

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-121338

(P2019-121338A)

(43) 公開日 令和1年7月22日(2019.7.22)

(51) Int.Cl.		F I		テーマコード (参考)
G06F 1/20 (2006.01)		G06F 1/20	D	
G06F 1/16 (2006.01)		G06F 1/16	3 1 2 M	

審査請求 有 請求項の数 10 O L 外国語出願 (全 16 頁)

(21) 出願番号	特願2018-127326 (P2018-127326)	(71) 出願人	508018934 廣達電腦股▲ふん▼有限公司 台湾桃園市龜山區文化二路188號
(22) 出願日	平成30年7月4日(2018.7.4)	(74) 代理人	110000154 特許業務法人はるか国際特許事務所
(31) 優先権主張番号	62/613, 855	(72) 発明者	周 樂生 台湾桃園市龜山區文化二路211號
(32) 優先日	平成30年1月5日(2018.1.5)	(72) 発明者	張 添榮 台湾桃園市龜山區文化二路211號
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	15/968, 382		
(32) 優先日	平成30年5月1日(2018.5.1)		
(33) 優先権主張国	米国 (US)		

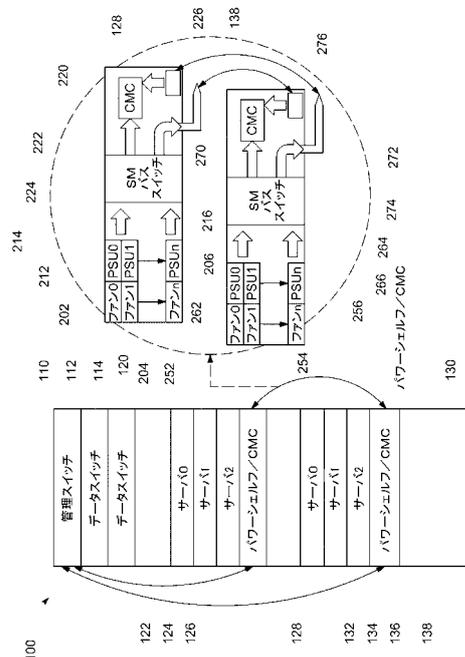
(54) 【発明の名称】 機器ラック及び機器ラックからの状態報告を保証する方法

(57) 【要約】 (修正有)

【課題】 シャーシ管理コントローラ(CMC)が故障するかオフラインの場合でも機器ラック及び機器ラックからの状態報告を保証する方法を提供する。

【解決手段】 リモート管理者は、CMC 220へのネットワーク接続を介して管理スイッチ110からパワーシェルフ状態を取得するコマンドを送信できる。次に、CMC 220は、管理スイッチ110を介してリモート管理者からの要求に応じてパワーシェルフ状態を報告する。CMC 220及び270は、それぞれのパワーシェルフ128及び138からの電力状態情報を交換できる。CMC 220又は270の一方が、他方のCMCが通信によるサービスを停止していることを発見した場合、そのCMCは、監視機能を引き継ぐ。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

機器ラックであって、
管理コントローラと、

第 1 の電源ユニット、前記第 1 の電源ユニットに結合された第 1 のシステムバス、前記第 1 のシステムバスに結合された第 1 のシャーシ管理コントローラ、及び第 1 のネットワーク装置を含み、かつ、前記第 1 の電源ユニットが前記第 1 のシステムバスに状態データを送る第 1 のシャーシと、

第 2 の電源ユニット、前記第 2 の電源ユニットに結合された第 2 のシステムバス、前記第 2 のシステムバスに結合された第 2 のシャーシ管理コントローラ、及び第 2 のネットワーク装置を含み、かつ、前記第 2 の電源ユニットが前記第 2 のシステムバスに状態データを送る第 2 のシャーシとを備え、

前記第 1 のシャーシ管理コントローラが、前記第 1 のシステムバスと通信して前記第 1 の電源ユニットから状態データを取得し、前記第 2 のシステムバスと通信して前記第 2 の電源ユニットから状態データを取得し、前記第 1 のシャーシ管理コントローラが、前記管理コントローラと通信して前記状態データを中継し、

前記第 2 のシャーシ管理コントローラが、前記第 1 のシステムバスと通信して前記第 1 の電源ユニットから状態データを取得し、前記第 2 のシステムバスと通信して前記第 2 の電源ユニットから状態データを取得し、

前記管理コントローラと前記第 2 のシャーシ管理コントローラの間の通信を確立して前記状態データを中継する働きをする機器ラック。

【請求項 2】

前記第 1 の電源ユニットが、複数の電源ユニットの 1 つであり、前記複数の電源ユニットがそれぞれ、前記第 1 のシステムバスに状態データを提供する、請求項 1 に記載の機器ラック。

【請求項 3】

前記第 1 のネットワーク装置が、前記第 1 のシャーシ上の複数のネットワーク装置の 1 つである、請求項 1 に記載の機器ラック。

【請求項 4】

前記第 1 のネットワーク装置に接続されたデータスイッチを更に有し、前記データスイッチが前記第 1 及び第 2 のネットワーク装置の一方と通信する、請求項 1 に記載の機器ラック。

【請求項 5】

前記シャーシ管理コントローラが、インターインテグレートッドサーキットプロトコル (I 2 C)、ユニバーサル非同期レシーバ/トランスミッタ (U A R T) 及びネットワークの少なくとも 1 つを介して管理スイッチと通信する、請求項 1 に記載の機器ラック。

【請求項 6】

前記第 1 のネットワーク装置が、サーバである、請求項 1 に記載の機器ラック。

【請求項 7】

前記第 1 の電源ユニットと前記第 1 のシャーシ管理コントローラが、パワーシェルフに取り付けられた、請求項 1 に記載の機器ラック。

【請求項 8】

前記管理スイッチが、前記第 1 及び第 2 のシャーシユニットから受け取った状態データのコンパイルに基づいてラック状態を決定する働きをする管理コントローラを有する、請求項 5 に記載の機器ラック。

【請求項 9】

前記管理コントローラが、前記第 1 のシャーシ管理コントローラからの状態データの通信の失敗を決定する働きをし、更に、前記管理コントローラは、前記第 1 のシャーシ管理コントローラからの状態データの通信の失敗が決定されたときに前記第 2 のシャーシ管理コントローラとの通信を確立する働きをする、請求項 1 に記載の機器ラック。

10

20

30

40

50

【請求項 10】

管理コントローラと、第 1 の電源ユニット、第 1 のシステムバス、第 1 のシャーシ管理コントローラ及び第 1 のネットワーク装置を有する第 1 のシャーシと、第 2 の電源ユニット、第 2 のシステムバス、第 2 のシャーシ管理コントローラ、及び第 2 のネットワーク装置を有する第 2 のシャーシを有する機器ラックからの状態報告を保証する方法であって、

