(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6975095号
(P6975095)

(45)発行日　令和3年12月1日(2021.12.1)　(24)登録日　令和3年11月9日(2021.11.9)

(51)Int.Cl.		F I		
G06F 3/06	(2006.01)	G06F	3/06	301N
G06F 3/08	(2006.01)	G06F	3/08	H
G06F 13/10	(2006.01)	G06F	13/10	340A
G06N 20/00	(2019.01)	G06N	20/00	

請求項の数 19　(全 17 頁)

(21)出願番号　特願2018-88873(P2018-88873)
(22)出願日　平成30年5月2日(2018.5.2)
(65)公開番号　特開2018-198054(P2018-198054A)
(43)公開日　平成30年12月13日(2018.12.13)
審査請求日　令和3年4月26日(2021.4.26)
(31)優先権主張番号　62/510,728
(32)優先日　平成29年5月24日(2017.5.24)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　15/672,223
(32)優先日　平成29年8月8日(2017.8.8)
(33)優先権主張国・地域又は機関
米国(US)

早期審査対象出願

(73)特許権者　390019839
三星電子株式会社
Ｓａｍｓｕｎｇ　Ｅｌｅｃｔｒｏｎｉｃｓ　Ｃｏ．，Ｌｔｄ．
大韓民国京畿道水原市霊通区三星路１２９　１２９，Ｓａｍｓｕｎｇ－ｒｏ，Ｙｅｏｎｇｔｏｎｇ－ｇｕ，Ｓｕｗｏｎ－ｓｉ，Ｇｙｅｏｎｇｇｉ－ｄｏ，Ｒｅｐｕｂｌｉｃ　ｏｆ　Ｋｏｒｅａ
(74)代理人　110000051
特許業務法人共生国際特許事務所
(72)発明者　カチェア，　ラムダス　ピー．
アメリカ合衆国，　９５０１４，　カリフォルニア州，　クパチーノ，　ノルマンディー　ウェイ　７６６５

最終頁に続く

(54)【発明の名称】マシンラーニングを実行するデータストレージ及び処理システムとその動作方法

(57)【特許請求の範囲】

【請求項１】

ホストサーバ及びストレージ部を備えるデータストレージ及び処理システムであって、

前記ストレージ部は、ドライブメモリー及びドライブプロセッサを含むドライブと、

前記ホストサーバと前記ドライブメモリーとの間でデータを伝送し受信するために前記ホストサーバを前記ドライブに連結させるように構成された外部スイッチと、

グラフィック処理装置と、を含み、

前記ドライブプロセッサは、処理命令及びデータを前記ドライブメモリーから前記グラフィック処理装置に伝送するように構成され、

前記グラフィック処理装置は、前記処理命令に従ってデータを処理して結果データを生成するように構成され、

前記ドライブは、前記グラフィック処理装置が変換を遂行可能なことを示すテーブルを含み、

前記ドライブプロセッサは、前記ドライブメモリーに格納されたデータチャンク及び前記データチャンクに適用される前記変換を識別し、前記データチャンクを前記グラフィック処理装置に伝送するように構成されることを特徴とするデータストレージ及び処理システム。

