

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5266590号
(P5266590)

(45) 発行日 平成25年8月21日(2013.8.21)

(24) 登録日 平成25年5月17日(2013.5.17)

(51) Int.Cl. F I
G 0 6 F 1 3 / 1 4 (2 0 0 6 . 0 1) G 0 6 F 1 3 / 1 4 3 3 0 E

請求項の数 15 (全 31 頁)

(21) 出願番号	特願2009-217437 (P2009-217437)	(73) 特許権者	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(22) 出願日	平成21年9月18日(2009.9.18)	(74) 代理人	100114236 弁理士 藤井 正弘
(65) 公開番号	特開2011-65551 (P2011-65551A)	(74) 代理人	100075513 弁理士 後藤 政喜
(43) 公開日	平成23年3月31日(2011.3.31)	(74) 代理人	100120260 弁理士 飯田 雅昭
審査請求日	平成23年2月4日(2011.2.4)	(72) 発明者	片野 真吾 神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業 部内

最終頁に続く

(54) 【発明の名称】 計算機システムの管理方法、計算機システム及びプログラム

(57) 【特許請求の範囲】

【請求項1】

プロセッサとメモリとP C I インターフェースとを有する複数の計算機と、
前記複数の計算機を前記P C I インターフェース経由で接続する1以上のP C I スイッチと、

前記P C I スイッチに接続する複数のP C I デバイスと、
前記P C I スイッチに接続された前記P C I デバイスを管理するP C I デバイステーブルと、

グループ識別子に対応して前記P C I デバイスの詳細な情報を格納するP C I デバイステーブルと、
を有して、

前記P C I デバイスと前記計算機の割り当てを制御する管理サーバと、を備えて前記計算機とP C I デバイスとの割り当てを管理する計算機システムの管理方法において、

前記P C I スイッチが、前記P C I デバイスの追加または変更、あるいは当該P C I スイッチの追加を前記管理サーバへ通知するステップと、

前記管理サーバが、前記通知を受信して前記P C I スイッチに対して前記P C I デバイスの基本情報の取得を要求するステップと、

前記P C I スイッチが前記基本情報の取得要求を受信して、前記P C I デバイスの基本情報を取得し、前記管理サーバに前記基本情報を通知するステップと、

前記管理サーバが、前記基本情報を受信して前記P C I デバイスの割り当て状態を占有

10

20

または共有の何れか一方に決定し、前記決定した割り当て状態で前記 P C I デバイスを管理サーバに割り当てるように前記 P C I スイッチに要求するステップと、

前記 P C I スイッチが前記管理サーバからの割り当て要求を受信し、前記管理サーバが決定した割り当て状態で、前記 P C I デバイスと管理サーバを接続するステップと、

前記管理サーバが、前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得するステップと、

前記管理サーバが、前記取得した P C I デバイスの詳細な情報を前記 P C I デバイステーブルに格納するステップと、

前記管理サーバは、前記 P C I デバイスの詳細な情報を取得した後に、当該 P C I デバイスと管理サーバの接続を解除する要求を前記 P C I スイッチに指令するステップと、

前記 P C I スイッチは、前記管理サーバから受信した接続解除の指令に基づいて前記 P C I デバイスと管理サーバの接続を解除するステップと、

前記 P C I デバイスの情報が前記 P C I デバイステーブルに追加された場合に、前記管理サーバが、前記 P C I デバイステーブルの各項目とともにグループ識別子を前記 P C I デバイスグループテーブルに格納するグループングステップと、

前記管理サーバが、前記 P C I デバイスグループテーブルの各項目を検索条件として表示装置に表示し、P C I デバイスを割り当てる計算機と、前記計算機に割り当てる P C I デバイスの検索条件とを受け付けるステップと、

前記検索条件に適合する P C I デバイスを、前記 P C I デバイステーブルから抽出し、当該 P C I デバイスを前記計算機に割り当てるステップと、

を含むことを特徴とする計算機システムの管理方法。

【請求項 2】

請求項 1 に記載の計算機システムの管理方法であって、

前記 P C I スイッチが、前記管理サーバから受信した接続解除の指令に基づいて前記 P C I デバイスと管理サーバの接続を解除するステップは、

前記 P C I デバイスの占有または共有の状態を、前記管理サーバに接続する以前の状態に復帰させることを特徴とする計算機システムの管理方法。

【請求項 3】

請求項 1 に記載の計算機システムの管理方法であって、

前記管理サーバは、仮想計算機を有し、

前記管理サーバが、前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得するステップは、

前記仮想計算機がドライバを読み込んで、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得することを特徴とする計算機システムの管理方法。

【請求項 4】

請求項 3 に記載の計算機システムの管理方法であって、

前記管理サーバは、複数の仮想計算機を有し、前記仮想計算機毎に異なる種類の O S を実行し、

前記管理サーバが、前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得するステップは、

前記仮想計算機毎にそれぞれドライバを読み込んで、当該ドライバを介して前記 P C I デバイスの詳細な情報をそれぞれ取得することを特徴とする計算機システムの管理方法。

【請求項 5】

請求項 1 に記載の計算機システムの管理方法であって、

前記管理サーバは、前記 P C I スイッチと前記計算機の接続状態と、前記計算機に割り当てた P C I デバイスを管理する管理テーブルを有し、

前記管理サーバが、前記 P C I デバイスの構成を抽出する抽出元の計算機を受け付ける

10

20

30

40

50

ステップと、

前記管理サーバが、前記抽出した P C I デバイスの構成を適用する適用先の計算機を受け付けるステップと、

前記管理サーバが、管理テーブルと前記 P C I デバイステーブルから前記抽出元の計算機に割り当てた P C I デバイスと同一の構成の P C I デバイスを抽出し、前記適用先の計算機に割り当てよう前記 P C I スイッチに指令するステップと、

をさらに有することを特徴とする計算機システムの管理方法。

【請求項 6】

プロセッサとメモリと P C I インターフェースとを有する複数の計算機と、

前記複数の計算機を前記 P C I インターフェースで接続する 1 以上の P C I スイッチと

10

、
前記 P C I スイッチに接続する複数の P C I デバイスと、

前記 P C I スイッチに接続された前記 P C I デバイスを管理する P C I デバイステーブルと、

グループ識別子に対応して前記 P C I デバイスの詳細な情報を格納する P C I デバイスグループテーブルと、

を有して、

前記 P C I デバイスと前記計算機の割り当てを制御する管理サーバと、を備えた計算機システムにおいて、

前記 P C I スイッチは、

20

前記 P C I デバイスの追加または変更、あるいは当該 P C I スイッチの追加を前記管理サーバへ通知し、前記管理サーバからの要求に基づいて前記 P C I デバイスの基本情報を取得して前記管理サーバへ通知し、前記管理サーバからの割り当て要求に基づいて前記 P C I デバイスを前記管理サーバまたは前記計算機に接続する制御部を有し、

前記管理サーバは、

前記 P C I スイッチから前記 P C I デバイスの追加または変更、あるいは当該 P C I スイッチの追加を示す通知を受信したときに、前記 P C I デバイスの詳細な情報を取得する P C I デバイス情報取得部と、前記計算機に P C I デバイスを割り当てる P C I デバイス割り当て部と、を有し、

前記 P C I デバイス情報取得部は、

30

前記 P C I スイッチから前記通知を受信したときに前記 P C I デバイスの基本情報を前記 P C I スイッチから取得し、当該基本情報に基づいて前記 P C I デバイスの割り当て状態を占有または共有の何れか一方に決定し、前記決定した割り当て状態で前記 P C I デバイスを管理サーバに割り当てるように前記 P C I スイッチに要求し、前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得し、前記取得した P C I デバイスの詳細な情報を前記 P C I デバイステーブルに格納し、前記 P C I デバイスの詳細な情報を取得した後に、当該 P C I デバイスと管理サーバの接続を解除する要求を前記 P C I スイッチに指令し、

前記 P C I スイッチは、前記管理サーバから受信した接続解除の指令に基づいて前記 P C I デバイスと管理サーバの接続を解除し、

40

前記 P C I デバイス割り当て部は、

前記 P C I デバイスグループテーブルの各項目を検索条件として表示装置に表示し、P C I デバイスを割り当てる計算機と、前記計算機に割り当てる P C I デバイスの検索条件とを受け付けて、前記検索条件に適合する P C I デバイスを、前記 P C I デバイステーブルから抽出し、当該 P C I デバイスを前記計算機に割り当てることを特徴とする計算機システム。

【請求項 7】

請求項 6 に記載の計算機システムであって、

前記 P C I スイッチは、前記管理サーバから受信した接続解除の指令に基づいて前記 P C I デバイスと管理サーバの接続を解除する際に、前記 P C I デバイスの占有または共有

50

の状態を、前記管理サーバに接続する以前の状態に復帰させることを特徴とする計算機システム。

【請求項 8】

請求項 6 に記載の計算機システムであって、

前記管理サーバは、前記 P C I デバイス情報取得部を含む仮想計算機を有し、

前記仮想計算機がドライバを読み込んで、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得することを特徴とする計算機システム。

【請求項 9】

請求項 8 に記載の計算機システムであって、

前記管理サーバは、前記 P C I デバイス情報取得部を含む複数の仮想計算機を有し、前記仮想計算機毎に異なる種類の O S を実行し、前記仮想計算機毎にドライバを読み込んで、当該ドライバを介して前記 P C I デバイスの詳細な情報をそれぞれ取得することを特徴とする計算機システム。

10

【請求項 10】

請求項 9 に記載の計算機システムであって、

前記管理サーバは、前記 P C I スイッチと前記計算機の接続状態と、前記計算機に割り当てた P C I デバイスを管理する管理テーブルを有し、

前記 P C I デバイス割り当て部は、

前記 P C I デバイスの構成を抽出する抽出元の計算機を受け付け、前記抽出した P C I デバイスの構成を適用する適用先の計算機を受け付け、前記管理テーブルと前記 P C I デバイステーブルから前記抽出元の計算機に割り当てた P C I デバイスと同一の構成の P C I デバイスを抽出し、前記適用先の計算機に割り当てるよう前記 P C I スイッチに指令することを特徴とする計算機システム。

20

【請求項 11】

プロセッサとメモリと P C I インターフェースとを有する複数の計算機と、前記複数の計算機を前記 P C I インターフェース経由で接続する 1 以上の P C I スイッチと、前記 P C I スイッチに接続する複数の P C I デバイスと、前記 P C I スイッチに接続された前記 P C I デバイスを管理する P C I デバイステーブルと、グループ識別子に対応して前記 P C I デバイスの詳細な情報を格納する P C I デバイスグループテーブルと、を有して、前記 P C I デバイスと前記計算機の割り当てを制御する管理サーバで実行されるプログラムであって、

30

前記 P C I スイッチから前記 P C I デバイスの追加または変更、あるいは当該 P C I スイッチの追加を示す通知を受信したときに、前記 P C I スイッチに対して前記 P C I デバイステーブルの基本情報の取得を要求する手順と、

前記 P C I スイッチから前記 P C I デバイスの基本情報を取得し、当該基本情報から前記 P C I デバイスの割り当て状態を占有または共有の何れか一方に決定し、前記決定した割り当て状態で前記 P C I デバイスを管理サーバに割り当てるように前記 P C I スイッチに要求する手順と、

前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得する手順と、

40

前記取得した P C I デバイスの詳細な情報を前記 P C I デバイステーブルに格納する手順と、

前記 P C I デバイスの詳細な情報を取得した後に、当該 P C I デバイスと管理サーバの接続を解除する要求を前記 P C I スイッチに指令し、前記 P C I スイッチに管理サーバと前記 P C I デバイスの接続を解除させる手順と、