し、
前記第 1 の電源ユニットから前記第 1 のシステムバスに状態データを提供するステップと、

前記第 2 の電源ユニットから前記第 2 のシステムバスに状態データを提供するステップと、

前記第 1 のシャーシ管理コントローラと前記第 1 のシステムバスの間の通信を確立して前記第 1 の電源ユニットから状態データを取得するステップと、

前記第 1 のシャーシ管理コントローラと前記第 2 のシステムバスの間の通信を確立して前記第 2 の電源ユニットから状態データを取得するステップと、

前記第 1 のシャーシ管理コントローラを介して前記管理コントローラに状態データを中継するステップと、

前記第 2 のシャーシ管理コントローラと前記第 2 のシステムバスの間の通信を確立して前記第 2 の電源ユニットから状態データを取得するステップと、

前記第 2 のシャーシ管理コントローラと前記第 1 のシステムバスの間の通信を確立して前記第 1 の電源ユニットから状態データを取得するステップと、

前記第 1 のシャーシ管理コントローラから状態データが中継されないという決定に基づいて、前記第 2 のシャーシ管理コントローラを介して前記管理コントローラに状態データの中継を開始するステップとを含む方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、一般に、複数のネットワーク装置を動作させるためのシャーシに関する。より詳細には、本開示の態様は、シャーシ取り付けネットワーク装置の動作を監視する複数のシャーシ管理コントローラのネットワーク接続に関する。

【背景技術】

【0002】

計算処理用途のためのクラウドの出現によって、リモート接続されたコンピュータ装置ユーザによってアクセスされるデータを記憶しアプリケーションを実行する、データセンタとして知られるオフサイト施設の需要が増えてきた。そのようなデータセンタは、典型的には、リモートコンピュータユーザが好都合にアクセスできるように、データを記憶し管理する膨大な数のサーバ、スイッチ及び記憶装置を有する。典型的なデータセンタは、電力及び通信接続が付随する物理ラック構造物を有する。このラックは、データセンタの部屋全体に配列される。各ラックは、サーバ、スイッチ、記憶装置などの複数の装置を保持できる横向きスロット又はシャーシを有するフレームを有する。最新のデータセンタでは、そのようなラック構造物内に積み重ねられた多数のそのような装置が見られる。例えば、幾つかのデータセンタには、何万台ものサーバと、それに付随する記憶装置とネットワークスイッチがある。したがって、典型的なデータセンタは、何万又は何十万の個別のラック内に何万台更には何十万台もの装置を含みうる。データセンタは、典型的には、機器の適切な動作を監視し保証するためにコントロールセンタ内に管理システムを有する。効率的に管理するため、管理者は、データセンタ内の各ラック内の機器の瞬間の状態に関する知識に依存する。

【0003】

図 1 に、典型的なラックシステム 10 を示す。ラックシステム 10 は、パワーシェルフ 14 に結合されたシャーシ管理コントローラ (CMC) 12 を有する。パワーシェルフ 14 は、幾つかの電源ユニットとそれに対応する冷却用ファンを有する。パワーシェルフ 1

10

20

30

40

50

4 は、ネットワーク装置バンク 16 及び 18 に接続され、バンク 16 及び 18 内のネットワーク装置のそれぞれに電力を提供する。この例では、ネットワーク装置バンク 16 及び 18 内のネットワーク装置は、ブレードサーバでよい。ラックシステム 10 は、また、管理スイッチ 20 と 2 台のデータスイッチ 22 及び 24 を含む。管理スイッチ 20 は、パワーシェルフ 14 上の電源を含むバンク 16 及び 18 に格納されたネットワーク装置の動作を監視する。データスイッチ 22 及び 24 は、バンク 16 及び 18 に格納されたネットワーク装置間のデータ通信を提供する。

【0004】

CMC 12 は、バンク 16 及び 18 内のネットワーク装置を管理する。CMC 12 は、設定及び監視タスクを実行し、バンク 16 及び 18 内のネットワーク装置への電力を制御し、バンク 16 及び 18 内のネットワーク装置に警告を提供する。CMC 12 は、マイクロプロセッサとメモリを有する。CMC 12 は、パワーシェルフ 14 上の電源によって電力供給される。

10

【0005】

CMC 12 は、データセンタ内のラック 10 の管理に重要な役割を果たす。CMC 12 は、データセンタ管理ソフトウェアのためにラック 10 からラックベースでの動作情報を収集する。動作情報には、(1) ラック全消費電力、(2) 各ネットワーク装置からの IP アドレスリスト、(3) 全体及び個別のラック構成要素の健全性に関する状態、並びに (4) ラックベースでのファン及び温度状態が含まれうる。そのような情報は、CMC 12 や管理スイッチ 20 上のラック管理コントローラなどのラックの全ての管理ユニットに対して、ラック電力キャッピングやラック ID 更新などの機能を実行するためにデータセンタ管理ソフトウェアによって使用されうる。したがって、データセンタ管理者は、そのような情報の収集によってデータセンタの各ラックシステム内の装置の動作をリモートで監視できる。

20

【0006】

CMC 12 のためにパワーシェルフ 14 上の構成要素との通信を実施する幾つかの周知の方法がある。例えば、CMC 12 は、インターインテグレートドサーキットプロトコル (I2C)、ユニバーサル非同期型レシーバ/トランスミッタ (UART)、又はパワーシェルフ 14 からの動作データを管理スイッチ 20 と通信するネットワークを使用できる。

30

【0007】

動作では、CMC 12 は、パワーシェルフ 14 の状態を定期的にポーリングする。リモート管理者は、管理スイッチ 20 などの管理スイッチを介してコマンドを送ってパワーシェルフ状態を取得できる。次に、CMC 12 は、リモート管理者からの要求に応じてパワーシェルフ 14 の状態を報告する。警告が出されたとき、CMC 12 は、潜在的故障などの潜在的状況への対応を強化するために、パワーシェルフ 14 の状態をリモート管理ノードに能動的に報告する。

【先行技術文献】

【特許文献】

【0008】

40

【特許文献 1】「Managing shared resources in a multi-computer system with failover support」と題する米国特許出願公開第 2012/0005348 号明細書

【発明の概要】

【発明が解決しようとする課題】

【0009】

一般に、図 1 で、各キャニスタには 1 台の CMC - たとえばラック 10 のようなラック - しかない。しかしながら、データセンタのリモート管理者は、CMC 12 が故障するかオフラインになると、キャニスタ又はラックパワーシェルフ 14 の状態監視を行う能力を失う。これにより、ラック 10 から提供される全ての電力及び支援からの問題が報告されなくなるので、ラック 10 の動作が妨げられる。

50

【 0 0 1 0 】

したがって、CMC 12の故障によって、ラック10内の全てのネットワーク装置の動作が妨げられる。正常状態では、キャニスタ又はラック10内のCMC 12は、パワーシェルフ14の状態を監視し、その状態を管理ノードに報告する。CMC 12の故障又は誤動作は、ネットワーク装置の監視を妨げ、したがって、ユニットが検査されうるまで動作を停止しなければならない。幾つかの現行のラック設計は、第1のCMCが故障した場合に、キャニスタ又はラック状態をリモート管理者に報告するバックアップCMCを備える。しかしながら、両方のCMCがリモート管理者から通信を遮断される可能性があるため、これは、完全な解決策ではない。更に、キャニスタ上の両方のCMCが故障すると、動作が停止されなければならない。