【請求項２】

前記グラフィック処理装置は、Ｕ．２コネクターを含み、前記Ｕ．２コネクターを介して前記ドライブに連結され、前記Ｕ．２コネクターを介して前記処理命令及びデータを受

信することを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項３】
前記グラフィック処理装置は、前記結果データを前記ドライブプロセッサに伝送するように構成され、
前記ドライブプロセッサは、前記ドライブメモリーに前記結果データを格納するように構成されることを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項４】
前記グラフィック処理装置は、前記外部スイッチに連結され、前記外部スイッチを利用して前記結果データを前記ホストサーバに伝送するように構成されることを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項５】
前記グラフィック処理装置は、前記処理命令に従ってデータの処理を完了した後、状態メッセージを前記ドライブプロセッサに伝送するように構成されることを特徴とする請求項４に記載のデータストレージ及び処理システム。
【請求項６】
前記ドライブプロセッサは、前記テーブルから前記グラフィック処理装置のアドレスを取得し、前記データチャンクを前記グラフィック処理装置のアドレスに伝送するように構成されることを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項７】
前記ストレージ部は、前記グラフィック処理装置の能力を判別して前記グラフィック処理装置の能力に基づいて前記テーブルをアップデートするように構成されたベースボード管理制御器、を更に含むことを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項８】
前記ストレージ部は、各ドライブが別途のテーブルを有する複数のドライブと、複数のグラフィック処理装置と、を含み、
前記ベースボード管理制御器は、前記複数のドライブの各ドライブの各テーブルをアップデートするように構成されることを特徴とする請求項７に記載のデータストレージ及び処理システム。
【請求項９】
前記ベースボード管理制御器及び前記グラフィック処理装置は、ＮＶＭｅ－ＭＩ（Ｎｏｎ Ｖｏｌａｔｉｌｅ Ｍｅｍｏｒｙ ｅｘｐｒｅｓｓ－Ｍａｎａｇｅｍｅｎｔ Ｉｎｔｅｒｆａｃｅ）のプロトコルを用いて通信するように構成され、
前記ベースボード管理制御器は、前記ＮＶＭｅ－ＭＩの識別コマンドを利用することで、前記グラフィック処理装置の能力を判別することを特徴とする請求項７に記載のデータストレージ及び処理システム。
【請求項１０】
前記ストレージ部は、複数のグラフィック処理装置を含み、
前記ベースボード管理制御器は、前記複数のグラフィック処理装置の各グラフィック処理装置に対する負荷を判別し、前記複数のグラフィック処理装置の各グラフィック処理装置に対する負荷に基づいて前記テーブルをアップデートするように構成されることを特徴とする請求項７に記載のデータストレージ及び処理システム。
【請求項１１】
前記外部スイッチは、イーサネット（登録商標）スイッチであり、
前記ドライブは、イーサネット（登録商標）ソリッドステートドライブであることを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項１２】
前記ホストサーバは、遠隔直接アクセスストレージプロトコルを用いて前記ストレージ部と通信することを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項１３】

前記ストレージ部は、シャシー及びベースボード管理制御器を更に含み、
前記グラフィック処理装置は、現場交換可能な装置であり、
前記ベースボード管理制御器は，前記グラフィック処理装置が前記シャシーに挿入されることに対応してプラグイン（ｐｌｕｇ－ｉｎ）イベントを感知するように構成されることを特徴とする請求項１に記載のデータストレージ及び処理システム。
【請求項１４】
ドライブプロセッサ及びドライブメモリーを含むドライブとグラフィック処理装置とを備えるデータストレージ及び処理システムの動作方法であって、
前記ドライブプロセッサが、ホストサーバからデータを受信するステップと、
前記データを前記ドライブメモリーに格納するステップと、
トリガーを感知するステップと、
前記トリガーに応答し、前記ドライブメモリーに格納されたデータチャンク及び前記データチャンクに適用する変換を識別するアルゴリズムを実行するステップと、
前記変換を遂行する装置に対応するアドレスを識別するために能力テーブルを検索するステップと、
前記データチャンク及び前記データチャンクを処理する処理命令を前記グラフィック処理装置のアドレスに伝送するステップと、を含むことを特徴とする方法。
【請求項１５】
前記グラフィック処理装置が、前記データチャンク及び前記処理命令を受信するステップと、
前記処理命令に従って前記データチャンクを処理して結果データを生成するステップと、を更に含むことを特徴とする請求項１４に記載の方法。
【請求項１６】
前記グラフィック処理装置が、前記結果データを前記ドライブプロセッサに伝送するステップと、
前記ドライブプロセッサが、前記結果データを前記ドライブメモリーに格納するステップと、を更に含むことを特徴とする請求項１５に記載の方法。
【請求項１７】
ベースボード管理制御器が、前記グラフィック処理装置の能力を感知するステップと、
前記グラフィック処理装置の能力に基づいて前記能力テーブルをアップデートするステップと、を更に含むことを特徴とする請求項１４に記載の方法。
【請求項１８】
前記データストレージ及び処理システムは、各ドライブが能力テーブルを含む複数のドライブを含み、
前記ベースボード管理制御器が、前記グラフィック処理装置の能力に基づいて前記複数のドライブの各ドライブの各能力テーブルをアップデートするステップを含むことを特徴とする請求項１７に記載の方法。
【請求項１９】
ドライブプロセッサ及びドライブメモリーを含むドライブとグラフィック処理装置とを備えるデータストレージ及び処理システムであって、
ホストサーバからデータを受信する手段と、
前記データを前記ドライブメモリーに格納する手段と、
トリガーを感知する手段と、
前記トリガーに応答し、前記ドライブメモリーに格納されたデータチャンク及び前記データチャンクに適用する変換を識別するアルゴリズムを実行する手段と、
前記変換を遂行する装置に対応するアドレスを識別するために能力テーブルを検索する手段と、
前記データチャンク及び前記データチャンクを処理する処理命令を前記グラフィック処理装置のアドレスに伝送する手段と、
前記処理命令に従って前記データチャンクを処理して結果データを生成し、前記結果デ