前記 P C I デバイスの情報が前記 P C I デバイステーブルに追加された場合に、前記 P C I デバイステーブルの各項目とともにグループ識別子を前記 P C I デバイスグループテーブルに格納する手順と、

前記 P C I デバイスグループテーブルの各項目を検索条件として表示装置に表示し、P C I デバイスを割り当てる計算機と、前記計算機に割り当てる P C I デバイスの検索条件

50

とを受け付ける手順と、

前記検索条件に適合する P C I デバイスを、前記 P C I デバイステーブルから抽出し、当該 P C I デバイスを前記計算機に割り当てる手順と、

を前記管理サーバに実行させることを特徴とするプログラム。

【請求項 1 2】

請求項 1 1 に記載のプログラムであって、

前記 P C I デバイスの詳細な情報を取得した後に、当該 P C I デバイスと管理サーバの接続を解除する要求を前記 P C I スイッチに指令し、前記 P C I スイッチに管理サーバと前記 P C I デバイスの接続を解除させる手順は、

前記 P C I デバイスの占有または共有の状態を、前記管理サーバに接続する以前の状態に復帰させることを特徴とするプログラム。

10

【請求項 1 3】

請求項 1 1 に記載のプログラムであって、

前記管理サーバは、仮想計算機を有し、

前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得する手順は、

前記仮想計算機でドライバを読み込んで、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得することを特徴とするプログラム。

【請求項 1 4】

請求項 1 3 に記載のプログラムであって、

前記管理サーバは、複数の仮想計算機を有し、前記仮想計算機毎に異なる種類の O S を実行し、

20

前記 P C I スイッチを介して接続した P C I デバイスのドライバを読み込み、当該ドライバを介して前記 P C I デバイスの詳細な情報を取得する手順は、

前記仮想計算機でそれぞれドライバを読み込んで、当該ドライバを介して前記 P C I デバイスの詳細な情報をそれぞれ取得することを特徴とするプログラム。

【請求項 1 5】

請求項 1 1 に記載のプログラムであって、

前記管理サーバは、前記 P C I スイッチと前記計算機の接続状態と、前記計算機に割り当てた P C I デバイスを管理する管理テーブルを有し、

30

前記 P C I デバイスの構成を抽出する抽出元の計算機を受け付ける手順と、

前記抽出した P C I デバイスの構成を適用する適用先の計算機を受け付ける手順と、

前記管理テーブルと前記 P C I デバイステーブルから前記抽出元の計算機に割り当てた P C I デバイスと同一の構成の P C I デバイスを抽出し、前記適用先の計算機に割り当てるよう前記 P C I スイッチに指令する手順と、

をさらに含むことを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、複数の計算機と複数の P C I デバイスを P C I スイッチで接続した計算機システムに関し、特に P C I スイッチの管理に関する。

40

【背景技術】

【0 0 0 2】

従来、P C I デバイスは計算機内のスロットに搭載されているため、P C I デバイスの搭載数は制限されており、複数の計算機の P C I デバイスを集中して管理する必要はなかった。しかし、近年では I / O スイッチとして P C I E x p r e s s S w i t c h (以降、P C I e S W) が実用化された事により、P C I e S W を採用した計算機システムでは、複数のサーバ(ブレード)を P C I e S W に接続し、この P C I e S W に多数の P C I デバイスを接続することで、各サーバが必要とする P C I デバイスを柔軟に割り当てることができる。そして、このような P C I e S W と複数のサーバ(ブレード)を搭載し

50

た計算機システムのPCI（またはPCI Express）デバイスの搭載数は従来に比して数10倍に増加する場合もある。なお、PCIデバイスは、PCIまたはPCI Expressの規格に準拠したI/Oデバイスである。

【0003】

計算機システムにおいては、各サーバが必要とするPCIデバイスを的確に割り当てるためには、PCIe SWに搭載された多数のPCIデバイスを管理し、PCIデバイスの情報や、サーバへの割り当て状態等を把握していなければならない。

【0004】

I/Oスイッチを備えた計算機システムで、PCIデバイスの情報を取得する手法としては、新たにI/OデバイスがPCIe SWに接続された場合やI/Oデバイスの障害が原因で交換された場合などに管理サーバがI/Oスイッチの管理インターフェースを介してI/Oデバイスの情報を取得する技術が知られている（例えば、特許文献1）。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2008-310489公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

上記従来技術において、管理サーバがPCIe SWの管理インターフェースから取得できるPCIデバイスの情報は、PCIバスプロトコルで規定された基本情報のみであり、管理者が管理を行うための情報としては不十分なため、管理者はより詳細な情報を手動で取得して管理しなければならない。その結果、設定作業時間が増大し、さらに設定ミスが発生するリスクが大幅に増加する、という問題があった。

【0007】

すなわち、PCIバスプロトコルで規定されたPCIデバイスの基本情報には、ベンダIDやデバイスID、クラスコードなどが含まれており、PCIデバイスの種別などを判別することができる。一方、計算機システムの管理者は、複数のサーバで各種のOSを稼働させ、複数のPCIデバイスのうちから各OSに割り当てる必要がある。この割り当ての際に、OSの種類やバージョン（カーネルバージョンやサービスパック・バージョン）の違いによっては、割り当てようとするPCIデバイスのドライバがOS標準のドライバに含まれていない場合がある。この場合、管理者は、対象としているサーバのOSに対応したPCIデバイスのドライバを探したり、他のPCIデバイスのうち当該OSでサポートしているPCIデバイスを探す必要があった。あるいは、複数の仮想サーバをVMM（Virtual Machine Manager）やハイパバイザといった仮想化技術で提供する場合、PCIデバイスのファームウェアのリビジョンが古いものでは仮想化技術を適用できない場合がある。

【0008】

管理者は、稼働させるOSの種類やバージョンの違いと、PCIデバイスの種類やバージョン、ファームウェアのリビジョンの違いに応じて、OSが確実に利用可能なPCIデバイスを割り当てる必要があるが、上記従来技術によるPCIデバイスの基本情報のみでは、対象としているOSの標準のドライバで動作するか否かを容易に判定することはできないという問題があった。また、仮想化技術を利用する場合では、PCIデバイスの基本情報のみでは、PCIデバイスのファームウェアのリビジョンなどが仮想化技術に対応しているか否かを管理者が容易に把握することができない、という問題があった。

【0009】

そこで本発明は、上記問題点に鑑みてなされたもので、PCIe SWを管理するにあたり、PCIデバイスの詳細な情報（種別、性能等、管理を行うために必要な機能レベルの情報）を自動的に取得して、サーバに割り当てるPCIデバイスを管理者が容易に探索可能な計算機システムを提供することを目的とする。

【課題を解決するための手段】

【0010】

本発明は、プロセッサとメモリとPCIインターフェースとを有する複数の計算機と、前記複数の計算機を前記PCIインターフェース経由で接続する1以上のPCIスイッチと、前記PCIスイッチに接続する複数のPCIデバイスと、前記PCIスイッチに接続された前記PCIデバイスを管理するPCIデバイステーブルと、グループ識別子に対応して前記PCIデバイスの詳細な情報を格納するPCIデバイスグループテーブルと、を有して、前記PCIデバイスと前記計算機の割り当てを制御する管理サーバと、を備えて前記計算機とPCIデバイスとの割り当てを管理する計算機システムの管理方法において、前記PCIスイッチが、前記PCIデバイスの追加または変更、あるいは当該PCIスイッチの追加を前記管理サーバへ通知するステップと、前記管理サーバが、前記通知を受信して前記PCIスイッチに対して前記PCIデバイスの基本情報の取得を要求するステップと、前記PCIスイッチが前記基本情報の取得要求を受信して、前記PCIデバイスの基本情報を取得し、前記管理サーバに前記基本情報を通知するステップと、前記管理サーバが、前記基本情報を受信して前記PCIデバイスの割り当て状態を占有または共有の何れか一方に決定し、前記決定した割り当て状態で前記PCIデバイスを管理サーバに割り当てるように前記PCIスイッチに要求するステップと、前記PCIスイッチが前記管理サーバからの割り当て要求を受信し、前記管理サーバが決定した割り当て状態で、前記PCIデバイスと管理サーバを接続するステップと、前記管理サーバが、前記PCIスイッチを介して接続したPCIデバイスのドライバを読み込み、当該ドライバを介して前記PCIデバイスの詳細な情報を取得するステップと、前記管理サーバが、前記取得したPCIデバイスの詳細な情報を前記PCIデバイステーブルに格納するステップと、前記管理サーバは、前記PCIデバイスの詳細な情報を取得した後に、当該PCIデバイスと管理サーバの接続を解除する要求を前記PCIスイッチに指令するステップと、前記PCIスイッチは、前記管理サーバから受信した接続解除の指令に基づいて前記PCIデバイスと管理サーバの接続を解除するステップと、前記PCIデバイスの情報が前記PCIデバイステーブルに追加された場合に、前記管理サーバが、前記PCIデバイステーブルの各項目とともにグループ識別子を前記PCIデバイスグループテーブルに格納するグルーピングステップと、前記管理サーバが、前記PCIデバイスグループテーブルの各項目を検索条件として表示装置に表示し、PCIデバイスを割り当ててる計算機と、前記計算機に割り当ててるPCIデバイスの検索条件とを受け付けるステップと、前記検索条件に適合するPCIデバイスを、前記PCIデバイステーブルから抽出し、当該PCIデバイスを前記計算機に割り当ててるステップと、を含む。

【発明の効果】

【0011】

したがって、本発明は、PCIeスイッチを管理する際に、PCIデバイスの新規追加や変更が生じる度に、PCIデバイスの詳細な情報（種別、性能等、管理を行うために必要な機能レベルの情報）を自動的に取得することが可能となつて、サーバに割り当ててるPCIデバイスを管理者が容易に探索可能な計算機システムを提供することができる。

【図面の簡単な説明】

【0012】

【図1】本発明の第1の実施形態を示し、計算機システムのブロック図である。

【図2】本発明の第1の実施形態を示し、管理サーバの構成を示すブロック図である。

【図3】本発明の第1の実施形態を示し、サーバ#0の構成を示すブロック図である。

【図4】本発明の第1の実施形態を示し、PCIeスイッチ#0の構成を示すブロック図である。

【図5】本発明の第1の実施形態を示し、PCIデバイステーブル217の一例を示す説明図である。

【図6】本発明の第1の実施形態を示し、PCIスイッチテーブル216の一例を示し、PCIeスイッチ#0のテーブルに関する説明図である。

【図 7】本発明の第 1 の実施形態を示し、サーバ管理テーブル 2 1 5 の一例を示す説明図である。

【図 8】本発明の第 1 の実施形態を示し、P C I デバイスグループテーブル 2 1 8 の一例を示す説明図である。

【図 9】本発明の第 1 の実施形態を示し、P C I デバイス管理部 2 0 0 の P C I デバイス情報取得機能 2 1 3 の処理の一例を示すシーケンス図である。

【図 1 0】本発明の第 1 の実施形態を示し、P C I デバイス 1 2 0 を一時的に管理サーバ 1 0 1 へ割り当てた状態を示すブロック図である。

【図 1 1】本発明の第 1 の実施形態を示し、P C I デバイス情報取得機能 2 1 3 の詳細な処理の一例を示すフローチャートである。

10

【図 1 2】本発明の第 1 の実施形態を示し、上記図 1 1 のステップ 1 1 0 1 で行われるグループピング機能 2 1 4 の処理の一例を示すフローチャートである。