10

【 0 0 1 1 】

したがって、CMCがオフラインになった場合でも、支援シェルフ状態のリモート監視を可能にするラックシステムが必要である。更に、CMCがラック上の装置の監視データを報告できない場合でも、ラックの動作を可能にする機構が必要である。更に、ラック内の複数の装置の状態監視データを収集する複数のCMCを有するラックシステムが必要である。

【 0 0 1 2 】

特許文献1は、複合コンピュータシステムに挿入されたコンピュータから、挿入されたコンピュータの優先順位を表わす優先順位検出信号を読み取るステップと、コンピュータから、複合コンピュータシステムに現在挿入されている全てのコンピュータの挿入状態を表わす平面検出信号を読み取るステップと、優先順位検出信号と平面検出信号に従って、コンピュータが、複合コンピュータシステムに挿入された全てのコンピュータの中で最高優先順位を有するかどうかを決定するステップと、コンピュータが最高優先順位を有するという決定に応じて、共用リソースを監視し、最高優先順位コンピュータと関連付けられた特定出力信号を出力するステップであって、特定出力信号が、複合コンピュータシステムに現在挿入されている他のコンピュータに対して最高優先順位コンピュータの識別を提供し、最高優先順位コンピュータによって共用リソースの制御を提示するステップを開示した。

20

【 課題を解決するための手段 】

【 0 0 1 3 】

1つの開示される例は、管理コントローラと、第1及び第2のシャーシとを含む機器ラックである。前記第1のシャーシは、第1の電源ユニットと、該第1の電源ユニットに結合された第1のシステムバスと、該第1のシステムバスに結合された第1のシャーシ管理コントローラと、第1のネットワーク装置とを含む。前記第1の電源ユニットは、前記第1のシステムバスに状態データを送る。前記第2のシャーシは、第2の電源ユニットと、該第2の電源ユニットに結合された第2のシステムバスと、該第2のシステムバスに結合された第2のシャーシ管理コントローラと、第2のネットワーク装置とを含む。前記第2の電源ユニットは、前記第2のシステムバスに状態データを送る。前記第1のシャーシ管理コントローラは、前記第1の電源ユニットから状態データを取得するために前記第1のシステムバスと通信する。前記第1のシャーシ管理コントローラは、前記第2の電源ユニットから状態データを取得するために前記第2のシステムバスと通信する。前記第1のシャーシ管理コントローラは、状態データを中継するために管理スイッチと通信する。第2のシャーシ管理コントローラは、前記第1の電源ユニットから状態データを取得するために前記第1のシステムバスと通信する。前記第2のシャーシ管理コントローラは、前記第2の電源ユニットから状態データを取得するために前記第2のシステムバスと通信する。管理コントローラは、状態データを中継するために前記管理コントローラと前記第2のシャーシ管理コントローラとの間の通信を確立する働きをする。

30

40

【 0 0 1 4 】

別の開示された例は、機器ラックからの状態報告を保証する方法である。前記機器ラックは、管理コントローラを有する。前記機器ラックは、第1の電源ユニット、第1のシス

50

テムバス、第1のシャーシ管理コントローラ及び第1のネットワーク装置を含む第1のシャーシを有する。前記機器ラックは、第2の電源ユニット、第2のシステムバス、第2のシャーシ管理コントローラ及び第2のネットワーク装置を含む第2のシャーシを有する。状態データが、前記第1の電源ユニットから前記第1のシステムバスに提供される。前記状態データが、前記第2の電源ユニットから前記第2のシステムバスに提供される。前記第1のシャーシ管理コントローラと前記第1のシステムバスの間に通信が確立されて、前記第1の電源ユニットから状態データが取得される。前記第1のシャーシ管理コントローラと前記第2のシステムバスの間に通信が確立されて、前記第2の電源ユニットから状態データが取得される。状態データは、前記第1のシャーシ管理コントローラによって管理スイッチに中継される。前記第2のシャーシ管理コントローラと前記第2のシステムバスの間に通信が確立されて、前記第2の電源ユニットから状態データが取得される。前記第2のシャーシ管理コントローラと前記第1のシステムバスの間に通信が確立されて、前記第1の電源ユニットから状態データが取得される。前記第1のシャーシ管理コントローラからの状態データが中継されないという決定に基づいて、前記第2のシャーシ管理コントローラによる状態データの中継が、前記管理コントローラに対して行われる。

10

20

30

40

50

【0015】

上記の要約は、本開示の各実施形態又は全ての態様を表わすものではない。より正確に言うと、前述の要約は、本明細書に記載された新規の態様及び特徴の幾つかの例を提供するに過ぎない。本開示の上記の特徴及び利点並びに他の特徴及び利点は、添付図面及び添付の特許請求の範囲に関連して行われるとき本発明を実現する代表的な実施形態及び形態の以下の詳細な説明から容易に明らかになるであろう。

【0016】

開示は、添付図面に関連した典型的実施形態の以下の記述からよりよく理解されるであろう。

【図面の簡単な説明】

【0017】

【図1】ラック動作データを監視する単一CMCを有する先行技術のラックシステムの例を示す図である。

【図2】マスタCMCが故障した場合に継続的な報告を可能にするパートナCMCを含む例示的ラック内の構成要素のブロック図である。

【図3】CMCがそれぞれのパワーシェルフを監視する能力を示す図2の例示的ラックの構成要素のブロック図である。

【図4】通常動作マスタCMCからの状態データの流れを示す図2の例示的ラックの構成要素のブロック図である。

【図5】情報を維持しマスタCMC機能をパートナCMCに切り替えるために管理コントローラによって実行されるプロセスの流れ図である。

【図6】本開示の様々な例による典型的システムを示す図である。

【図7】本開示の様々な例による典型的システムを示す図である。

【発明を実施するための形態】

【0018】

本開示は、様々な修正及び代替形態が可能である。幾つかの代表的な実施形態は、図面に例として示され、本明細書で詳述される。しかしながら、本発明が、開示された特定の形態に限定されないことを理解されたい。より正確に言うと、開示は、添付の特許請求の範囲で定義されたような本発明の趣旨及び範囲内にある修正物、等価物、及び代替物を全て対象として含む。

【0019】

本発明は、様々な形態で実施されうる。代表的な実施形態が図面に示され、本明細書で詳述される。本開示は、本開示の原理の例又は説明であり、開示の広い態様を、示された実施形態に限定するものではない。その範囲で、例えば、要約 (Abstract)、課題を解決するための手段 (Summary) 及び発明を実施するための形態 (Detailed Description) の

節に開示されているが、特許請求の範囲に明示的に記載されていない要素及び制限は、含意、参照又は他の方法によって、単独又は集合的に、特許請求の範囲に組み込まれるべきでない。この詳細な説明のため、具体的に否定されない限り、単数は複数を含み逆の場合も同様であり、用語「含む (including)」は、「限定なしに含む」ことを意味する。更に、「約 (about)」、「ほぼ (almost)」、「実質的に (substantially)」、「近似的に (approximately)」などの近似の用語は、本明細書では、「における (at)」、「近い (near)」、「に近い (nearly at)」、「～の3～5%以内 (within 3-5% of)」、「許容可能な製造公差内」、又はこれらの任意の論理的組み合わせを意味するために使用されうる。

