

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成 23 年 3 月 17 日 (2011.3.17)

【公開番号】特開 2010-108211 (P2010-108211A)
 【公開日】平成 22 年 5 月 13 日 (2010.5.13)
 【年通号数】公開・登録公報 2010-019
 【出願番号】特願 2008-279101 (P2008-279101)
 【国際特許分類】

G 0 6 F 3/06 (2006.01)

【 F I 】

G 0 6 F 3/06 3 0 5 Z

G 0 6 F 3/06 5 4 0

【手続補正書】
 【提出日】平成 23 年 1 月 27 日 (2011.1.27)
 【手続補正 1】
 【補正対象書類名】特許請求の範囲
 【補正対象項目名】全文
 【補正方法】変更
 【補正の内容】
 【特許請求の範囲】
 【請求項 1】

ストレージコントローラ内のコンポーネントが PCI Express で接続された内部ネットワークを持つストレージ装置であって、

前記ストレージコントローラ及びその内部ネットワークは、

全体で複数のルートポートを有し、それぞれが 1 つ以上のルートポートを有する複数のルートコンプレックスと、

前記ルートコンプレックス毎に接続される複数のプロセッサと、

所定のインターフェイスを備え、前記複数のルートポートのそれぞれからアクセス可能な複数のバーチャルファンクションを有するマルチルートアウェアデバイスである、複数のエンドポイントデバイスと、

前記複数のルートポートと前記複数のエンドポイントデバイスとを接続してデータ経路を構成し複数のポートを備えるマルチルートアウェアスイッチであるスイッチと、を有し、

前記複数のエンドポイントデバイスの少なくとも 1 つは、前記バーチャルファンクションと前記所定のインターフェイスとのマッピングを変更するバーチャルファンクションマイグレーションを実行可能であり、

前記ストレージコントローラ及びその内部ネットワークは、すべての前記複数のルートポートが、すべての前記複数のエンドポイントデバイス内のバーチャルファンクションへアクセス可能なように、前記スイッチ及び前記エンドポイントデバイスが設定されており、

前記ストレージコントローラ及びその内部ネットワークは、前記複数のルートポートの 1 つから前記スイッチを介して前記複数のエンドポイントデバイスの 1 つへのデータ経路の 1 つに障害が発生した場合に、データ経路のフェイルオーバーを実行する処理として、当該障害状態の第 1 のデータ経路から代替の第 2 のデータ経路への切り替えにおいて、当該データ経路のアクセス元ルートポートを切り替える処理と、当該アクセス先エンドポイントデバイスのバーチャルファンクションと前記所定のインターフェイスとのマッピングを前記バーチャルファンクションマイグレーションによって切り替える処理とを行うこと、を特徴とするストレージ装置。

【請求項 2】

請求項 1 記載のストレージ装置において、

前記エンドポイントデバイスとして、

上位装置を接続するためのフロントエンドインターフェイスを備え、前記バーチャルファンクションマイグレーションを実行可能な第 1 のエンドポイントデバイスと、

記憶装置群を接続するためのバックエンドインターフェイスを備え、前記バーチャルファンクションマイグレーションを実行可能な第 2 のエンドポイントデバイスと、を有し、

前記スイッチは、前記複数のルートポートと、前記第 1 及び第 2 のエンドポイントデバイスとを接続し、

前記ストレージコントローラ及びその内部ネットワークは、すべての前記複数のルートポートが、すべての前記複数のエンドポイントデバイス内のバーチャルファンクションへアクセス可能なように、前記スイッチ及び前記第 1 及び第 2 のエンドポイントデバイスが設定されており、