【図 1 3】本発明の第 1 の実施形態を示し、P C I デバイス割り当て機能 2 1 2 の処理の概要を示すシーケンス図である。

【図 1 4】本発明の第 1 の実施形態を示し、管理サーバ 1 0 1 の P C I デバイス割り当て機能 2 1 2 の処理の詳細を示すフローチャートである。

【図 1 5】本発明の第 1 の実施形態を示し、管理サーバ 1 0 1 の P C I デバイス割り当て機能 2 1 2 の処理の詳細を示すフローチャートである。

【図 1 6】本発明の第 1 の実施形態を示し、グループ割り当て画面 1 5 0 1 の画面イメージである。

20

【図 1 7】本発明の第 2 の実施形態を示し、管理サーバのソフトウェア構成を示すブロック図である。

【図 1 8】本発明の第 2 の実施形態を示し、管理サーバで行われる P C I デバイス情報取得機能の処理の一例を示すシーケンス図である。

【図 1 9】本発明の第 3 の実施形態を示し、管理サーバ 1 0 1 の P C I デバイス割り当て機能 2 1 2 の処理の一例を示すフローチャートである。

【図 2 0】本発明の第 3 の実施形態を示し、グループ抽出画面の画面イメージである。

【発明を実施するための形態】

【 0 0 1 3 】

以下、本発明の一実施形態を添付図面に基づいて説明する。

30

【 0 0 1 4 】

図 1 は本発明の第 1 の実施形態による計算機システムのブロック図である。図 1 において計算機システムは、計算機である 1 以上のサーバ（ブレード）1 0 2 - 1、1 0 2 - 2 と、I / O デバイスとしての P C I デバイス 1 2 0 ~ 1 2 8 とサーバ 1 0 2 - 1、1 0 2 - 2 の接続状態を変更可能な 1 以上の P C I e スイッチ（または P C I スイッチ）1 0 3 - 1、1 0 3 - 2 と、I / O デバイスとして、1 以上の P C I デバイス 1 2 0 ~ 1 2 8 と、計算機システムのサーバ 1 0 2 - 1、1 0 2 - 2 及び P C I e スイッチ 1 0 3 - 1、1 0 3 - 2 の管理を行う管理サーバ 1 0 1 を主体に構成されている。

【 0 0 1 5 】

なお、図 1 においては、物理計算機としてのサーバ 1 0 2 - 1、1 0 2 - 2 が 2 つ、P C I e スイッチ 1 0 3 - 1、1 0 3 - 2 が 2 つの構成を示し、2 つのサーバ 1 0 2 - 1、1 0 2 - 2 は、サーバ # 0、サーバ # 1 として識別し、2 つの P C I スイッチ 1 0 1 1、1 0 1 2 は、P C I e スイッチ # 0、P C I e スイッチ # 1 として識別する。

40

【 0 0 1 6 】

管理サーバ 1 0 1 とサーバ # 0、# 1 及び P C I e スイッチ # 0、# 1 はネットワークインターフェース（N I C）を介して L A N スイッチ 1 0 4 に接続される。管理サーバ 1 0 1 は、L A N スイッチ 1 0 4 を介してサーバ # 0、# 1 へ P C I e スイッチ # 0、# 1 に接続された P C I デバイス 1 2 0 ~ 1 2 8 を割り当てる。

【 0 0 1 7 】

また、サーバ # 0、# 1 と管理サーバ 1 0 1 は、P C I e スイッチインターフェース 3

50

04、204を備えており、PCIeスイッチインターフェースのポートをPCIeスイッチ#0、#1のポートにピアツーピアで接続して、PCIデバイス120～128を利用する。

【0018】

サーバ#0、#1は、各サーバの初期化や電源の投入、遮断などの稼動に関する制御を行うBMC(Baseboard Management Controller)120を備える。BMC120はLANスイッチ104を介して管理サーバ101に接続され、管理サーバ101はサーバ#0、#1の障害の検知や、電源の投入または遮断を行うことができる。

【0019】

ここで、PCIeスイッチ#0、#1は、マルチルートI/O仮想化技術(Multi Root I/O Virtualization: MR-IOV)またはシングルルートI/O仮想化技術(Single Root I/O Virtualization: SR-IOV)に対応する。上記MR-IOVまたはSR-IOVはPCI-Sigで標準化されたものを適用すればよい。また、サーバ#0、#1、管理サーバ101のPCIeスイッチインターフェース304、204、PCIeスイッチ#0、#1及びPCIデバイス120～128は、上記PCI-SIGが策定したPCI Expressの規格に準拠したものである。

【0020】

管理サーバ101は、後述するように、サーバ#0、#1とPCIeスイッチ#0、#1の接続状態と、PCIeスイッチ#0、#1とPCIデバイス120～127の接続状態を管理するPCIデバイス管理部を備え、各サーバ#0、#1にPCIデバイス120～127を割り当てる。管理サーバ101はLANスイッチ104を介してPCIeスイッチ#0、#1の管理テーブルを設定することで、サーバ#0、#1とPCIデバイス120～127の接続状態を更新する。そして、管理サーバ101は、LANスイッチ104を介してPCIデバイス120～127及びPCIeスイッチ#0、#1の状態をサーバ#0、#1に通知することができる。LANスイッチ104は、計算機システムの管理ネットワークとして機能する。なお、サーバ#0、#1が外部の計算機と通信を行うためのLANスイッチを別途設けても良い。

【0021】

以下に、計算機システムの各部の詳細について説明する。

【0022】

<サーバ>

図3は、サーバ#0(102-1)の構成を示すブロック図である。サーバ#0、#1は、同一の構成の計算機であるので、サーバ#0(102-1)についてのみ説明する。

【0023】

サーバ#0は、1以上のCPU(プロセッサ)302と、メモリ301と、記憶媒体としてのローカルディスク303と、PCIeスイッチ#0、#1と接続するためのPCIeスイッチインターフェース304と、LANスイッチ104に接続するためのネットワークインターフェース305と、サーバ#0の稼動状態の監視と、電源の制御を行うBMC120と、を主体にして構成される。なお、図3の例では、PCIeスイッチインターフェース304は、2つのポート#0、#1(304-1、304-2)を備え、ポート#0をPCIeスイッチ#0のポート#0に接続し、ポート#1をPCIeスイッチ#1のポート#0に接続した例を示す。

【0024】

メモリ301にはOS310がロードされてCPU302により実行される。

【0025】

<PCIeスイッチ>

図4は、PCIeスイッチ#0(103-1)の構成を示すブロック図である。PCIeスイッチ#0、#1は、同一の構成のI/Oスイッチであるので、PCIeスイッチ#0(103-1)についてのみ説明する。

【0026】

10

20

30

40

50

PCIeスイッチ#0は、ポート#0～#7(30-1～30-8)をピアツーピアで接続するクロスバススイッチ2006と、クロスバススイッチ2006を制御する制御プロセッサ2002と、メモリ2001と、LANスイッチ104に接続するネットワークインターフェース2003と、を主体に構成される。

【0027】

本実施形態では、ポート#0～7のうち、ポート#0～3がサーバ#0、#1、管理サーバ101等の計算機に接続される上流ポートを構成し、ポート#4～#7がPCIデバイスに接続される下流ポートを構成する。なお、下流ポート#4～#7は、複数のPCIデバイスを収容可能なI/Oドロワ(図示省略)内のスロットに接続されても良い。

【0028】

メモリ2001には、クロスバススイッチ2006を制御するソフトウェアとしてのクロスバススイッチ制御部2004がロードされ、制御プロセッサ2002により実行される。また、メモリ2001にはクロスバススイッチ制御部2004が利用するポート管理テーブル2005が格納される。ポート管理テーブル2005は、図示はしないが、上流ポートとしてのポート#0～#3と、下流ポートとしてのポート#4～#7の接続状態と、共有または占有の状態を管理する。

【0029】

クロスバススイッチ制御部2004は、ネットワークインターフェース2003を介して管理サーバ101からの指令を受け付けて、上流ポートとしてのポート#0～#3と、下流ポートとしてのポート#4～#7の接続関係を決定し、ポート管理テーブル2005を更新する。クロスバススイッチ制御部2004は、管理サーバ101からの要求に応じてポート管理テーブル2005の内容を送信する。

【0030】

また、クロスバススイッチ制御部2004は、ポート#0～#7の接続状態が変化すると、管理サーバ101に変化が発生したことを通知する。接続状態の変化は、例えば、下流ポートに新たなPCIデバイスが接続されたり、既存のPCIデバイスが撤去された場合や、上流ポートに新たなサーバが接続されたり、既存のサーバが撤去された場合を示す。これらの状態が変化したときに、クロスバススイッチ制御部2004は、管理サーバ101に状態の変化の内容(例えば、ポートの識別子とサーバまたはPCIデバイスの識別子)を通知する。なお、クロスバススイッチ制御部2004は、PCIeスイッチ#0の下流ポートに新たなPCIeスイッチが接続されたり、既存のPCIeスイッチが撤去された場合も、PCIデバイスの変更と同様にして、管理サーバ101に接続状態の変化の内容を通知する。このように、PCIeスイッチ#0はポート#0～#7の接続状態を監視して、接続状態に変化が生じたときには管理サーバ101に変化の内容を通知する。

【0031】

<管理サーバ>

図2は、管理サーバ101の構成を示すブロック図である。管理サーバ101は、1以上のCPU(プロセッサ)202と、メモリ201と、記憶媒体としてのローカルディスク203と、PCIeスイッチ#0、#1と接続するためのPCIeスイッチインターフェース204と、LANスイッチ104に接続するためのネットワークインターフェース205と、マウスやキーボードで構成されて入力や操作を受け付ける入力装置230と、ディスプレイ等で構成されてユーザインタフェースを提供する表示装置220と、を主体にして構成される。なお、図1の例では、PCIeスイッチインターフェース204は、2つのポート#0、#1(204-1、204-2)を備え、ポート#0をPCIeスイッチ#0のポート#2に接続し、ポート#1をPCIeスイッチ#1のポート#2に接続した例を示す。

【0032】

メモリ201には、OS210がロードされてCPU202により実行される。OS210上ではPCIデバイス管理部200が実行される。PCIデバイス管理部200はローカルディスク203に格納されたプログラムで構成され、CPU202がローカルディ

10

20

30

40

50

スク 203 からメモリ 201 に P C I デバイス管理部 200 をロードし、C P U 202 により実行される。

【 0033 】

O S 210 は、P C I デバイス管理部 200 の実行を管理し、後述するように、P C I デバイスの詳細な情報を取得する際には、予め用意したデバイスドライバ 211 を読み込んで、デバイスドライバ 211 を介して P C I デバイスの詳細な情報を取得する。なお、デバイスドライバ 211 は、O S 210 に標準で搭載されるもの以外についてはローカルディスク 203 に格納しておき、適宜読み出すことができる。

【 0034 】

P C I デバイス管理部 200 は、P C I e スイッチ # 0、# 1 の設定の変更と、サーバ # 0、# 1 への P C I デバイス 120 ~ 127 の割り当てを管理する。このため、P C I デバイス管理部 200 は、後述する各テーブルの情報に基づいて P C I デバイス 120 ~ 127 とサーバ # 0、# 1 の割り当て状態を管理する P C I デバイス割り当て機能 212 と、P C I デバイス 120 ~ 127 の詳細な情報を取得して P C I デバイステーブル 217 と P C I デバイスグループテーブル 218 を更新する P C I デバイス情報取得機能 213 と、P C I デバイスグループテーブル 218 の情報から管理サーバ 101 を利用する管理者に P C I デバイス 120 ~ 127 を種別毎等に提供するグループピング機能 214 を備える。