【0020】

図2は、単一CMC又は単一ボード上の複数のCMCが故障した場合でもラック100上の全ての装置の動作データの監視を可能にするために、異なるシャーシユニット内の異なるシェルフ又はボード上に複数のCMCを有する例示的な複式シャーシキャニスタラック100を示す。動作データ状態を取得する別個シャーシ上に少なくとも1つのパートナーCMCを追加することにより、マスタCMCが状態情報を提供できない場合でも、連続監視が可能になり、したがってラック100の動作が可能になる。

【0021】

ラック100は、管理スイッチ110と、2つのデータスイッチ112及び114を有する。ラック100は、第1のシャーシユニット120と第2のシャーシユニット130を含む。管理スイッチ110は、ラック100上の全ての装置の動作を監視する管理コントローラとして働くコントローラを含みうる。第1のシャーシユニット120は、サーバ122、124及び126などの様々なネットワーク装置を含む。シャーシユニット120は、また、シャーシユニット120のサーバ122、124及び126用の電源ユニット及びその他の支援電子回路を収容するパワーシェルフ128を含む。他のシャーシユニット130は、サーバ132、134及び136などの様々なネットワーク装置を含む。シャーシユニット130は、また、シャーシユニット130のネットワーク装置132、134及び136用の電源ユニット及び他の支援電子回路を収容するパワーシェルフ138を含む。例えば、シャーシユニット120及び130は、スロットを画定する取付け機構を保持する平行な側壁を備えうる。パワーシェルフは、シャーシユニット内の任意の位置に取り付けられうる。それぞれのネットワーク装置は、スロットに挿入され、それぞれのパワーシェルフ上の電源ユニットに接続されうる。

【0022】

図2は、また、パワーシェルフ128及び138のブロック図を示す。パワーシェルフ128は、一連の電源ユニット202、204及び206を有する。電源ユニット202、204及び206はそれぞれ、対応するファン212、214及び216を有する。パワーシェルフ128は、パワーシェルフ128の動作を監視するCMC220を有する。この例では、パワーシェルフ128は、サーバ122、124及び126などの関連ネットワーク装置に電力を提供する。システム管理(SM)バススイッチ222は、システムバス224を介して電源ユニット202、204及び206並びにファン212、214及び216と通信する。SMバススイッチ222は、動作データを受け取り、電源ユニット202、204及び206並びにファン212、214及び216にオン/オフコマンドを伝える働きをする。CMC220は、ネットワークインタフェース226に結合され、SMバススイッチ222に結合される。

【0023】

同様に、パワーシェルフ138は、一連の電源ユニット252、254及び256を有する。電源ユニット252、254及び256はそれぞれ、対応するファン262、264及び266を有する。パワーシェルフ138は、パワーシェルフ138の動作を監視するCMC270を有する。この例では、パワーシェルフ138は、サーバ132、134及び136などの関連ネットワーク装置に電力を提供する。システム管理(SM)バススイッチ272は、システムバス274を介して、電源ユニット252、254及び256

10

20

30

40

50

並びにファン 262、264 及び 266 と通信する。SMバススイッチ 272 は、動作データを受け取り、電源ユニット 252、254 及び 256 並びにファン 262、264 及び 266 にオン/オフコマンドを通信する働きをする。CMC 270 は、ネットワークインタフェース 276 に結合され、また SMバススイッチ 272 に結合される。

【0024】

図 2 に示されたように、CMC 220 及び 270 は、それぞれのネットワークインタフェース 226 及び 276 を介してネットワークによって管理スイッチ 110 に接続される。ネットワークインタフェース 226 は、SMバススイッチ 272 にも接続される。ネットワークインタフェース 276 は、SMバススイッチ 222 に接続される。後述されるように、この機構は、CMC 220 又は 270 がパワーシェルフ 128 又は 138 を監視し制御することを可能にする。

10

【0025】

この例では、CMC 220 は、ラック 100 のマスタ CMC であり、CMC 270 は、パートナ CMC である。当然ながら、CMC 270 は、マスタ CMC として働くことができ、CMC 220 は、ラック 100 のパートナ CMC として働く。CMC を有する追加のシャーシユニットが、ラック 100 内に取り付けられた例では、そのような他の CMC が、マスタ CMC として働きうる。この例では、CMC 220 は、SMバススイッチ 222 を介してパワーシェルフ 128 のパワーシェルフ状態を定期的にポーリングする。このようにして、CMC 220 は、電源ユニット 202、204、206 並びにファン 212、214 及び 216 の動作状態データを収集する。CMC 220 は、また、パワーシェルフ 138 の状態に関してパートナ CMC 270 をポーリングする。CMC 270 は、SMバススイッチ 272 を介してそのような状態データを、電源ユニット 252、254、256 並びにファン 262、264 及び 266 の動作状態データに関して収集する。

20

【0026】

リモート管理者は、CMC 220 へのネットワーク接続を介して管理スイッチ 110 からパワーシェルフ状態を取得するコマンドを送信できる。次に、CMC 220 は、管理スイッチ 110 を介したリモート管理者からの要求に応じてパワーシェルフ状態を報告する。冗長 CMC 220 及び 270 は、I2C、CPIO、UART 又は類似のネットワークプロトコルなどの接続によって互いに通信できる。2つの CMC 220 と 270 の間のハンドシェイクが、適切なプロトコルによるハートビートネゴシエーションによって行われる。この例では、このネゴシエーションは、I2C、CPIO、UART 又は類似のネットワークなどのプロトコルによって行われうる。したがって、CMC 220 及び 270 は、それぞれのパワーシェルフ 128 及び 138 からの電力状態情報を交換できる。CMC 220 又は 270 の一方が、他方の CMC が通信によるサービスを停止していることを発見した場合、その CMC は、監視機能を引き継ぐ。

30

【0027】

CMC 220 が故障した場合、パートナ CMC 270 は、元のマスタ CMC が故障したか管理スイッチ 110 へのネットワーク接続が絶たれたことを検出できる。あるいは、管理スイッチ 110 のコントローラなどの管理コントローラは、元のマスタ CMC からの通信の失敗を検出できる。次に、パートナ CMC 270 は、状態データをリモート管理者に報告する役割を引き継ぎ、パートナ CMC が既にそのような機能を引き継いだという警告をリモート管理者に送る。したがって、リモート管理者に報告される情報は、元のマスタ CMC がまだ機能している場合と同じになる。

40

【0028】

図 3 は、CMC 220 がマスタ CMC として指定されたパワーシェルフ 128 及びパワーシェルフ 138 を示す。図 3 の類似要素には、図 2 の類似要素の番号が付けられている。通常動作では、CMC 220 及び 270 は両方とも、それぞれのパワーシェルフ 128 及び 138 上のそれぞれの電源ユニットから全てのセンサ情報を能動的に読み取る。CMC 220 などの CMC の 1 つが、マスタ CMC として指定される。この例では、CMC 270 は、パートナ CMC であり、SMバススイッチ 272 から収集した状態情報を、図 3