前記ストレージコントローラ及びその内部ネットワークは、前記複数のルートポートの 1 つから、前記第 1 または第 2 のエンドポイントデバイスの 1 つへのデータ経路の 1 つに障害が発生した場合に、データ経路のフェイルオーバーを実行する処理として、当該障害状態の第 1 のデータ経路から代替の第 2 のデータ経路への切り替えにおいて、当該データ経路のアクセス元ルートポートを切り替える処理と、当該アクセス先エンドポイントデバイスのバーチャルファンクションと前記所定のインターフェイスとのマッピングを前記バーチャルファンクションマイグレーションによって切り替える処理とを行うこと、を特徴とするストレージ装置。

【請求項 3】

請求項 2 記載のストレージ装置において、

前記エンドポイントデバイスとして、キャッシュメモリを接続するためのキャッシュメモリインターフェイスを備える第 3 のエンドポイントデバイスを有し、

前記スイッチは、前記複数のルートポートと、前記第 1 及び第 2 及び第 3 のエンドポイントデバイスを接続し、

すべての前記複数のルートポートが、すべての前記複数のエンドポイントデバイス内のバーチャルファンクションへアクセス可能なように、前記スイッチ及び前記第 1 及び第 2 及び第 3 のエンドポイントデバイスが設定されており、

前記複数のルートポートの 1 つから、前記第 1 または第 2 または第 3 のエンドポイントデバイスの 1 つへのデータ経路の 1 つに障害が発生した場合に、データ経路のフェイルオーバーを実行する処理として、当該障害状態の第 1 のデータ経路から代替の第 2 のデータ経路への切り替えにおいて、当該データ経路のアクセス元ルートポートを切り替える処理と、当該アクセス先エンドポイントデバイスのバーチャルファンクションと前記所定のインターフェイスとのマッピングを前記バーチャルファンクションマイグレーションによって切り替える処理とを行うこと、を特徴とするストレージ装置。

【請求項 4】

請求項 3 記載のストレージ装置において、

前記第 3 のエンドポイントデバイスは、前記バーチャルファンクションに対してマッピングされるミッションバーチャルファンクションを備え、前記ミッションバーチャルファンクションに対して前記キャッシュメモリインターフェイスが対応付けられ、当該第 3 のエンドポイントデバイスに備えるミッションバーチャルファンクションの数は、当該第 3 のエンドポイントデバイスにアクセスするルートポートの数以上であること、を特徴とするストレージ装置。

【請求項 5】

請求項 3 記載のストレージ装置において、

前記ストレージコントローラ及びその内部ネットワークにおいて、

前記第 1 のエンドポイントデバイスのミッションバーチャルファンクションとなる前記フロントエンドインターフェイスと前記第 3 のエンドポイントデバイスのミッションバー

チャルファンクションとなる前記キャッシュメモリインターフェイスとが同じルートポートからアクセス可能なように設定され、

前記第2のエンドポイントデバイスのミッションバーチャルファンクションとなる前記バックエンドインターフェイスと前記第3のエンドポイントデバイスのミッションバーチャルファンクションとなる前記キャッシュメモリインターフェイスとが同じルートポートからアクセス可能なように設定されていること、を特徴とするストレージ装置。

【請求項6】

請求項3記載のストレージ装置において、

前記複数のルートコンプレックスは、それぞれ、第1のルートポートと第2のルートポートとを有し、

前記ストレージコントローラ及びその内部ネットワークにおいて、

各ルートコンプレックスの前記第1のルートポートは、前記第1のエンドポイントデバイスのミッションバーチャルファンクションとなる前記フロントエンドインターフェイス及び前記第3のエンドポイントデバイスのミッションバーチャルファンクションとなる前記キャッシュメモリインターフェイスへアクセス可能なように設定され、

各ルートコンプレックスの前記第2のルートポートは、前記第2のエンドポイントデバイスのミッションバーチャルファンクションとなる前記バックエンドインターフェイス及び前記第3のエンドポイントデバイスのミッションバーチャルファンクションとなる前記キャッシュメモリインターフェイスへアクセス可能なように設定されていること、を特徴とするストレージ装置。

