリカ合衆国・カリフォルニア・94110・サン・フランシスコ・ハンプシャー・ストリート
・1040
- (72)発明者 グレゴリー・ピー・インウォール
アメリカ合衆国・カリフォルニア・94085・サニーヴェール・ノース・ウルフ・ロード・35
5・アパートメント・#633

審査官 多賀 実

- (56)参考文献 米国特許出願公開第2008/0046766 (US, A1)
特開2007-179437 (JP, A)
特開平10-143385 (JP, A)
特表2004-507806 (JP, A)
温度や電流値のセンサーをリモート監視するラックマネジメントユニット RMS9000S /
9000, インターネットデータセンター完全ガイド 2007年秋号, 日本, 株式会社インブ
レスR&D, 2007年10月29日, p. 52

(58)調査した分野(Int.Cl., DB名)

G06F1/00
G06F1/20
G06F9/46-9/54
G06F11/00
G06F11/22
G06F11/30-11/34
G06F13/00