10

【 0035 】

上記 P C I デバイス管理部 200 の各機能部については後述する。

20

【 0036 】

次に、P C I デバイス管理部 200 が利用する各テーブルの詳細について以下に説明する。各テーブルはメモリ 201 に格納され、P C I デバイス管理部 200 の各機能部によって設定され、更新される。

【 0037 】

図 7 は、サーバ管理テーブル 215 の一例を示す説明図である。サーバ管理テーブル 215 は、P C I デバイス割り当て機能 212 がサーバ # 0、# 1 (以下、サーバ # 0、# 1 の総称を単にサーバとする) と P C I e スイッチ # 0、# 1 (以下、P C I e スイッチ # 0、# 1 の総称を単に P C I e スイッチとする) から取得した情報と、管理サーバ 101 を利用する管理者が設定した情報に基づいて設定される。また、管理サーバ 101 の管理者がサーバ管理テーブル 215 を設定するようにしても良い。

30

【 0038 】

サーバ管理テーブル 215 は、サーバの識別子を格納するサーバ I D 611 と、サーバの P C I e スイッチインターフェース 304 のポート 304 - 1 または 304 - 2 の識別子を格納する P C I e ポート番号 612 と、サーバのポートに接続された P C I e スイッチ # 0、# 1 の識別子を格納する接続 P C I スイッチ I D 613 と、サーバのポートが接続されている P C I e スイッチのポートの識別子を格納する接続 P C I スイッチポート番号 614 と、P C I e スイッチの下流ポートに接続された P C I デバイス 120 ~ 127 (以下、P C I デバイス 120 ~ 127 の総称を単に P C I デバイスとする) に割り当てられるグループの識別子 (グループ I D) を格納する適用グループ I D 615 からひとつのエントリが構成される。なお、各識別子は、図 1 において「#」を付した符号を識別子とし、例えば、サーバ 102 - 1 は、サーバ # 0 であるので、識別子は「0」とする。

40

【 0039 】

適用グループ I D 615 は、後述するように、管理者がサーバに割り当てたい P C I デバイスのグループ I D を指定したものである。P C I デバイスのグループ I D は、P C I デバイスの種類や性能、ベンダ、適合するドライバなどの違いに応じて後述するように指定されたものである。

【 0040 】

サーバ I D 611 や P C I e ポート番号 612、接続 P C I スイッチ I D 613、接続 P C I スイッチポート番号 614 は、サーバの起動時やリセットをかけた際に、P C I e

50

スイッチインターフェース304が取得したPCIバスツリーの情報を、管理サーバ101のPCIデバイス割り当て機能212が読み込むことでサーバ管理テーブル215を更新することができる。

【0041】

図6は、PCIスイッチテーブル216の一例を示し、PCIeスイッチ#0のテーブルに関する説明図である。PCIスイッチテーブル216は、PCIデバイス割り当て機能212がサーバにPCIデバイスを割り当てるとき等に更新される。また、管理サーバ101の管理者がPCIスイッチテーブル216を設定するようにしても良い。なお、図示はしないが、PCIeスイッチ#1についても同様のテーブルが構成される。

【0042】

PCIスイッチテーブル216は、PCIeスイッチの識別子を格納するPCIeスイッチID511と、PCIeスイッチの下流ポートの識別子を格納するポート番号512と、当該ポート番号512に接続されたPCIデバイスの識別子を格納する搭載PCIデバイスID513と、当該下流ポートに接続されたPCIデバイスが共有または占有の何れであることを示す割り当て状態514と、当該下流ポートに接続されたPCIデバイスを共有することができるか否かを示す共有可否516と、当該PCIデバイスの接続先の装置の情報を格納するトポロジ情報515と、からひとつのエントリが構成される。

【0043】

PCIeスイッチID511と、ポート番号512と、搭載PCIデバイスIDは、PCIデバイス割り当て機能212がPCIeスイッチから取得する。割り当て状態514と共有可否516及びトポロジ情報515は、管理サーバ101を利用する管理者が設定することができる。割り当て状態514が「共有」の場合には、PCIデバイスが複数のサーバで共有されていることを示し、「占有」または「-」の場合にはPCIデバイスがひとつのサーバに割り当てられていることを示す。共有可能516が「可」の場合には、搭載PCIデバイスIDでのPCIデバイスが複数のサーバで共有することが可能であることを示す。なお、共有が不可能な場合には、ブランクまたは「否」が設定される。トポロジ情報515は、PCIデバイスにはどのような装置が接続されているかを管理者などによって設定することができる。例えば、PCIデバイスの種類がNICの場合、ネットワークのセグメントの情報等であり、HBA(Host Bus Adaptor)の場合では接続されているストレージ装置の識別子などが格納され、管理者が当該PCIデバイスを他のサーバに割り当てる際に参考となる情報が格納される。

【0044】

図5は、PCIデバイステーブル217の一例を示す説明図である。PCIデバイステーブル217は、PCIデバイス情報取得機能213が後述する処理により各PCIデバイスから取得した情報を設定したものである。また、管理サーバ101の管理者が設定してもよい。

【0045】

PCIデバイステーブル217は、PCIデバイスの識別子を格納するPCIデバイスID411と、PCIデバイスの種別を格納するカード種別412と、PCIデバイスのベンダ名を格納するベンダ413と、PCIデバイスの転送速度や対応規格などの性能情報を格納する性能414と、PCIデバイスが備えるポートの数を格納するポート数415と、PCIデバイスを動作させることが可能なデバイスドライバ211の情報を格納する適合ドライバ416と、PCIデバイスのファームウェアのバージョンを格納するファームウェアリビジョン417と、からひとつのエントリが構成される。

【0046】

PCIデバイス情報取得機能213は、後述する処理により、PCIデバイスID411、カード種別412、ベンダ413、性能414、ポート数415、ファームウェアリビジョン417を取得してPCIデバイステーブル217を更新する。適合ドライバ416については、管理サーバ101を利用する管理者が設定することができる。適合ドライバ416の情報としては、OSに標準で添付されているデバイスドライバ211で当該P

10

20

30

40

50

PCIデバイスを認識可能なOSの種類やバージョン、カーネルバージョン、サービスパック・バージョン等を設定するのが望ましい。あるいは、デバイスドライバ211のバージョン情報をOSの種類毎に設定してもよい。

【0047】

図8は、PCIデバイスグループテーブル218の一例を示す説明図である。PCIデバイスグループテーブル218は、後述するように、グルーピング機能214により作成されるものである。

【0048】

PCIデバイスグループテーブル218は、グループIDを格納するグループID711と、PCIデバイスの種別を示すカード種別712と、PCIデバイスのベンダ名を格納するベンダ713と、PCIデバイスの性能情報を格納する性能714と、PCIデバイスに実装されたポート数715と、PCIデバイスを動作させるのに必要なデバイスドライバ211を備えたOSの種類やバージョンを格納する適合ドライバ716、PCIデバイスのファームウェアのバージョンを格納するファームウェアリビジョン717と、PCIデバイスに接続された装置の情報を格納するトポロジ情報718と、からひとつのエントリが構成される。

【0049】

例えば、グループID=0の例では、PCIデバイスの種類がネットワークインターフェース(NIC)のものを検索することになり、グループID=Nの場合では、PCIデバイスの種類がネットワークインターフェースで、かつ、デバイスドライバ211がWin2k3sp2に標準で添付されているドライバを利用可能なものを検索する。

【0050】

<PCIデバイス管理部の処理>

次に、PCIデバイス管理部200の各機能で行われる処理の一例について、以下に説明する。

【0051】

<PCIデバイス情報取得機能>

図9は、PCIデバイス管理部200のPCIデバイス情報取得機能213の処理の一例を示すシーケンス図である。まず、ステップ811では、PCIeスイッチがPCIデバイスの追加または変更を検知して、LANスイッチ104を介して管理サーバ101に通知する。変更の検知は、クロスバススイッチ制御部2004が、クロスバススイッチ2006の下流ポート#4~#7に接続されたPCIデバイスが、ポート管理テーブル2005と一致しないときに追加または変更を検知して、接続位置として下流ポートを特定する。そしてクロスバススイッチ制御部2004は管理サーバ101にPCIデバイスの変更が発生したことを通知する。また、クロスバススイッチ制御部2004は、ポート管理テーブル2005を更新する。この更新は、変更のあった下流ポートに接続されたPCIデバイスを更新する。

【0052】

管理サーバ101は、ステップ801でPCIeスイッチからPCIデバイスの変更または追加の通知を受信し、PCIデバイスに変更が生じたことを検知する。なお、PCIeスイッチは、当該PCIeスイッチが新規に計算機システムへ追加されたときにもステップ811と同様の通知を発行する。

【0053】

ステップ802では、管理サーバ101は、変更のあったPCIデバイスの基本情報と接続位置をPCIeスイッチに対して要求する。基本情報としては、PCIデバイスID(デバイス番号)、カード種別、ベンダ名などである。接続位置は、下流ポートの識別子である。

【0054】

ステップ812では、PCIeスイッチのクロスバススイッチ制御部2004が、管理サーバ101の要求を受信して、クロスバススイッチ2006を介して変更が生じたPCIデ

10

20

30

40

50

バースにアクセスし、基本情報を取得する。基本情報の取得はP C Iバスプロトコルに基づいて取得する。そして、P C I eスイッチは、取得した基本情報を管理サーバ101に送信する。

【0055】

ステップ803では、基本情報を受信した管理サーバ101が、P C I eスイッチに対して追加または変更が生じた下流ポートのP C Iデバイスを、管理サーバ101に割り当てるように要求する。なお、管理サーバ101は、P C IデバイスID等からP C Iスイッチテーブル216を参照して、変更が生じたP C Iデバイスが新規(追加)または変更の何れであるかを判定し、新規または追加のP C Iデバイスの場合には、当該P C Iデバイスを占有で割り当てるように要求し、P C Iデバイスの変更の場合には共有で割り当てるように要求しても良い。

10

【0056】

ステップ813では、P C I eスイッチが管理サーバ101からの割り当て要求を受信し、割り当て対象のP C Iデバイスを管理サーバ101に割り当てる。すなわち、P C I eスイッチは、ポート管理テーブル2005を更新してクロススイッチ2006の設定を変更し、下流ポートと上流ポートの接続関係を変更する。例えば、図1に示したP C I eスイッチ#0のポート#4に接続されたP C Iデバイス120に変更があった場合、P C I eスイッチ#0は、ポート#4をポート#2(30-3)に接続し、図10で示すように管理サーバ101のポート#0とP C Iデバイス120を通信する。図10は、P C Iデバイス120を一時的に管理サーバ101へ割り当てた状態を示すブロック図である。このとき、P C Iデバイス120が既に他のサーバに割り当てられている場合には、P C I eスイッチ#0は、ポート管理テーブル2005を更新してP C Iデバイス120の割り当て状態を共有に変更する。そして、P C I eスイッチ#0は、ポート管理テーブル2005を更新する。なお、P C Iデバイス120が共有で割り当てることができない場合には、管理サーバ101にエラーを通知することができる。