50

に線 300 によって示されたように、CMC 220 に直接送らせる。また、CMC 220 は、SM バススイッチ 222 から収集した状態情報を、線 302 によって示されたように、CMC 270 に送らせる。マスタ CMC として指定された CMC 220 は、両方のパワーシェルフ 128 及び 138 上の全ての構成要素の状態を、管理スイッチ 110 (図 2) へのネットワーク接続によって、リモート管理に報告する。接続は、管理スイッチ 110 への能動的なデータ伝送を示す線 310 によって示されうる。両方の CMC 220 及び 270 が、それぞれの SM バススイッチ 222 及び 272 を介して情報を取得できるので、CMC 220 及び 270 の両方が常に同期される。CMC 220 から管理スイッチに伝送される動作又は状態データは、インターインテグレートッドサーキットプロトコル (I2C)、ユニバーサル非同期レシーバ/トランスミッタ (UART)、又はネットワークを介したものでよい。

10

【0029】

図 4 は、パワーシェルフ 128 とパワーシェルフ 138 を示し、この場合、CMC 220 は、管理スイッチ 110 と通信していない。図 4 の類似要素は、図 2 の類似要素の番号で示される。CMC 220 からの通信がない理由は、CMC 220 の故障又は通信路の切断であることある。CMC 220 及び 270 は両方とも、それぞれのパワーシェルフ 128 及び 138 上のそれぞれの電源ユニットから全てのセンサ情報を能動的に読み取る。この例では、CMC 270 は、パートナ CMC であり、SM バススイッチ 272 から収集された状態情報を、線 400 によって示されたように、CMC 220 に直接送らせる。CMC 220 は、また、SM バススイッチ 222 から収集された状態情報を、線 402 によって示されたように、CMC 270 に送らせる。この場合、システムは、管理スイッチ 110 (図 2) との通信を開始するパートナ CMC (CMC 270 など) に切り替わる。この接続は、管理スイッチ 110 へのデータの能動的伝送を示す線 410 によって示される。CMC 270 から管理スイッチ 110 に送られる動作又は状態データは、インターインテグレートッドサーキットプロトコル (I2C)、ユニバーサル非同期レシーバ/トランスミッタ (UART) 又はネットワークを介したものでよい。

20

【0030】

SM バススイッチ 222 及び 272 には、取り付けられた SM バスがあるものから別のものに切り替えることができる限り他の装置を使用できることを理解されたい。例えば、様々な SM バスからの入力に同時に通信できる SM バスアービタが使用されうる。SM バスアービタは、どのバスが最初にターゲットにトークできるかを定めるアービレータの役割をすることができる。あるいは、ターゲットに通信するバスを選択するためにマルチプレクサも使用できる。

30

【0031】

図 5 の流れ図は、図 2 のラック 100 が図 2 の CMC の一方からの状態情報の流れを保証する例示的な機械可読命令を表す。この例では、機械可読命令は、(a) プロセッサ、(b) コントローラ、及び/又は (c) 1 つ以上の他の適切な処理装置によって実行されるアルゴリズムを含む。アルゴリズムは、例えば、フラッシュメモリ、CD-ROM、フロッピディスク、ハードディスク、デジタルビデオ (多用途) ディスク (DVD)、又は他のメモリ装置などの有形媒体上に記憶されたソフトウェアで実施されうる。しかしながら、当業者は、アルゴリズム全体及び/又はその一部分が、代替として、プロセッサ以外の装置によって実行されかつ/又はファームウェア又は専用ハードウェアで周知の方法で実施されうることを容易に理解するであろう (例えば、特定用途向け集積回路 [ASIC]、プログラム可能論理回路 [PLD]、フィールドプログラム可能論理回路 [FPLD]、フィールドプログラマブルゲートアレイ [FPGA]、個別ロジックなどによって実現されうる)。例えば、インタフェースの構成要素のどれか又は全てが、ソフトウェア、ハードウェア、及び/又はファームウェアによって実現されうる。また、図 5 のフローチャートによって表わされた機械可読命令の幾つか又は全てが、手動で実行されうる。更に、図 5 に示されたフローチャートに関連して例示的なアルゴリズムを記述するが、当業者は、代替として、例示的な機械可読命令を実行する他の多くの方法が使用されうることを

40

50

容易に認識するであろう。例えば、ブロックの実行順序が変更されてもよく、及び/又は記述されたブロックの幾つかが変更、削除又は組み合わせられてもよい。

【0032】

図5は、シャreshyunitt 120及び130(図2)の監視を保証するために例示的ラック100を操作するプロセスの例示的な流れ図を示す。この例では、図2のCMCのそれぞれ(CMC220及びCMC270など)が、それぞれのパワーシェルフ状態をポーリングする(500)。CMCはそれぞれ、図2のバススイッチ222などのバススイッチから動作状態データを収集する(502)。CMCは、また、パートナーCMCによって監視されたそれぞれのパワーシェルフの状態に関して全てのパートナーCMCをポーリングする(504)。管理スイッチ110(図2)の管理コントローラは、マスタCMC220が動作しているかどうかを決定する(506)。マスタCMC220が動作している場合、管理スイッチ110は、パワーシェルフ状態を取得するために要求をマスタCMC220に送る(508)。要求を受け取ると、CMC220は、各シェルフの収集された動作データを送る(510)。

10

【0033】

現在のマスタCMCが動作できない場合、管理スイッチ110は、マスタCMCの役割を引き受けるようにCMC270の状態を切り替える(512)。次に、管理スイッチ110は、パワーシェルフ状態を取得する要求を新しいマスタCMC270に送る(514)。CMC270は、要求を受け取ると、各シェルフの収集された動作データを管理スイッチ110に送る(516)。

20

【0034】

図6は、例示的な計算処理システム600であり、計算処理システムの構成要素は、バス602を使用して互いに電気通信する。システム600は、処理ユニット(CPU又はプロセッサ)630と、システムメモリ604(例えば、読み取り専用メモリ(ROM)606及びランダムアクセスメモリ(RAM)608)を含む様々なシステム構成要素をプロセッサ630に結合するシステムバス602とを含む。システム600は、プロセッサ630と直接接続されるか、プロセッサ630と接近するか、プロセッサ630の一部として一体化された高速メモリのキャッシュを含みうる。システム600は、メモリ604及び/又は記憶装置612からのデータを、プロセッサ630による迅速アクセスされるキャッシュ628に複写できる。このようにして、キャッシュは、データを待っている間にプロセッサ630の性能向上を提供できる。以上その他のモジュールは、様々なアクションを実行するために、プロセッサ630を制御するかそのプロセッサ630を制御するように構成されうる。他のシステムメモリ604も使用可能でよい。メモリ604は、異なる性能特性を有する複数の様々なタイプのメモリを含みうる。プロセッサ630は、任意の汎用プロセッサと、記憶装置612に埋め込まれたモジュール1614、モジュール2616、モジュール3618などのハードウェアモジュール又はソフトウェアモジュールも含みうる。ハードウェアモジュール又はソフトウェアモジュールは、プロセッサ630、並びにソフトウェア命令が実際のプロセッサ設計に組み込まれた専用プロセッサを制御するように構成される。プロセッサ630は、本質的に、複数のコア又はプロセッサ、バス、メモリコントローラ、キャッシュなどを含む完全自立式の計算処理システムでよい。マルチコアプロセッサは、対称的でも非対称的でもよい。