ータを前記ホストサーバに伝送する手段と、
現場交換可能なグラフィック処理装置のプラグインイベントを感知し、前記現場交換可能なグラフィック処理装置の能力を判別し、前記現場交換可能なグラフィック処理装置の能力に基づき前記プラグインイベントの感知に応答して前記能力テーブルをアップデートする手段と、を含むことを特徴とするデータストレージ及び処理システム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、マシンラーニングのためのシステムに関し、より詳しくは、ＳＳＤフォームファクターで内蔵されたＧＰＵを有するＮＶＭｅ－ｏＦｅＳＳＤシャシーに実装されたストレージシステムにおいて、複数のグラフィック処理装置を利用してマシンラーニングアルゴリズムを実行するデータストレージ及び処理システムとその動作方法に関する。
【背景技術】
【０００２】
収集及び格納されるデータ量が増加している。モノのインターネット（Ｉｎｔｅｒｎｅｔ　ｏｆ　Ｔｈｉｎｇｓ）からソーシャルネットワーク、デジタルヘルスの領域に至るまで数多くのアプリケーションは膨大な量のデータを生成する。このようなデータはデータセンターのようなデータシステムでホスト（ｈｏｓｔ）される。データは、しばしば格納、処理、及び分析が要求される。マシンラーニングアルゴリズムのようなアルゴリズムはデータから特徴又は有用な情報を抽出するためにビッグデータ（ｂｉｇ　ｄａｔａ）セットに適用される。このデータを、迅速、効率的、経済的、弾力的に、格納、処理、及び分析するための技法及びアーキテクチャが必要である。
【０００３】
背景技術で開示する上記の情報は、単に本発明の背景に対する理解を増進させるためのものであり、従って通常の技術者にとって既に知られた先行技術ではない情報を含む。
【先行技術文献】
【特許文献】
【０００４】
特許文献１：米国特許第８８１９３３５号明細書
特許文献２：米国特許第９３１７２０４号明細書
特許文献３：米国特許第９４３０４１２号明細書
特許文献４：米国特許第９４８３４３１号明細書
特許文献５：米国特許出願公開第２０１７／０２１００７６号明細書
特許文献６：米国特許出願公開第２０１７／００１０９９５号明細書
特許文献７：米国特許出願公開第２０１７／００１９３１２号明細書
特許文献８：米国特許出願公開第２０１７／００６０４４２号明細書
【発明の概要】
【発明が解決しようとする課題】
【０００５】
本発明は、上記従来技術に鑑みてなされたものであって、本発明の目的は、ストレージシステムでマシンラーニングアルゴリズムを実行するデータストレージ及び処理システムとその動作方法を提供することにある。
【課題を解決するための手段】
【０００６】
上記目的を達成するためになされた本発明の一態様によるとデータストレージ及び処理システムは、ホストサーバ及びストレージ部を備え、前記ストレージ部は、ドライブメモリー及びドライブプロセッサを含むドライブと、前記ホストサーバと前記ドライブのメモリーとの間でデータを伝送し受信するために前記ホストサーバを前記ドライブに連結させる外部スイッチと、グラフィック処理装置と、を含み、前記ドライブプロセッサは、処理命令及びデータを前記ドライブメモリーから前記グラフィック処理装置に伝送し、前記グ