20

【0057】

P C Iデバイス120を管理サーバ101に割り当てたP C Iスイッチは、割り当ての完了と、割り当ての内容を管理サーバ101に通知する。

【0058】

次に、ステップ804では、管理サーバ101は割り当ての完了通知と、割り当て内容を受信し、P C Iデバイスと接続されたので、ステップ803で受信した基本情報に基づいてデバイスドライバ211をロードする。このとき、管理サーバ101は、OS210に該当するデバイスドライバ211がない場合には、管理者にデバイスドライバ211を要求する。

30

【0059】

そして、管理サーバ101は、割り当てられたP C Iデバイスのデバイスドライバ211を介してP C Iデバイスの詳細な情報を取得する。詳細な情報は、P C Iデバイステーブル217に格納するための情報であり、例えば、性能情報(転送速度や対応する規格)、ポート数、ファームウェアリビジョンなどを取得する。なお、P C Iデバイスに適合するデバイスドライバ211については、OS210が保持している複数のデバイスドライバの一覧を表示装置220に出力して管理者に選択させるようにしてもよい。また、当該P C Iデバイスの接続先の情報を取得できる場合には、管理サーバ101は接続先の情報を読み込んでトポロジ情報としてメモリ201に記憶する。

40

【0060】

次に、ステップ805では、上記ステップ804で取得したP C Iデバイスの詳細な情報と、上記ステップ802で取得した基本情報からP C Iデバイステーブル217を更新する。管理サーバ101は、基本情報からP C IデバイスID411、カード種別412、ベンダ413を設定し、詳細な情報から性能414、ポート数415、適合ドライバ416、ファームウェアリビジョン417を設定する。なお、追加や変更の生じたP C Iデバイスと同一のP C Iデバイスの情報がP C Iデバイステーブル217に存在する場合に

50

は、新たなエントリーを生成せず、また、更新も行わない。

【 0 0 6 1 】

次に、ステップ 8 0 6 では、追加または変更が生じた P C I デバイスの詳細な情報の取得が完了したので、管理サーバ 1 0 1 は P C I e スイッチに対して、現在管理サーバ 1 0 1 が接続している P C I デバイスの割り当てを解除する要求を送信する。

【 0 0 6 2 】

ステップ 8 1 4 では、P C I e スイッチは管理サーバ 1 0 1 の割り当て解除要求を受信し、管理サーバ 1 0 1 と P C I デバイスの接続を解除する。このため、P C I e スイッチはクロスバスイッチ 2 0 0 6 を制御して管理サーバ 1 0 1 に接続された上流ポートと、接続対象の P C I デバイスが接続された下流ポートの通信経路を切断する。そして、管理サーバ 1 0 1 と接続していた P C I デバイスが共有の場合には、元の状態（例えば、占有）に戻す。そして、P C I e スイッチはポート管理テーブル 2 0 0 5 を更新して、管理サーバ 1 0 1 と接続する以前の状態に戻して割り当ての変更（接続解除）を完了する。P C I e スイッチは、管理サーバ 1 0 1 と P C I デバイスの接続の解除が完了すると、完了の通知を管理サーバ 1 0 1 に送信する。こうして、P C I e スイッチは、P C I デバイスの接続状態を、管理サーバ 1 0 1 に接続する以前の状態に復帰する。

【 0 0 6 3 】

ステップ 8 0 7 では、管理サーバ 1 0 1 が P C I e スイッチから接続解除の完了通知を受信し、P C I デバイスグループテーブル 2 1 8 を必要に応じて更新し、処理を終了する。P C I デバイスグループテーブル 2 1 8 の更新は、上記ステップ 8 0 5 で P C I デバイステーブル 2 1 7 に新たなエントリーを追加した場合は、新たな P C I デバイスが P C I e スイッチに接続されたので、この新たな P C I デバイスの基本情報と詳細情報を、後述するように P C I デバイスグループテーブル 2 1 8 に加える。

【 0 0 6 4 】

以上の処理により、P C I e スイッチに P C I デバイスが加えられたり変更（P C I デバイス取り外しや、P C I e スイッチの追加または削除）が生じると、P C I e スイッチからの通知により管理サーバ 1 0 1 の P C I デバイス情報取得機能 2 1 3 が起動して、管理サーバ 1 0 1 と変更が生じた P C I デバイスを一時的に接続する。

【 0 0 6 5 】

管理サーバ 1 0 1 の O S 2 1 0 は、接続された P C I デバイスのデバイスドライバ 2 1 1 を読み込んで、このデバイスドライバ 2 1 1 によって P C I デバイスの詳細な情報を取得することができ、取得した詳細な情報を P C I デバイステーブル 2 1 7 に設定する。P C I デバイステーブル 2 1 7 を更新すると、管理サーバ 1 0 1 と P C I デバイスの接続を解除して元の状態へ戻すことにより、変更の生じた P C I デバイスをサーバが利用することができる。

【 0 0 6 6 】

図 1 1 は、図 9 のステップ 8 0 1 ~ 8 0 7 で行われる P C I デバイス情報取得機能 2 1 3 の詳細な処理の一例を示すフローチャートである。

【 0 0 6 7 】

ステップ 1 0 0 1 では、P C I デバイス情報取得機能 2 1 3 が P C I e スイッチから変更があったことを示す通知を受信し、P C I デバイスまたは P C I e スイッチの追加、削除などの変更が生じたことを検知する。

【 0 0 6 8 】

ステップ 1 0 0 2 では、上記ステップ 8 0 2、8 1 2 で示したように、管理サーバ 1 0 1 は通知を受けた P C I デバイスの基本情報と接続位置を P C I e スイッチに要求し、P C I e スイッチから基本情報（P C I デバイス I D）と接続位置を取得する。

【 0 0 6 9 】

ステップ 1 0 0 3 では、P C I デバイス情報取得機能 2 1 3 が P C I e スイッチから取得した P C I デバイスについて P C I デバイステーブル 2 1 7 等に登録または更新する必要があるか否かを判定する。この判定は、ステップ 1 0 0 2 で取得した P C I デバイスに

10

20

30

40

50

ついて詳細な情報を取得していなければステップ1004に進み、既に詳細な情報を取得していればステップ1011のPCIデバイスグループテーブル218の作成へ進む。なお、ステップ1001で検知した変更がPCIデバイスの削除（取り外し）である場合には、ステップ1006～1008、100Aの処理を行わず、PCIデバイステーブル217とPCIスイッチテーブル216の更新のみを行えばよい。

【0070】

ステップ1004では、PCIデバイス情報取得機能213がPCIデバイステーブル217とPCIスイッチテーブル216を参照し、ステップ1005に進んで、ステップ1002で取得したPCIデバイスID及び接続位置のPCIデバイスが、新たに加えられたPCIデバイスであるか否かを判定する。すなわち、PCIデバイス情報取得機能213は、PCIデバイステーブル217に上記取得したPCIデバイスIDが存在しなければ、新たに加えられたPCIデバイスと判定して、ステップ1006に進んで、当該PCIデバイスを管理サーバ101が占有するようPCIeスイッチに割り当てを要求する。

10

【0071】

一方、上記取得したPCIデバイスIDがPCIスイッチテーブル216に存在し、接続位置（下流ポートの識別子）の変更の場合で、共有可否516が「可」であれば、ステップ1007へ進んで、当該PCIデバイスを管理サーバ101と他のサーバで共有するようPCIeスイッチに割り当てを要求する。なお、共有可否516が「可」でない場合には、表示装置220へエラーを通知するようにしても良い。

20

【0072】

この要求により、上記図8のステップ813で述べたようにPCIeスイッチは、変更の生じたPCIデバイスの割り当て状態を変更し、管理サーバ101に占有または共有で割り当てる。

【0073】

次に、ステップ100Bでは、上記図8のステップ804の処理を行い、管理サーバ101に接続されたPCIデバイスをOS210が認識し、OS210がデバイスドライバ211を検索する。PCIデバイス情報取得機能213は、OS210がPCIデバイスのデバイスドライバ211を読み込んだか否かを判定する。

【0074】

OS210がPCIデバイスのデバイスドライバ211を読み込めなかった場合には、ステップ100Cへ進んでPCIデバイス情報取得機能213が、表示装置220へデバイスドライバを要求する通知を送信し、管理者が手動でデバイスドライバ211をインストールするように通知する。デバイスドライバ211がインストールされた後に、ステップ1008へ進む。

30

【0075】

ステップ1008では、PCIデバイス情報取得機能213が、デバイスドライバ211を介して、PCIデバイスの詳細な情報を上記ステップ804で述べたように取得する。

【0076】

ステップ1009では、PCIデバイス情報取得機能213が、ステップ1008で取得したPCIデバイスの詳細な情報をPCIデバイステーブル217に格納する。

40

【0077】

ステップ100Aでは、PCIデバイス情報取得機能213が、詳細な情報の取得が完了したPCIデバイスの管理サーバ101との接続を遮断するため、PCIeスイッチに当該PCIデバイスの管理サーバ101への割り当てを解除する要求を送信する。

【0078】

そして、ステップ1002の処理に戻って、追加または更新するPCIデバイスがなくなるまで上記ステップ1002～100Aを繰り返す。

【0079】

50

ステップ1002の処理で、変更が生じた全てのPCIデバイスについて詳細な情報の取得が完了すると、グルーピング機能214によるステップ1101の処理へ進む。

【0080】

以上の処理によって、PCIeスイッチにPCIデバイスが追加または変更された場合には、この変更のあったPCIデバイスを一時的に管理サーバ101に割り当てる。そして、OS210が当該PCIデバイスのデバイスドライバ211を読み込んで、デバイスドライバ211を介してPCIデバイス情報取得機能213がPCIデバイスの詳細な情報を取得することができる。

【0081】

<グルーピング機能機能>

図12は、上記図11のステップ1101で行われるグルーピング機能214の処理の一例を示すフローチャートである。

【0082】

ステップ1101では、グルーピング機能214がPCIデバイステーブル217を参照し、ステップ1102では、新たにエントリが追加されているか否かを判定する。新たにエントリが追加されている場合には、PCIデバイスの情報が追加されているのでステップ1103の処理へ進み、そうでなければステップ1104で終了する。

【0083】

ステップ1103では、グルーピング機能214が新たなPCIデバイスの情報からPCIデバイスをサーバに割り当てるときの条件として、所定の項目を抽出して、PCIデバイスグループテーブル218に追加する。所定の項目としては、カード種別+ベンダ名の検索条件や、カード種別+性能の検索条件、あるいは、カード種別+適合ドライバなどをそれぞれPCIデバイステーブル217から取得し、PCIデバイスグループテーブル218にグループIDを付与して追加する。

【0084】

<PCIデバイス割り当て機能>

図13は、PCIデバイス割り当て機能212の処理の概要を示すシーケンス図である。管理者などがサーバにPCIデバイスを割り当てる際に、管理サーバ101でPCIデバイス割り当て機能212の起動を指令することで開始される。

【0085】

PCIデバイス割り当て機能212は、起動するとステップ1221で図16に示すような、グループ割り当て画面を表示装置220に出力し、管理者などが操作する入力装置230からの入力を受け付ける。

【0086】

ステップ1211では、管理者などが入力装置230を操作して、表示装置220のグループ割り当て画面1501上で、PCIデバイスを割り当てるサーバの選択と、割り当てるPCIデバイスのグループ(検索条件)を指定する。