【請求項7】

請求項1記載のストレージ装置において、

前記エンドポイントデバイスは、当該ストレージ装置のデータ読み出しあるいは書き込みに関連するデータ転送のための第1のファンクションと、前記第1のファンクションに付随するバーチャルファンクションの状態遷移を制御可能な第2のファンクションと、を備えること、を特徴とするストレージ装置。

【請求項8】

請求項7記載のストレージ装置において、

前記ストレージコントローラまたはそれに接続される保守端末装置に、マルチルートPCIマネージャのプログラムを備え、

前記ルートポートからエンドポイントデバイスへのバーチャルハイアラキの構成において、

前記エンドポイントデバイスは、前記マルチルートPCIマネージャのプログラムからの前記第2のファンクションへの、データ経路障害が未発生の際のバーチャルハイアラキにおけるインバンド通信により、前記第1のファンクションのバーチャルファンクションを状態遷移させること、を特徴とするストレージ装置。

【請求項9】

請求項7記載のストレージ装置において、

前記ストレージコントローラまたはそれに接続される保守端末装置に、マルチルートPCIマネージャのプログラムを備え、

前記エンドポイントデバイスは、前記マルチルートPCIマネージャのプログラムからの前記第2のファンクションへのアウトバンド通信により、前記第1のファンクションのバーチャルファンクションを状態遷移させること、を特徴とするストレージ装置。

【請求項10】

請求項7記載のストレージ装置において、

前記ストレージコントローラに、シングルルートPCIマネージャのプログラムを備え、

、

前記ルートポートからエンドポイントデバイスへのバーチャルハイアラキの構成において、

前記エンドポイントデバイスは、自律機能により、データ経路障害が発生したバーチャ

ルハイアラキにおける前記シングルルートPCIマネージャーのプログラムと、前記第2のファンクションとの通信可否を判定し、否の場合に前記第1のファンクションのバーチャルファンクションを状態遷移させること、を特徴とするストレージ装置。

【請求項11】

請求項1記載のストレージ装置において、

前記ストレージコントローラ及びその内部ネットワークは、

前記ルートポートとエンドポイントデバイス内バーチャルファンクションとの対応関係を管理するテーブルと、

前記バーチャルファンクションと前記所定のインターフェイスとのマッピングを管理するテーブルと、

前記フェイルオーバーの際の前記バーチャルファンクションのマイグレーションを管理するテーブルと、を含む複数のテーブルを管理し、

前記ストレージコントローラ及びその内部ネットワークは、前記データ経路のフェイルオーバーを実行する処理の際、前記複数のテーブルを参照して、前記データ経路のアクセス元ルートポートを切り替える処理と、前記アクセス先エンドポイントデバイスのバーチャルファンクションと前記所定のインターフェイスとのマッピングを前記バーチャルファンクションマイグレーションによって切り替える処理とを行うこと、を特徴とするストレージ装置。

【請求項12】

ストレージコントローラ内のコンポーネントがPCI Expressで接続されたストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法であって、

前記ストレージコントローラの内部ネットワークは、全体で複数のルートポートと、前記ルートポートを有するルートコンプレックスに接続されるプロセッサと、所定のインターフェイスを備え前記複数のルートポートのそれぞれからアクセス可能な複数のバーチャルファンクションを有する複数のエンドポイントデバイスと、前記複数のルートポートと前記複数のエンドポイントデバイスとを接続してデータ経路を構成し複数のポートを備えるスイッチと、を有し、

前記複数のエンドポイントデバイスの少なくとも1つは、前記バーチャルファンクションと前記所定のインターフェイスとのマッピングを変更するバーチャルファンクションマイグレーションを実行可能であり、

前記ストレージコントローラの内部ネットワークは、すべての前記複数のルートポートが、すべての前記複数のエンドポイントデバイス内のバーチャルファンクションへアクセス可能なように、前記スイッチ及び前記エンドポイントデバイスが設定されており、