30

40

【0035】

計算処理装置600とのユーザ対話を可能にするため、入力装置620が、入力機構として提供される。入力装置620は、会話用のマイクロフォン、ジェスチャ又は図形入力のタッチセンシティブスクリーン、キーボード、マウス、モーション入力などを含みうる。幾つかの例では、多モードシステムは、ユーザが、システム600と通信するために複数タイプの入力を提供することを可能にできる。この例では、出力装置622も提供される。通信インタフェース624は、ユーザ入力とシステム出力を支配し管理できる。

【0036】

記憶装置612は、コンピュータによってアクセス可能なデータを記憶する不揮発性メ

50

メモリでよい。記憶装置 6 1 2 は、磁気カセット、フラッシュメモリカード、ソリッドステートメモリ装置、デジタル多機能ディスク、カートリッジ、ランダムアクセスメモリ (RAM) 6 0 8、読み取り専用メモリ (ROM) 6 0 6、及びこれらの複合物でよい。

【0037】

コントローラ 6 1 0 は、BMC (ベースボード管理コントローラ) などのシステム 6 0 0 上の専用マイクロコントローラ又はプロセッサでよい。場合によって、コントローラ 6 1 0 は、インテリジェントプラットフォーム管理インタフェース (IPMI) の一部でよい。更に、場合によって、コントローラ 6 1 0 は、システム 6 0 0 のマザーボード又は主回路基板に埋め込まれる。コントローラ 6 1 0 は、システム管理ソフトウェアとプラットフォームハードウェアの間のインタフェースを管理できる。また、コントローラ 6 1 0 は、更に後述されるように、コントローラや周辺構成要素などの様々なシステム機器及び構成要素 (内部及び / 又は外部) と通信できる。

10

【0038】

コントローラ 6 1 0 は、通知、警告、及び / 又はイベントに対する固有の応答を生成し、リモート装置又は構成要素 (例えば、電子メールメッセージ、ネットワークメッセージなど) と通信して、自動ハードウェア回復手順などのための命令又はコマンドを生成できる。管理者は、また、更に後述されるように、コントローラ 6 1 0 とリモートで通信して、特定のハードウェア回復手順又は動作を開始又は行うことができる。

【0039】

コントローラ 6 1 0 は、また、コントローラ 6 1 0 が受け取ったイベント、警告、及び通知を管理及び維持するためにシステムイベントログコントローラ及び / 又は記憶装置を含みうる。例えば、コントローラ 6 1 0 又はシステムイベントログコントローラは、1 つ以上の装置及び構成要素から警告又は通知を受け取り、システムイベントログ記憶構成要素内に警告又は通知を維持できる。

20

【0040】

フラッシュメモリ 6 3 2 は、記憶及び / 又はデータ転送のためにシステム 6 0 0 によって使用される電子式不揮発性コンピュータ記憶媒体又はチップでよい。フラッシュメモリ 6 3 2 は、電氣的消去及び / 又は再プログラムされる。例えば、フラッシュメモリ 6 3 2 には、EPROM (消去可能 PROM)、EEPROM (電氣的消去可能プログラム可能な読み取り専用メモリ)、ROM、NVRAM、又は CMOS (相補型金属酸化膜半導体) が含まれる。フラッシュメモリ 6 3 2 は、ファームウェア 6 3 4 に専用化された 1 組の構成に加えて、システム 6 0 0 が最初に起動されたときにシステム 6 0 0 によって実行されるファームウェア 6 3 4 を記憶できる。フラッシュメモリ 6 3 2 は、また、ファームウェア 6 3 4 によって使用される構成を記憶できる。

30

【0041】

ファームウェア 6 3 4 は、EFI (拡張可能ファームウェアインタフェース) や UEFI (統合拡張可能ファームウェアインタフェース) などの基本入出力システム又は等価物を含みうる。ファームウェア 6 3 4 は、システム 6 0 0 が起動されるたびにシーケンスプログラムとしてロード及び実行される。ファームウェア 6 3 4 は、1 組の構成に基づいてシステム 6 0 0 内にあるハードウェアを認識、初期化、試験できる。ファームウェア 6 3 4 は、システム 6 0 0 上で POST (電源投入時自己試験) などの自己試験を実行できる。この自己試験は、ハードディスクドライブ、光学読み取り装置、冷却装置、メモリモジュール、拡張カードなどの様々なハードウェア構成要素の機能を試験できる。ファームウェア 6 3 4 は、オペレーティングシステム (OS) を記憶するメモリ 6 0 4、ROM 6 0 6、RAM 6 0 8 及び / 又は記憶装置 6 1 2 内の領域をアドレス指定し割り当てることができる。ファームウェア 6 3 4 は、ブートルード及び / 又は OS をロードし、システム 6 0 0 の制御を OS に提供できる。

40

【0042】

システム 6 0 0 のファームウェア 6 3 4 は、ファームウェア 6 3 4 がシステム 6 0 0 内の様々なハードウェア構成要素をどのように制御するかを定義するファームウェア構成を

50

含みうる。ファームウェア構成は、システム 600 内の様々なハードウェア構成要素が始動される順序を決定できる。ファームウェア 634 は、ファームウェアデフォルト構成がパラメータと異なりうる様々な異なるパラメータを設定できるようにする UEFI などのインタフェースを提供できる。例えば、ユーザ（例えば、管理者）は、ファームウェア 634 を使用してクロックとバス速度を指定し、どの周辺装置がシステム 600 に取り付けられるかを定義し、健全状態（例えば、ファン回転速度及び CPU 温度制限）の監視を設定し、及び / 又はシステム 600 の全体性能及び消費電力に影響を及ぼす様々な他のパラメータを提供できる。ファームウェア 634 は、フラッシュメモリ 632 に記憶されるように示されているが、当業者は、ファームウェア 634 がメモリ 604 や ROM 606 などの他のメモリ構成要素に記憶されうることを容易に理解するであろう。

10

【0043】

システム 600 は、1つ以上のセンサ 626 を含みうる。1つ以上のセンサ 626 には、例えば、1つ以上の温度センサ、サーマルセンサ、酸素センサ、化学センサ、ノイズセンサ、ヒートセンサ、電流センサ、検電器、空気流量センサ、流量センサ、赤外線放射温度計、熱流束センサ、温度計、高温計などが含まれうる。例えば、1つ以上のセンサ 626 は、例えば、バス 602 を介して、プロセッサ、キャッシュ 628、フラッシュメモリ 632、通信インタフェース 624、メモリ 604、ROM 606、RAM 608、コントローラ 610 及び記憶装置 612 と通信できる。1つ以上のセンサ 626 は、システム内の他の構成要素と、インターインテグレートッドサーキット（I2C）、汎用出力（GPIO）などの1つ以上の異なる手段によって通信できる。また、システム 600 上の様々なタイプのセンサ（例えば、センサ 626）は、冷却ファン速度、電力状態、オペレーティングシステム（OS）状態、ハードウェア状態などのパラメータについてコントローラ 610 に報告できる。表示装置 636 は、コントローラ 610 によって実行されるアプリケーションに関連付けられたグラフィックを提供するために、システム 600 によって使用されうる。