【0087】

図16は、グループ割り当て画面1501の画面イメージである。グループ割り当て画面1501は、サーバ選択部1502とポート選択部103と、検索条件指定部1510と、PCIデバイスグループテーブル表示部1520から構成される。管理者などが入力装置230を操作して、サーバ選択部1502でPCIデバイスを割り当てるサーバを指定し、ポート選択部1503でPCIデバイスを割り当てるポートを指定する。検索条件指定部1510は、複数の検索条件にチェックボックス1511と項目選択部1512を備え、管理者などが入力装置230を操作して適宜選択することができる。PCIデバイスグループテーブル表示部1520は、チェックボックス1521を備え、図8に示したPCIデバイスグループテーブル218のエントリを検索条件として選択することができる。

【0088】

図13のステップ1211で、管理者などがサーバ選択部1502でサーバを選択し、

10

20

30

40

50

ポート選択部 1503 でポートを選択し、チェックボックス 1521 を指定すると、ステップ 1222 で P C I デバイス割り当て機能 212 は、入力を受け付けて選択されたサーバに対してチェックボックス 1521 で指定された P C I デバイスのグループの適用を開始する。

【 0089 】

ステップ 1223 では、P C I デバイス割り当て機能 212 が、選択されたチェックボックス 1521 のエントリに対応する P C I デバイスグループテーブル 218 からグループ I D 711 と検索条件を取得する。

【 0090 】

ステップ 1224 では、P C I デバイス割り当て機能 212 が、サーバ管理テーブル 215 を参照し、サーバ選択部 1502 で指定されたサーバと、ポート選択部 1503 で指定されたポートから P C I デバイスを検索する P C I e スイッチの識別子を接続 P C I スイッチ I D 613 から取得する。

【 0091 】

ステップ 1225 では、上記ステップ 1224 で取得した接続 P C I スイッチ I D 613 の P C I e スイッチの P C I デバイスから、上記ステップ 1223 で取得した検索条件に適合する P C I デバイスを、P C I デバイステーブル 217 及び P C I スイッチテーブル 216 から抽出する。

【 0092 】

ステップ 1226 では、P C I デバイス割り当て機能 212 が上記ステップ 1225 で抽出した P C I デバイスを、指定されたサーバに割り当てるように、ステップ 1224 で取得した接続 P C I スイッチ I D 613 の P C I e スイッチに割り当て要求を送信する。

【 0093 】

ステップ 1231 では、P C I e スイッチが管理サーバ 101 からの割り当て要求を受信して、要求で指定された下流ポートの P C I デバイスを、上流ポートのサーバに接続するようクロスバスイッチ 2006 の接続を変更し、ポート管理テーブル 2005 を更新する。そして、P C I e スイッチは、割り当てが完了したことを管理サーバ 101 に通知する。

【 0094 】

ステップ 1227 では、管理サーバ 101 の P C I デバイス割り当て機能 212 が P C I e スイッチからの割り当て完了通知を受信し、割り当てを行った P C I デバイスの I D、ポートの識別子、割り当て状態を P C I スイッチテーブル 216 に設定し、サーバ管理テーブル 215 に上記グループ割り当て画面 1501 で指定したチェックボックス 1521 に対応する P C I デバイスグループテーブル 218 のグループ I D 711 を設定する。そして、P C I デバイス割り当て機能 212 は、割り当てが完了したことを表示装置 220 に出力する。

【 0095 】

以上の処理により、P C I デバイスの詳細な情報として、デバイスドライバ 211 に関する情報やファームウェアリビジョンを含めることで、P C I e スイッチで P C I デバイスをサーバ # 0、# 1 に割り当ての際に、各サーバの O S の種類やバージョンに対応する多数の P C I デバイスの中から最適な P C I デバイスを容易に提供することができる。

【 0096 】

これにより、P C I e スイッチに多数の P C I デバイスを接続している環境において、管理者は所望の P C I デバイスを、P C I デバイスグループテーブル 218 のグループ I D 711 を選択するだけでサーバに割り当てることが可能となる。

【 0097 】

図 14、図 15 は、管理サーバ 101 の P C I デバイス割り当て機能 212 の処理の詳細を示すフローチャートである。

【 0098 】

図 14 のステップ 1301 では、P C I デバイス割り当て機能 212 が P C I デバイス

10

20

30

40

50

グループテーブル 218 を参照し、ステップ 1302 で図 16 に示したグループ割り当て画面 1501 を表示装置 220 に出力する。そして、管理者などの入力装置 230 の操作を受け付けて指定されたサーバ、ポート、検索条件を取得する。

【0099】

次に、ステップ 1303 では、図 16 のグループ割り当て画面 1501 の「OK」ボタンまたは「キャンセル」ボタンがクリックされるのを待ち、「OK」ボタンがクリックされればステップ 1304 に進む。「キャンセル」ボタンがクリックされればそのまま終了する。

【0100】

ステップ 1304 では、PCI デバイスの割り当て処理を実行し、ステップ 1305 では割り当てに利用したグループ ID 711 をサーバ管理テーブル 215 に設定して一連の処理を完了する。

10

【0101】

図 15 は、上記図 14 のステップ 1304 で行われる PCI デバイス割り当て処理の詳細を示すフローチャートである。

【0102】

ステップ 1411 では、PCI デバイス割り当て機能 212 が、上記ステップ 1302 で指定されたチェックボックス 1521 のエントリに対応する PCI デバイスグループテーブル 218 からグループ ID 711 と検索条件を取得する。

20

【0103】

ステップ 1412 では、PCI デバイス割り当て機能 212 が、サーバ管理テーブル 215 を参照し、サーバ選択部 1502 で指定されたサーバと、ポート選択部 1503 で指定されたポートから PCI デバイスを検索する PCI e スイッチの識別子を接続 PCI スイッチ ID 613 から検索する。

【0104】

ステップ 1413 では、上記ステップ 1412 で取得した接続 PCI スイッチ ID 613 の PCI e スイッチの PCI デバイスから、上記ステップ 1411 で取得した検索条件に適合する PCI デバイスを、PCI デバイステーブル 217 及び PCI スイッチテーブル 216 から検索する。

30

【0105】

ステップ 1414 では、検索条件に一致した PCI デバイスのうち、未割り当ての PCI デバイスがあるか否かを判定する。未割り当ての PCI デバイスが存在する場合には、ステップ 1415 に進む。未割り当ての PCI デバイスが存在しない場合には、ステップ 1416 に進む。

【0106】

ステップ 1416 では、ステップ 1413 の検索結果の PCI デバイスのうち、既に他のサーバに割り当て済みである PCI デバイスについて PCI スイッチテーブル 216 を参照して、共有可否 516 が「可」の PCI デバイスの有無を判定する。共有可否 516 が「可」の PCI デバイスが存在する場合には、ステップ 1417 へ進んで、共有が可能な PCI デバイスで割り当て可能な PCI デバイスがあるか否かを判定する。割り当て可能な PCI デバイスが存在する場合には、ステップ 1418 へ進んで、当該 PCI デバイスを共有で割り当てよう設定する。一方、共有可能な PCI デバイスがない場合、または、割り当て可能な PCI デバイスがない場合には、ステップ 1419 に進んで、割り当て可能な PCI デバイスがないことを表示装置 220 に通知して処理を終了する。

40

【0107】

ステップ 1415 では、PCI デバイス割り当て機能 212 が、ステップ 1414 またはステップ 1415 で抽出した PCI デバイスを、指定されたサーバに割り当てるように、接続 PCI スイッチ ID 613 の PCI e スイッチに割り当て要求を送信する。

【0108】

その後、ステップ 141B で PCI e スイッチから割り当て完了の通知を受信してから

50

、PCIスイッチテーブル216を更新する。

【0109】

以上のように、本実施形態では、PCIeスイッチに新たなPCIデバイスが新たに追加された場合などでは、管理サーバ101とPCIデバイスを一時的に接続するようPCIeスイッチに指令して、管理サーバ101に新たなPCIデバイスを割り当てる。そして、管理サーバ101のOS210が接続されたPCIデバイスのデバイスドライバ211を読み込んで、PCIデバイス情報取得機能がPCIデバイスの詳細な情報を取得して、PCIデバイステーブル217とPCIデバイスグループテーブル218を更新しておく。そして、管理サーバ101と新たなPCIデバイスの接続を切断し、元の状態に戻しておく。

10

【0110】

サーバにPCIデバイスを割り当てる際には、PCIデバイス割り当て機能212がグループ割り当て画面1501を提供し、予めPCIデバイスグループテーブル218に設定した検索条件を管理者に選択させることで、極めて容易にサーバが必要とするPCIデバイスを割り当てることができるのである。

【0111】

<第2実施形態>

図17は、本発明の第2の実施形態を示し、管理サーバ101のソフトウェア構成を示すブロック図である。本実施形態では、前記第1実施形態の管理サーバ101で仮想化技術を用いて複数種のゲストOSを実行し、各種のゲストOSでPCIデバイスの詳細情報をそれぞれ取得するものである。その他の構成については、前記第1実施形態と同様である。

20

【0112】

本実施形態では、計算機システムではサーバ#0、#1毎に異なる種類のOSを実行でき、同一のPCIデバイスであっても各OS毎に使用するデバイスドライバは異なる。このため、管理サーバ101では、計算機システムで実行予定のOSの種類を、ゲストOS2200、2300として実行しておき、各ゲストOS2200、2300上でPCIデバイス情報取得機能を実行し、各OSの種類毎にデバイスドライバの情報など詳細情報を取得する。そして、各ゲストOS2200、2300は、それぞれ取得したPCIデバイスの詳細情報を、ホストOS2100で実行するPCIデバイス管理部200のPCIデバイステーブル217に格納する。これにより、計算機システムで利用可能なOSの種類に応じたPCIデバイスの詳細情報を取得しておくことで、PCIデバイスの割り当てを効率よく行うことができる。

30

【0113】

図17において、管理サーバ101では、図2に示したメモリ201上に仮想化ホストOS2100をロードしてCPU202が実行する。仮想化ホストOS2100上では、仮想化ゲストOS2200と、仮想化ゲストOS2300が実行される。仮想化ホストOS2100と、仮想化ゲストOS2200、仮想化ゲストOS2300は異なる種類のOSである。なお、仮想化技術としては、VMM(Virtual Machine Manager)やハイパーバイザを適宜使用することができる。

40

【0114】

仮想化ホストOS2100では、前記第1実施形態と同様にPCIデバイス管理部200を実行し、PCIデバイスを一時的に管理サーバ101へ割り当てたときに、デバイスドライバ2110によりPCIデバイスの詳細情報を前記第1実施形態と同様に取得する。

【0115】

仮想化ゲストOS2200、2300では、PCIデバイス情報取得機能2213、2313を実行し、PCIデバイスを一時的に管理サーバ101へ割り当てたときに、それぞれのOSの種類に応じたデバイスドライバ2210、2310によってPCIデバイスの詳細情報を前記第1実施形態と同様に取得する。そして、仮想化ゲストOS2200、

50

2300上のPCIデバイス情報取得機能2213、2313は、取得したPCIデバイスの詳細情報をホストOS2100上のPCIデバイス情報取得機能213に通知する。仮想化ホストOS2100上のPCIデバイス情報取得機能213は、ゲストOS2200、2300から受け付けたPCIデバイスの詳細情報をPCIデバイステーブル217に格納する。

【0116】

図18は、管理サーバ101で行われるPCIデバイス情報取得機能の処理の一例を示すシーケンス図である。

【0117】

まず、ステップ1631では、PCIeスイッチが前記第1実施形態と同様にしてPCIデバイスの追加または変更を検知して、LANスイッチ104を介して管理サーバ101に通知する。