前記ストレージコントローラの内部ネットワークは、前記複数のルートポートの1つから前記スイッチを介して前記複数のエンドポイントデバイスの1つへのデータ経路の1つに障害が発生した場合に、データ経路のフェイルオーバーを実行する処理ステップとして、当該障害状態の第1のデータ経路から代替の第2のデータ経路への切り替えにおいて、当該データ経路のアクセス元ルートポートを切り替える処理ステップと、当該アクセス先エンドポイントデバイスのバーチャルファンクションと前記所定のインターフェイスとのマッピングを前記バーチャルファンクションマイグレーションによって切り替える処理ステップとを有すること、を特徴とする、ストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法。

【請求項13】

請求項12記載のストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法において、

前記ストレージコントローラの内部ネットワークに、シングルルートPCIマネージャーのプログラム、及びフェイルオーバー制御プログラムを備え、

前記ストレージコントローラの内部ネットワークまたはそれに接続される保守端末装置に、マルチルートPCIマネージャーのプログラムを備え、

前記ストレージコントローラの内部ネットワークは、前記複数のルートポートの1つか

ら前記スイッチを介して前記複数のエンドポイントデバイスの１つへのデータ経路の１つに障害が発生した場合に、データ経路のフェイルオーバーを実行する処理として、

前記マルチルートPCIマネージャーのプログラムが、前記フェイルオーバー制御プログラムへデータ経路障害発生を通知する処理ステップと、

前記フェイルオーバー制御プログラムが、前記内部ネットワーク内にあるフェイルオーバー管理テーブルを参照し、上記通知された情報に基づいて前記マルチルートPCIマネージャーのプログラムに前記バーチャルファンクションマイグレーションを要求する処理ステップと、

前記マルチルートPCIマネージャーが、上記要求に対し、前記バーチャルファンクションマイグレーションを実行し、前記内部ネットワーク内にあるバーチャルファンクションと前記所定のインターフェイスとのマッピングを管理するテーブルを更新する処理ステップと、

前記フェイルオーバー制御プログラムが、上記バーチャルファンクションマイグレーションの実行に伴い、前記フェイルオーバー管理テーブルを更新する処理ステップと、を有すること、を特徴とする、ストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法。

【請求項１４】

請求項１３記載のストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法において、

前記データ経路における障害が未発生 of データ経路に関して、前記マルチルートPCIマネージャーのプログラムが、前記バーチャルファンクションマイグレーションを実行し、前記バーチャルファンクションマッピング管理テーブルを更新する処理ステップと、

前記フェイルオーバー制御プログラムが、上記バーチャルファンクションマイグレーションの実行に伴い、前記フェイルオーバー管理テーブルを更新する処理ステップとを有すること、を特徴とする、ストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法。

【請求項１５】

請求項１２記載のストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法において、

前記エンドポイントデバイスは、当該ストレージ装置のデータ読み出しあるいは書き込みに関連するデータ転送のための第１のファンクションと、前記第１のファンクションに付随するバーチャルファンクションの状態遷移を制御可能な第２のファンクションと、を備え、

前記エンドポイントデバイスは、前記シングルルートPCIマネージャーのプログラムまたは前記マルチルートPCIマネージャーのプログラムからの前記第２のファンクションへの通信に基づき、前記第１のファンクションのバーチャルファンクションを状態遷移させる処理ステップを有すること、を特徴とする、ストレージコントローラ内部ネットワークのデータ経路フェイルオーバー方法。

【手続補正２】

【補正対象書類名】明細書

【補正対象項目名】００２０

【補正方法】変更

【補正の内容】

【００２０】

第２の特徴は、VFマイグレーション制御手段に対する、データ経路障害発生時等の対策として、SR-PCIM以外による代替手段を設ける。EPDに、その代替手段となる機能を持たせる。例えば第１の方法（インバンド通信による方法）、第２の方法（アウトバンド通信による方法）、第３の方法（自律機能による方法）がある。