20

【0044】

図 7 は、記述された方法又は動作を実行し、グラフィックカルユーザインタフェース（GUI）を生成し表示する際に使用されうる、チップセットアーキテクチャを有する例示的なコンピュータシステム 700 を示す。コンピュータシステム 700 は、開示された技術を実施するために使用されうるコンピュータハードウェア、ソフトウェア及びファームウェアを含みうる。システム 700 は、識別された計算を実行するように構成されたソフトウェア、ファームウェア及びハードウェアを実行できる様々な物理的及び / 又は論理的に別個のリソースを表すプロセッサ 710 を含みうる。プロセッサ 710 は、プロセッサ 710 への入出力を制御できるチップセット 702 と通信できる。この例では、チップセット 702 は、表示装置などの出力装置 714 に情報を出し、また記憶装置 716 に情報を読み書きできる。例えば、記憶装置 716 には、磁気媒体とソリッドステート媒体が含まれうる。チップセット 702 は、また、RAM 718 からデータを読み出しデータを書き込みできる。チップセット 702 とインタフェースするために、様々なユーザインタフェース構成要素 706 とインタフェースするブリッジ 704 が提供されうる。ユーザインタフェース構成要素 706 には、キーボード、マイクロフォン、タッチ検出及び処理回路、及びマウスなどのポインティング装置が含まれうる。

30

40

【0045】

チップセット 702 は、また、様々な物理インタフェースを有しうる1つ以上の通信インタフェース 708 とインタフェースできる。そのような通信インタフェースには、有線及び無線ローカルエリアネットワーク、広帯域無線ネットワーク、及びパーソナルエリアネットワークのためのインタフェースが含まれうる。更に、機械は、ユーザインタフェース構成要素 706 を介してユーザから入力を受け取り、プロセッサ 710 を使用してそのような入力を解釈することによってブラウズ機能などの適切な機能を実行できる。

【0046】

更に、チップセット 702 は、また、起動時にコンピュータシステム 700 によって実

50

行されうるファームウェア 712 と通信できる。ファームウェア 712 は、1組のファームウェア構成に基づいてコンピュータシステム 700 内にあるハードウェアを理解、初期化、試験できる。ファームウェア 712 は、システム 700 上で POST などの自己試験を実行できる。自己試験は、様々なハードウェア構成要素 702 ~ 718 の機能を試験できる。ファームウェア 712 は、OS を記憶するメモリ 718 内の領域をアドレス指定し割り当てることができる。ファームウェア 712 は、ブートローダ及び / 又は OS をロードし、システム 700 の制御を OS に提供できる。場合によって、ファームウェア 712 は、ハードウェア構成要素 702 ~ 710 及び 714 ~ 718 と通信できる。ここで、ファームウェア 712 は、チップセット 702 及び / 又は 1 つ以上の他の構成要素によって、ハードウェア構成要素 702 ~ 710 及び 714 ~ 718 と通信できる。場合によって、ファームウェア 712 は、ハードウェア構成要素 702 ~ 710 及び 714 ~ 718 と直接通信できる。

10

【0047】

例示的なシステム 600 (図 6) 及び 700 は、複数のプロセッサ (例えば、630、710) を有してもよく、より高い処理性能を提供するためにネットワーク接続された一群又はクラスタの計算処理装置の一部でもよい。

【0048】

本明細書で使用される用語は、特定の実施形態について述べるためだけのものであり、本発明を限定するものではない。本明細書で使用される時、単数形「a」、「an」及び「the」は、文脈で特に断らない限り、複数形も含む。更に、用語「含む (including)」、「含む (includes)」、「有する (having)」、「有する (has)」、「備える (with)」、又はこれらの変形は、詳細な説明及び / 又は特許請求の範囲で使用される範囲で、そのような用語は、「含む (comprising)」と同じように包括的である。

20

【0049】

特に定義されない限り、本明細書で使用される全ての用語 (技術及び科学用語を含む) は、当業者によって一般に理解される意味と同じ意味を有する。更に、一般に使用されている辞書で定義されたような用語は、関連技術の文脈での意味と一致する意味を有すると解釈されるべきであり、本明細書で明示的に定義されない限り、理想的又は過度に形式的な意味で解釈されない。

【0050】

以上、本発明の様々な実施形態について述べたが、それらの実施形態が、単なる例であり限定ではないことを理解されたい。開示された実施形態に対する多数の変更が、本発明の趣旨又は範囲から逸脱せずに、本明細書の開示に従って行われうる。したがって、本発明の広さ及び範囲は、前述の実施形態のいずれによっても限定されるべきでない。より正確に言うと、本発明の範囲は、以下の特許請求の範囲とその等価物によって定義されるべきである。

30

【0051】

本発明を 1 つ以上の実施態様に関して図示し述べてきたが、本明細書及び添付図面の読み取りと理解に基づいて、等価な変更及び修正が行われる又は当業者に知られる。更に、本発明の特定の特徴が、幾つかの実施態様の 1 つのみに関して開示されたが、そのような特徴が、任意の所定又は特定用途に望ましく有利になるように、他の実施態様の 1 つ以上の他の特徴と組み合わせられうる。