【0118】

管理サーバ101は、ステップ1621でPCIeスイッチからPCIデバイスの変更または追加の通知を受信し、PCIデバイスに変更が生じたことを検知する。

【0119】

ステップ1622では、管理サーバ101は、変更のあったPCIデバイスの基本情報と接続位置をPCIeスイッチに対して要求する。基本情報としては、PCIデバイスID(デバイス番号)、カード種別、ベンダ名などである。接続位置は、下流ポートの識別子である。

【0120】

ステップ1632では、PCIeスイッチのクロスバススイッチ制御部2004が、管理サーバ101の要求を受信して、クロスバススイッチ2006を介して変更が生じたPCIデバイスにアクセスし、基本情報を取得する。基本情報の取得はPCIバスプロトコルに基づいて取得する。そして、PCIeスイッチは、取得した基本情報を管理サーバ101に送信する。

【0121】

ステップ1623では、基本情報を受信した管理サーバ101が、PCIeスイッチに対して追加または変更が生じた下流ポートのPCIデバイスを、管理サーバ101に割り当てるように要求する。

【0122】

ステップ1633では、PCIeスイッチが管理サーバ101からの割り当て要求を受信し、割り当て対象のPCIデバイスを管理サーバ101に割り当てる。PCIデバイス120を管理サーバ101に割り当てたPCIeスイッチは、割り当ての完了と、割り当ての内容を管理サーバ101に通知する。

【0123】

次に、ステップ1624では、管理サーバ101はPCIデバイスの割り当ての完了通知と、ステップ1623で受信した基本情報に基づいて仮想化ホストOS2100のデバイスドライバ211をロードする。このとき、管理サーバ101は、OS2100に該当するデバイスドライバ2110がない場合には、前記第1実施形態と同様にして管理者にデバイスドライバ211を要求する。

【0124】

次に、ステップ1625では、ホストOS2100上のPCIデバイス情報取得機能213が、ゲストOS2200、2300上のPCIデバイス情報取得機能2213、2313に対して、管理サーバ101に割り当てられたPCIデバイスの詳細情報の取得を指令する。

【0125】

ステップ1611では、ホストOS2100上のPCIデバイス情報取得機能213から指令を受けたゲストOS2200、2300は、それぞれのデバイスドライバ2210、2310をロードして、PCIデバイス情報取得機能2213、2313が、それぞれ

10

20

30

40

50

PCIデバイスの詳細情報を取得する。

【0126】

ゲストOS 2200、2300上のPCIデバイス情報取得機能2213、2313は、PCIデバイスの詳細情報を取得した後に、ホストOS 2100上のPCIデバイス情報取得機能213に取得した詳細情報を通知する。

【0127】

ステップ1626では、ホストOS 2100上のPCIデバイス情報取得機能213が、ゲストOS 2200、2300のPCIデバイス情報取得機能2213、2313からそれぞれPCIデバイスの詳細情報を受信する。

【0128】

ステップ1627では、ホストOS 2100上のPCIデバイス情報取得機能213が、自OS 2100のデバイスドライバ2110を介して取得したPCIデバイスの詳細情報と、ゲストOS 2200、2300のPCIデバイス情報取得機能2213、2313からそれぞれ取得した詳細情報をPCIデバイステーブル217に格納する。

【0129】

次に、ステップ1628では、追加または変更が生じたPCIデバイスの詳細な情報の取得が完了したので、管理サーバ101はPCIeスイッチに対して、現在管理サーバ101が接続しているPCIデバイスの割り当てを解除する要求を送信する。

【0130】

ステップ1634では、PCIeスイッチは管理サーバ101の割り当て解除要求を受信し、管理サーバ101とPCIデバイスの接続を前記第1実施形態と同様にして解除する。PCIeスイッチは、管理サーバ101とPCIデバイスの接続の解除が完了すると、完了の通知を管理サーバ101に送信する。

【0131】

ステップ1629では、管理サーバ101がPCIeスイッチから接続解除の完了通知を受信し、PCIデバイスグループテーブル218を前記第1実施形態と同様に更新して、一連の処理を終了する。

【0132】

以上の処理により、PCIeスイッチにPCIデバイスが加えられたり変更（PCIデバイス取り外しや、PCIeスイッチの追加または削除）が生じると、PCIeスイッチからの通知により管理サーバ101のPCIデバイス情報取得機能213が起動して、管理サーバ101と変更が生じたPCIデバイスを一時的に接続する。

【0133】

管理サーバ101では、ホストOS 2100上のPCIデバイス情報取得機能213と、ゲストOS 2200、2300のPCIデバイス情報取得機能2213、2313がそれぞれPCIデバイスの詳細情報を取得することで、OSの種類に応じた詳細情報を取得することができる。

【0134】

<第3実施形態>

図19は、本願発明の第3の実施形態を示し、管理サーバ101のPCIデバイス割り当て機能212の処理の一例を示すフローチャートである。

【0135】

本第3実施形態では、既存のサーバが利用中のPCIデバイスの構成を取得し、これと同じ条件でPCIデバイスのグループを作成するようにしたもので、その他の構成は前記第1実施形態と同様である。

【0136】

図19において、ステップ1801では、図20に示すユーザインターフェース（グループ抽出画面1901）を表示装置220に出力し、管理サーバ101を操作する管理者などから、PCIデバイスの抽出元のサーバ1910と、抽出したPCIデバイスの条件を適用する適用先サーバ1911を受け付ける。そして、PCIデバイス割り当て機能2

10

20

30

40

50

12は、サーバ管理テーブル215を参照して、抽出元サーバ1910に接続されたPCIデバイスの情報(適用グループID等)を取得する。

【0137】

ステップ1802では、PCIデバイス割り当て機能212がPCIデバイスグループテーブル218を参照し、ステップ1801で取得したPCIデバイスの適用グループIDに対応する検索条件を取得する。

【0138】

ステップ1803では、ステップ1801で受け付けた抽出元サーバ1910と抽出先サーバ1911を表示する。そして、抽出元サーバ1910と同一の適用グループID711で抽出先サーバ1911に割り当て可能なPCIデバイスをPCIスイッチテーブル216から抽出する。このとき、抽出元サーバ1910のポート数が抽出先サーバ1911のポート数を超えない範囲で、PCIデバイスを抽出する。

10

【0139】

ステップ1804では、ステップ1803の検索結果を図20の適用グループ1920に表示する。図20は、グループ抽出画面の画面イメージを示す。

【0140】

次に、ステップ1805では、図20のグループ抽出画面1901の「OK」ボタンまたは「キャンセル」ボタンがクリックされるのを待ち、「OK」ボタンがクリックされればステップ1806に進む。「キャンセル」ボタンがクリックされればそのまま終了する。

20

【0141】

ステップ1806では、図20のPCIデバイスのグループのうち選択されたPCIデバイスの割り当て条件(適用グループID711)でサーバ管理テーブル215の適用先サーバ1911のエントリを更新する。

【0142】

ステップ1807では、図20のPCIデバイスのグループで選択されたグループを適用先サーバ1911に割り当ててる処理を実行する。この処理は、前記第1実施形態との図15と同様である。

【0143】

上記処理によって、既存のサーバで動作しているPCIデバイスと同一のグループ(PCIデバイスの構成)を、他のサーバにも適用することができる。

30

【0144】

なお、上記各実施形態では、管理サーバ101でPCIスイッチ管理部200を実行した例を示したが、複数のサーバ#0、#1のうちのいずれかひとつでPCIスイッチ管理部200を実行してもよい。

【産業上の利用可能性】

【0145】

以上のように、本発明はPCIeスイッチを備えた計算機システムに適用することができ、特にPCIデバイスの管理システム、管理方法、管理プログラムに適用することができる。

40

【符号の説明】

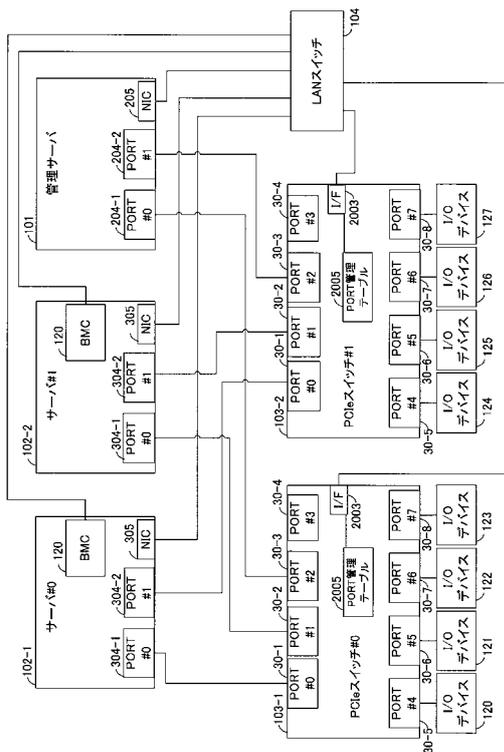
【0146】

- 101 管理サーバ
- 102 - 1 ~ 102 - 2 サーバ
- 103 - 1 ~ 103 - 2 PCIeスイッチ
- 104 LANスイッチ
- 200 PCIデバイス管理部
- 210 OS
- 211 デバイスドライバ
- 212 PCIデバイス割り当て機能

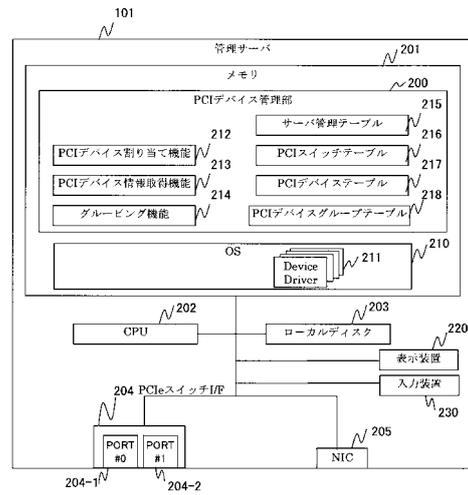
50

- 2 1 3 P C I デバイス情報取得機能
- 2 1 4 グルーピング機能
- 2 1 5 サーバ管理テーブル
- 2 1 6 P C I スイッチテーブル
- 2 1 7 P C I デバイステーブル
- 2 1 8 P C I デバイスグループテーブル

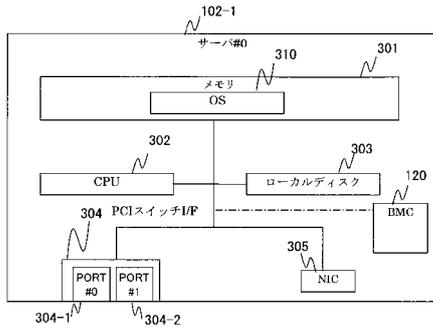
【 図 1 】



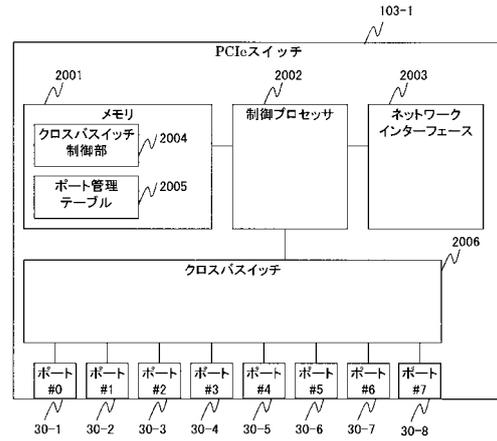
【 図 2 】



【図3】



【図4】



【図5】

217 PCIデバイステーブル

PCI Dev. ID	カード種別	ベンダ	性能 (転送速度)	ポート数	適合ドライバ	F/W Rev.
0	NIC	I社	1Gbps	1	W2KSP4.W2k3R2.W2k8.LX2.30.3	1.02
1	NIC	I社	1Gbps	2	W2KSP4.W2k3R2.W2k8.LX2.3.3	1.13
2	HBA	H社	10Gbps	2	W2k3R2.W2k8.LX2.3.3	2.66
3	HBA	G社	1Gbps	1	W2k3R2.W2k8.LX2.3.3	1.0.1