【手続補正３】

【補正対象書類名】明細書

【補正対象項目名】 0 1 4 0

【補正方法】 変更

【補正の内容】

【 0 1 4 0 】

第 1 の方法（インバンド通信による方法）： V F 状態制御部 2 1 0 4 は、P F 2 1 0 3 と接続することにより、S C 2 0 内で動作するプログラム（例えば、当該 E P D 5 1 と正常に接続されるデータ経路における R P 上のプロセッサ 2 1 が実行する F P 1 1 0 ）からの指令により、V F 2 0 4（「V F 0:0,1」,「V F 1:0,1」,「V F 2:0,1」,「V F 3:0,1」）を状態遷移させることができる。

【手続補正 4】

【補正対象書類名】 明細書

【補正対象項目名】 0 1 4 2

【補正方法】 変更

【補正の内容】

【 0 1 4 2 】

第 2 の方法（アウトバンド通信による方法）： また、V F 状態制御部 2 1 0 4 は、管理ポート 2 1 0 5 とネットワーク N 0 0 を介して S V P 1 9 0 と接続することにより、P C I e ファブリック外で動作するプログラムからのアウトバンド通信による命令を受けて、V F 2 0 4（「V F 0:0,1」,「V F 1:0,1」,「V F 2:0,1」,「V F 3:0,1」）を状態遷移させることができる。なお第 2 の方法のみの構成とする場合、B F 2 1 0 2 , P F 2 1 0 3 は必要無い。

【手続補正 5】

【補正対象書類名】 明細書

【補正対象項目名】 0 1 6 3

【補正方法】 変更

【補正の内容】

【 0 1 6 3 】

< V F マイグレーション処理（ 2 ） >

図 2 3 は、第 2 の方法を用いて、E P D 5 1 において、M R - P C I M が管理ポート 2 1 0 5 とネットワーク N 0 0 を介したアウトバンド通信を用いて、ファンクション番号 1 の機能によりファンクション番号 0 の V F を状態遷移させ、V F マイグレーションを実行する処理（ 2 3 0 0 ）のフローである。

【手続補正 6】

【補正対象書類名】 明細書

【補正対象項目名】 0 1 6 5

【補正方法】 変更

【補正の内容】

【 0 1 6 5 】

S 2305 において、M R - P C I M 等（M R - P C I M、または M R - P C I M と通信可能なソフトウェア（フェイルオーバー制御プログラム等））は、アクセス障害が生じている E P D 5 1 内 V F 状態制御部 2 1 0 4 とネットワーク N 0 0 を介したアウトバンド通信を行い、V F 状態制御部 2 1 0 4 に V F（「V F h:f,s」）に対して F L R を発行させる。

【手続補正 7】

【補正対象書類名】 明細書

【補正対象項目名】 0 1 6 6

【補正方法】 変更

【補正の内容】

【 0 1 6 6 】

S 2306 において、M R - P C I M 等は、アクセス障害が生じている E P D 5 1 の V F 状態制御部 2 1 0 4 とネットワーク N 0 0 を介したアウトバンド通信を行い、V F（「V F h

:f,s」)のVF状態を、「A.M」から「I.U」へ遷移させる。この遷移では、MR VF Migration Status bitをセットし、MR - PCIMへの割り込みを起動する。以降の処理は前述同様である。

【手続補正8】

【補正対象書類名】明細書

【補正対象項目名】0167

【補正方法】変更

【補正の内容】

【0167】

上記のように、実施の形態2(第2の方法)のフェイルオーバー方法及びVFマイグレーション処理では、S2305～S2306において、MR - IOV標準仕様ではSR - PCIMが処理すべきステップを、EPD51内に設けたVF状態制御部2104と[アウトバンド](#)通信を用いて処理する点が特徴である。

【手続補正9】

【補正対象書類名】図面

【補正対象項目名】図23

【補正方法】変更

【補正の内容】

【図 23】