ラフィック処理装置は、前記処理命令に従ってデータを処理して結果データを生成する。
【０００７】
前記グラフィック処理装置は、Ｕ．２コネクターを含み、前記Ｕ．２コネクターを介して前記ドライブに連結され、前記Ｕ．２コネクターを介して前記処理命令及びデータを受信し得る。
前記グラフィック処理装置は、前記結果データを前記ドライブプロセッサに伝送し、前記ドライブプロセッサは、前記ドライブメモリーに前記結果データを格納し得る。
前記グラフィック処理装置は、前記外部スイッチに連結され、前記外部スイッチを利用して前記結果データを前記ホストサーバに伝送し得る。
前記グラフィック処理装置は、前記処理命令に従ってデータの処理を完了した後、状態メッセージを前記ドライブプロセッサに伝送し得る。
前記ホストサーバは、トリガー命令を前記ドライブプロセッサに伝送し、前記ドライブプロセッサは、前記トリガー命令に応答し、実行時に前記処理命令及びデータを前記グラフィック処理装置に伝送する過程を含むデータ処理アルゴリズムを実行し得る。
前記ドライブは、能力テーブルを含み、前記ドライブプロセッサは、前記ドライブメモリーに格納されたデータチャンク及び前記データチャンクに適用される変換を識別し、前記グラフィック処理装置が前記変換を遂行可能なことを示す前記能力テーブルから前記グラフィック処理装置のアドレスを検索し、前記データチャンクを前記グラフィック処理装置のアドレスに伝送し得る。
前記ストレージ部は、前記グラフィック処理装置の能力を判別し、前記グラフィック処理装置の能力に基づいて前記能力テーブルをアップデートするベースボード管理制御器（ｂａｓｅｂｏａｒｄ ｍａｎａｇｅｍｅｎｔ ｃｏｎｔｒｏｌｌｅｒ）を、更に含み得る。
前記ストレージ部は、各ドライブが別途の能力テーブルを有する複数のドライブと、複数のグラフィック処理装置と、を含み、前記ベースボード管理制御器は、前記複数のドライブの各ドライブの各能力テーブルをアップデートし得る。
前記ベースボード管理制御器及び前記グラフィック処理装置は、ＮＶＭｅ－ＭＩ（Ｎｏｎ Ｖｏｌａｔｉｌｅ Ｍｅｍｏｒｙ ｅｘｐｒｅｓｓ－Ｍａｎａｇｅｍｅｎｔ Ｉｎｔｅｒｆａｃｅ）のプロトコルを用いて通信し、前記ベースボード管理制御器は、前記ＮＶＭｅ－ＭＩの識別（Ｉｄｅｎｔｉｆｙ）コマンドを利用することで、前記グラフィック処理装置の能力を判別し得る。
前記ストレージ部は、複数のグラフィック処理装置を含み、前記ベースボード管理制御器は、前記複数のグラフィック処理装置の各グラフィック処理装置に対する負荷を判別し、前記複数のグラフィック処理装置の各グラフィック処理装置に対する負荷に基づいて前記能力テーブルをアップデートし得る。
前記外部スイッチは、イーサネット（登録商標）スイッチであり、前記ドライブは、イーサネット（登録商標）ソリッドステートドライブ（Ｅｔｈｅｒｎｅｔ（登録商標） ｓｏｌｉｄ ｓｔａｔｅ ｄｒｉｖｅ）であり得る。
前記ホストサーバは、遠隔直接アクセスストレージ（ｒｅｍｏｔｅ ｄｉｒｅｃｔ ａｃｃｅｓｓ ｓｔｏｒａｇｅ）プロトコルを用いて前記ストレージ部と通信し得る。
前記ストレージ部は、シャシー（ｃｈａｓｓｉｓ）及びベースボード管理制御器を更に含み、前記グラフィック処理装置は、現場交換可能な装置であり、前記ベースボード管理制御器は、前記グラフィック処理装置が前記シャシーに挿入されることに対応してプラグイン（ｐｌｕｇ－ｉｎ）イベントを感知し得る。
【０００８】
上記目的を達成するためになされた本発明の一態様によるドライブプロセッサ及びドライブメモリーを含むドライブとグラフィック処理装置とを備えるデータストレージ及び処理システムの動作方法は、前記ドライブプロセッサが、ホストサーバからデータを受信するステップと、前記データを前記ドライブメモリーに格納するステップと、トリガーを感知するステップと、前記トリガーに応答し、前記ドライブメモリーに格納されたデータチ