【図6】

216 PCIeスイッチテーブル

PCIeSW ID	Port#	搭載 PCI Dev ID	割り当て状態	共有可否	トポロジ情報
0	4	0	占有	可	(接続先の機器等の情報)
	5	1	占有	可	
	6	2	-	-	
	7	3	共有	可	

【図7】

215 サーバ管理テーブル

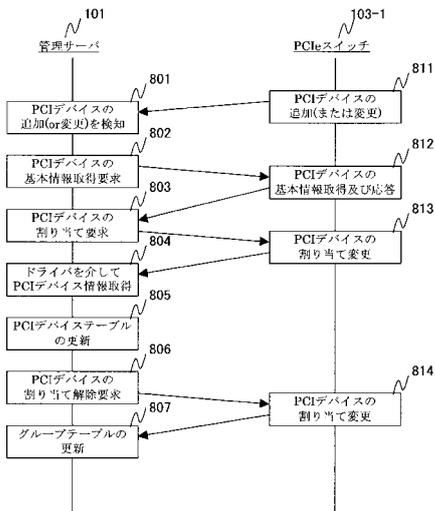
サーバID	PCIe Port#	接続 PCI SW ID	接続 PCI SW Port#	適用 グループID
0	0	0	0	2
	1	1	1	3
1	0	0	0	2
	1	1	1	2

【図8】

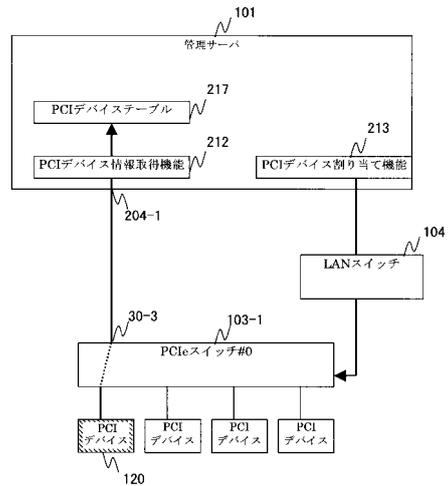
218 PCIデバイスグループテーブル

Group ID	カード種別	ベンダ	機能 (転送速)	ポート 数	適合ドライバ	P/W Rev	メモロシ 情報
0	NIC	-	-	-	-	-	-
1	HBA	-	-	-	-	-	-
2	NIC	住友	-	-	-	-	-
3	HBA	HPE	-	-	-	-	-
4	NIC	-	10Gbps	-	-	-	-
N	NIC	-	-	-	Win2k3 ep2の デフォルトドライバ	-	-

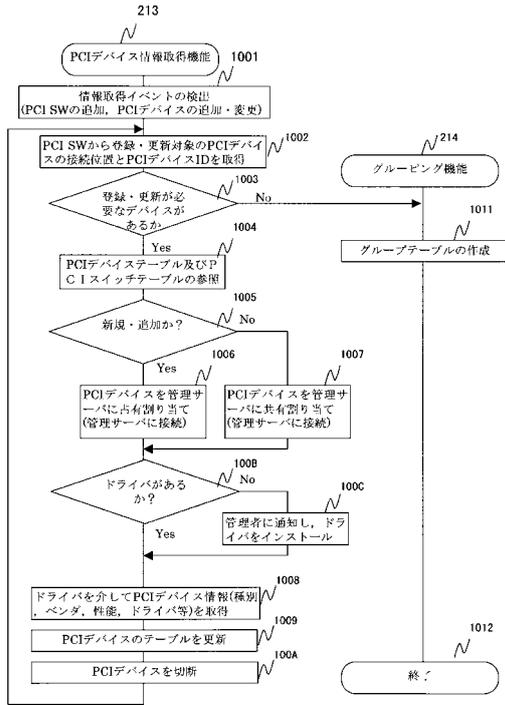
【図9】



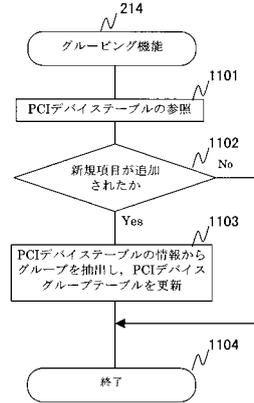
【図10】



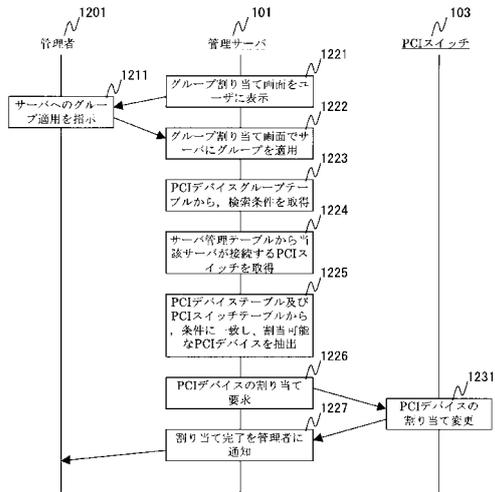
【図 1 1】



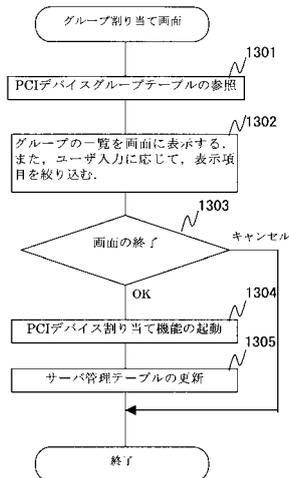
【図 1 2】



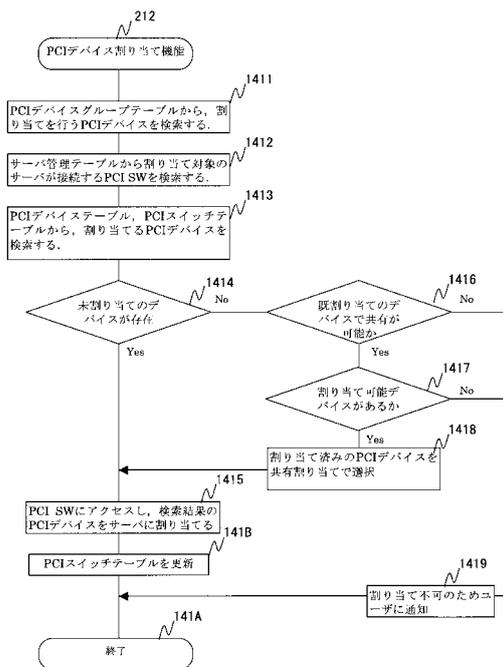
【図 1 3】



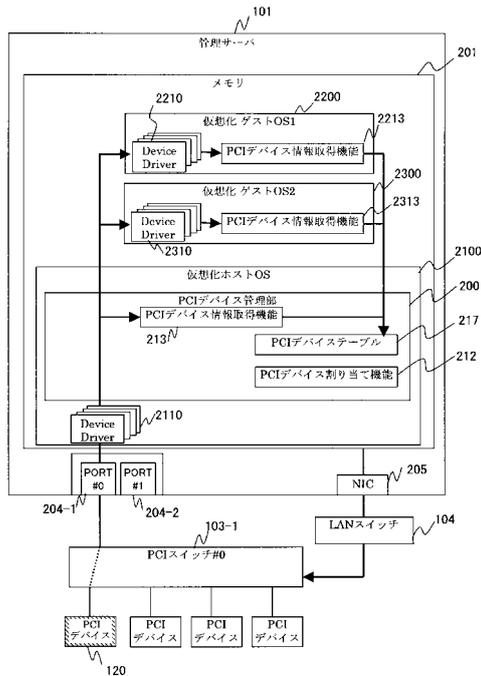
【図 1 4】



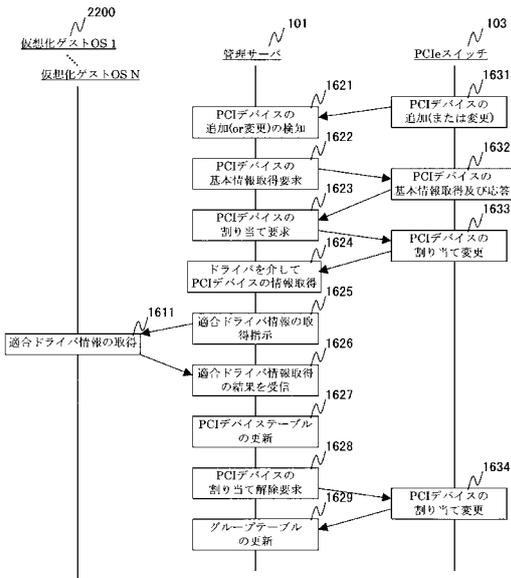
【図15】



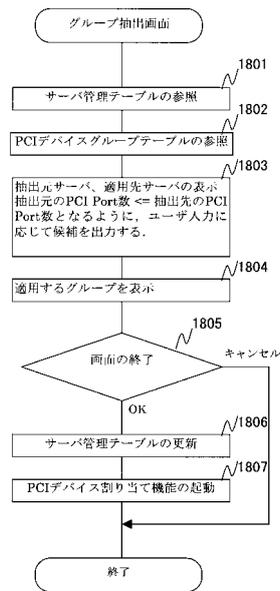
【図17】



【図18】



【図19】



【図16】

グループ割り当て画面

サーバの選択 Server1 PCI ポートの選択 #0

検索条件

1511 種別 指定なし ポート数 1
 ベンダ 指定なし 適合ドライバ 指定なし
 性能 指定なし F/W Rev 指定なし

PCI デバイスの共有を許可する
 割り当て可能な候補のみ表示する

種別	ベンダ	性能 (転送速度)	ポート数	適合ドライバ	F/W Rev
<input checked="" type="radio"/> NIC	-	-	-	-	-
<input type="radio"/> HBA	-	-	-	-	-
<input type="radio"/> NIC	I社	-	-	-	-
<input type="radio"/> HBA	H社	-	-	-	-
<input type="radio"/> NIC		10Gbps	-	-	-
<input type="radio"/> NIC		20Gbps			
<input type="radio"/> NIC	H社	10Gbps			

OK キャンセル

【 図 2 0 】

1901

グループ抽出画面

抽出元サーバの選択 ▼ 1910

適用先サーバの選択 ▼ 1911

1920

PCI slot#	種別	ベンダ	性能 (転送速度)	ポート数	適合ドライバ	F/W Rev
0	NIC	I社	1Gbps	2	(ドライバを識別する情報)	01-00
1	HBA	H社	4Gbps	2		02-10

OK キャンセル

フロントページの続き

(72)発明者 若松 峻彦

神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業部内

審査官 横山 佳弘

(56)参考文献 特開2005-317021(JP,A)

特開2008-257597(JP,A)

特開2000-339268(JP,A)

特開2009-187490(JP,A)

特開平09-272243(JP,A)

特開2008-152787(JP,A)

特開2002-158732(JP,A)

特開2007-094470(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 13/14

G06F 13/36

G06F 3/00