40

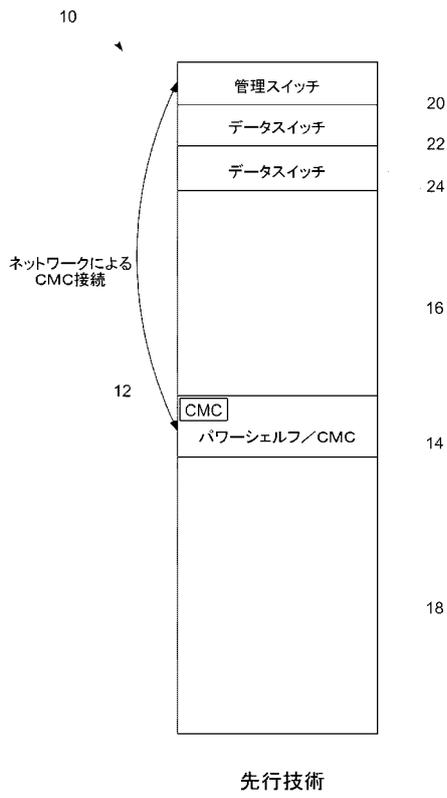
【符号の説明】

【0052】

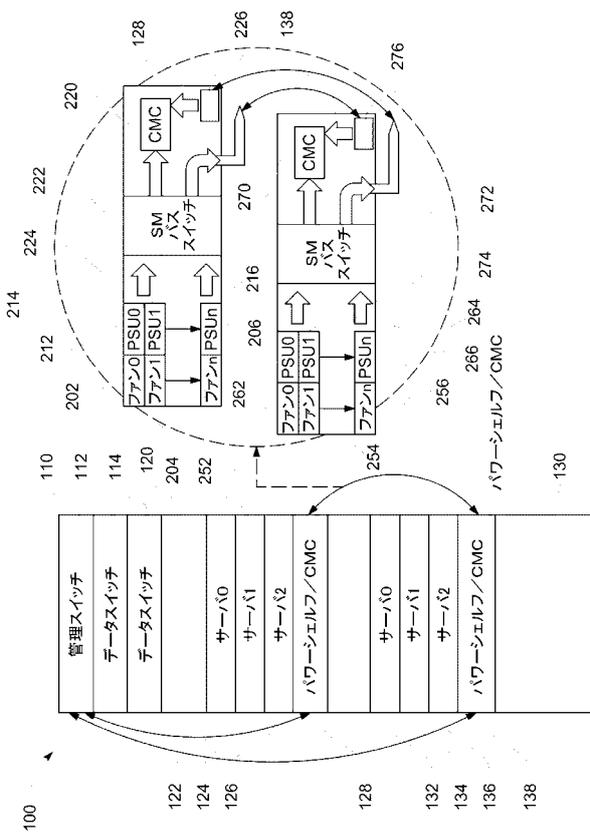
110 管理スイッチ
 112, 114 データスイッチ
 122, 124, 126, 132, 134, 136 サーバ
 128, 138 パワーシェルフ / CMC
 202, 204, 206, 252, 254, 256 ファン
 222, 272 SMバススイッチ

50

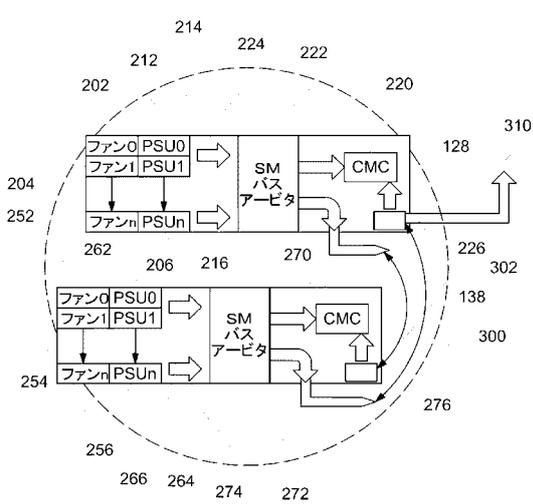
【 図 1 】



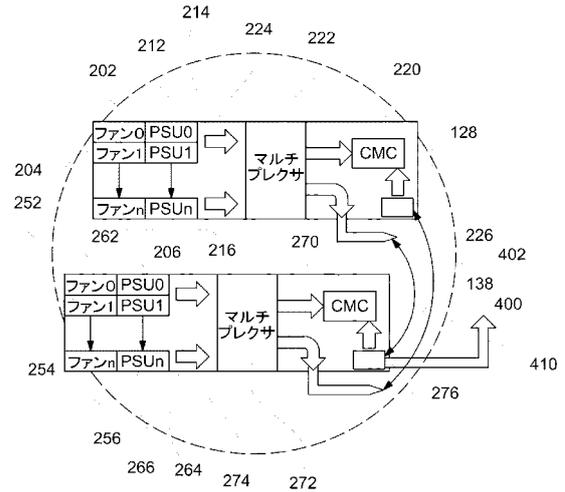
【 図 2 】



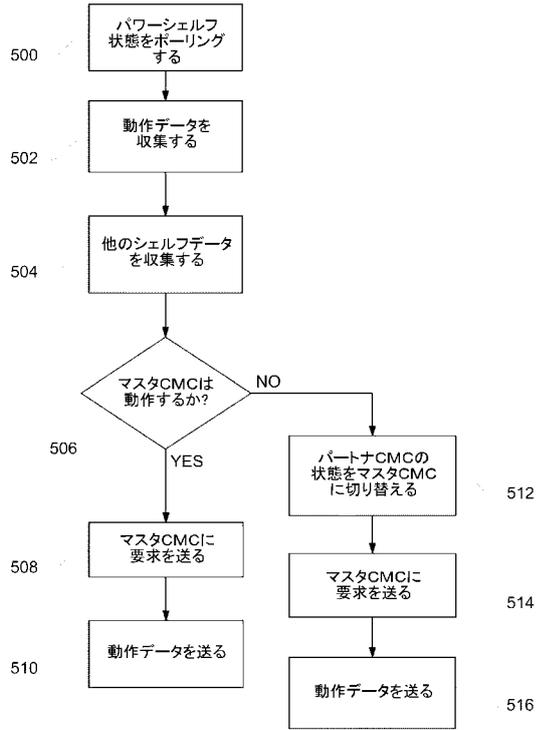
【 図 3 】



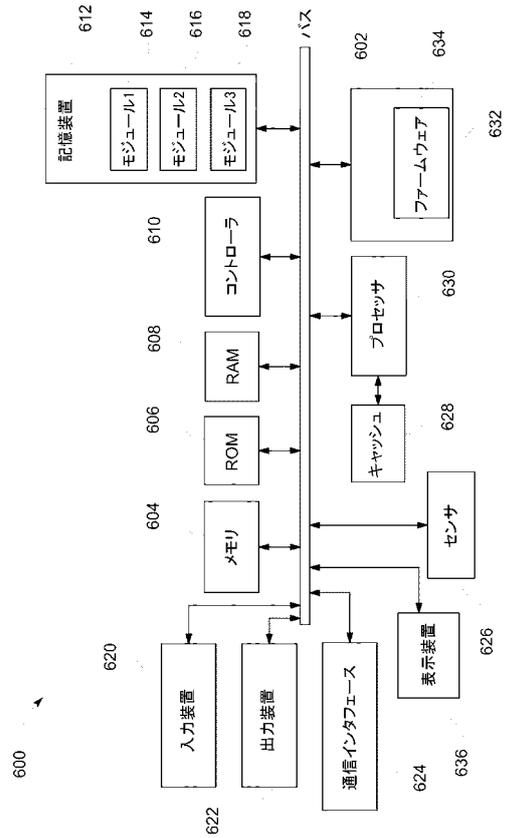
【 図 4 】



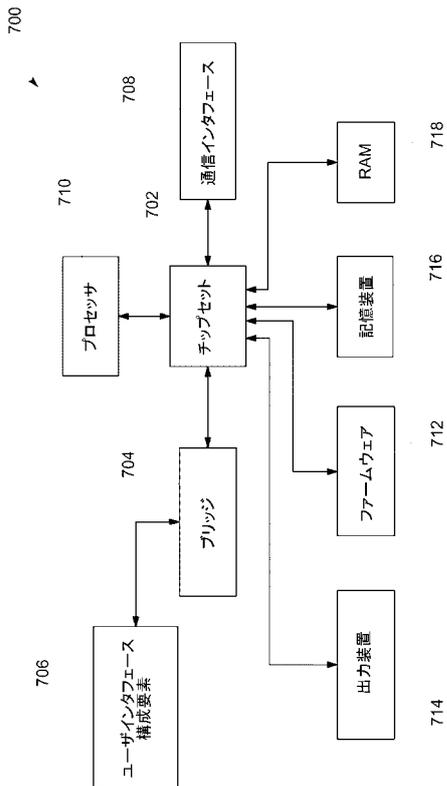
【 図 5 】



【 図 6 】



【 図 7 】



【外国語明細書】

2019121338000001.pdf