ャンク及び前記データチャンクに適用する変換を識別するアルゴリズムを実行するステップと、前記変換を遂行する装置に対応するアドレスを識別するために能力テーブルを検索するステップと、前記データチャンク及び前記データチャンクを処理する処理命令を前記グラフィック処理装置のアドレスに伝送するステップと、を含むことを特徴とする。

【０００９】

前記方法は、前記グラフィック処理装置が、前記データチャンク及び前記処理命令を受信するステップと、前記処理命令に従って前記データチャンクを処理して結果データを生成するステップと、を更に含み得る。

前記方法は、前記グラフィック処理装置が、前記結果データを前記ドライブプロセッサに伝送するステップと、前記ドライブプロセッサが、前記結果データを前記ドライブメモリーに格納するステップと、を更に含み得る。

前記データストレージ及び処理システムは、ベースボード管理制御器を更に備え、前記方法は、前記ベースボード管理制御器が、前記グラフィック処理装置の能力を感知するステップと、前記グラフィック処理装置の能力に基づいて前記能力テーブルをアップデートするステップと、を更に含み得る。

前記データストレージ及び処理システムは、各ドライブが能力テーブルを含む複数のドライブを含み、前記方法は、前記ベースボード管理制御器が、前記グラフィック装置の能力に基づいて前記複数のドライブの各ドライブの各能力テーブルをアップデートするステップと、を含み得る。

【００１０】

上記目的を達成するためになされた本発明の他の態様によるデータストレージ及び処理システムは、ドライブプロセッサ及びドライブメモリーを含むドライブとグラフィック処理装置とを備え、ホストサーバからデータを受信する手段と、前記データを前記ドライブメモリーに格納する手段と、トリガーを感知する手段と、前記トリガーに応答し、前記ドライブメモリーに格納されたデータチャンク及び前記データチャンクに適用する変換を識別するアルゴリズムを実行する手段と、前記変換を遂行する装置に対応するアドレスを識別するために能力テーブルを検索する手段と、前記データチャンク及び前記データチャンクを処理する処理命令を前記グラフィック処理装置のアドレスに伝送する手段と、前記処理命令に従って前記データチャンクを処理して結果データを生成し、前記結果データを前記ホストサーバに伝送する手段と、現場交換可能なグラフィック処理装置のプラグインイベントを感知し、前記現場交換可能なグラフィック処理装置の能力を判別し、前記現場交換可能なグラフィック処理装置の能力に基づき前記プラグインイベントの感知に応答して前記能力テーブルをアップデートする手段と、を含むことを特徴とする。

【発明の効果】

【００１１】

本発明によると、データストレージ及び処理システムは、複数のグラフィック処理装置を利用してマシンラーニングアルゴリズムを効率的に実行することができ、複数のグラフィック処理装置の負荷の均衡をとることができる。

【図面の簡単な説明】

【００１２】

【図１】図１は、関連技術によるストレージシステムのブロック図である。

【図２】図２は、本発明の一実施形態によるストレージシステムのブロック図である。

【図３】図３は、本発明の一実施形態によるグラフィック処理装置のブロック図である。

【図４】図４は、本発明の一実施形態によるグラフィック処理装置と通信するドライブに対する手順を示すフローチャートである。

【図５】図５は、本発明の一実施形態による能力テーブルの一例を示す図である。

【図６】図６は、本発明の一実施形態によるグラフィック処理装置を管理するベースボード管理制御器に対する手順を示すフローチャートである。

【発明を実施するための形態】

【００１３】
以下、本発明を実施するための形態の具体例を、図面を参照しながら詳細に説明する。
【００１４】
次の詳細な説明では、本発明の特定の実施形態のみを、図面を通じて示して説明する。当業者（通常の技術者）が認識するように、本発明は多様な形態で具現され、本明細書に記載する実施形態に制限されるものと解釈すべきではない。各実施形態に有る特徴又は態様の説明は、典型的に他の実施形態で他の類似した特徴及び態様に利用されるものとして考慮されなければならない。本明細書で同一の参照番号は同一の要素を示す。
【００１５】
図１は、関連技術によるストレージシステムのブロック図である。ホストサーバ１１０は、ネットワーク１２０を通じて一つ以上のストレージ部１３０に連結される。ホストサーバ１１０は、一つ以上のアプリケーション１１２とオペレーティングシステム（ＯＳ）及びファイルシステム１１４を実行する。ホストサーバ１１０は、一つ以上のストレージ部と相互作用するために利用される遠隔直接アクセスストレージ（ｒＤＡＳ：ｒｅｍｏｔｅ　Ｄｉｒｅｃｔ　Ａｃｃｅｓｓ　Ｓｔｏｒａｇｅ）ドライバーを含む。
【００１６】
ストレージ部１３０は、シャシー（ｃｈａｓｓｉｓ）１３１、イーサネット（登録商標）（Ｅｔｈｅｒｎｅｔ（登録商標）：以下、「イーサネット」の（登録商標）の記載は省略する。）スイッチ１３２、ベースボード管理制御器（ＢＭＣ：ｂａｓｅｂｏａｒｄ　ｍａｎａｇｅｍｅｎｔ　ｃｏｎｔｒｏｌｌｅｒ）１３４、ＰＣＩｅスイッチ１３６、及び複数のイーサネットソリッドステートドライブ（ｅＳＳＤ：ｅｔｈｅｒｎｅｔ　Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅｓ）（１４２Ａ～１４２Ｃ）を含む。ＰＣＩｅスイッチ１３６はシャシーミッドプレーン（ｍｉｄｐｌａｎｅ）１４０を通じてＢＭＣ１３４をｅＳＳＤ（１４２Ａ～１４２Ｃ）に連結する。ＢＭＣ１３４はｅＳＳＤ（１４２Ａ～１４２Ｃ）を管理するためにＰＣＩｅスイッチ１３６を利用する。イーサネットスイッチ１３２はｅＳＳＤ（１４２Ａ～１４２Ｃ）をネットワーク１２０に連結する。ホストサーバ１１０は、イーサネットスイッチ１３２を通じてデータをｅＳＳＤ（１４２Ａ～１４２Ｃ）に送信し、ｅＳＳＤ（１４２Ａ～１４２Ｃ）からデータを受信するためにｒＤＡＳドライバーを利用する。
【００１７】
マシンラーニングのようなプロセスは、大量のデータが処理されるように要求する。計算リソース（例えば、プロセッサ）はデータに対してアルゴリズム（例えば、マシンラーニングアルゴリズム）を実行する。ストレージ部１３０に格納されたデータにこのような処理を遂行する際、ホストサーバ１１０はストレージ部１３０から処理されるデータを要請しなければならない。ストレージ部１３０はネットワーク１２０を通じてデータをホストサーバ１１０に伝送しなければならない。ホストサーバ１１０は、例えばマシンラーニングアルゴルズムを実行することにより、ホストサーバ１１０に位置するＣＰＵ又ＧＰＵを利用してデータを処理し、格納のためにストレージ部１３０に処理結果を再び伝送する必要がある。
【００１８】
処理のためにストレージ部１３０からホストサーバ１１０にデータを移動することは、相当な電気的エネルギーを消費し、ネットワーク１２０を通じてデータ及び結果を前後に伝播することに関連する処理遅延を伴う。アーキテクチャは、ネットワーク１２０を通じて、データ及び結果を伝送するための適切な帯域幅とデータを処理するためのホストサーバ１１０における処理リソース及びシステムメモリーとを要求するため、アーキテクチャの費用が高くなる。更に、ホストサーバ１１０でデータ処理における処理リソースの速度はデータセットの処理にボトルネック現象を引き起こし、処理に用いられる処理リソースは追加、除去、又は交換が容易でない。
【００１９】
図２は、本発明の一実施形態によるストレージシステムのブロックである。図２を参照

すると、ホストサーバ２１０はネットワーク２２０を通じて一つ以上のストレージ部２３０に連結される。ホストサーバ２１０はアプリケーション２１２とオペレーティングシステム（ＯＳ）及びファイルシステム２１４を実行する。また、ホストサーバ２１０は一つ以上のストレージ部２３０と相互に作用するために利用されるストレージドライバー２１６を含む。

【００２０】

ストレージ部２３０は、シャシー２３１、外部スイッチ２３２（例えば、イーサネットスイッチ）、ベースボード管理制御器（ＢＭＣ）２３４、内部スイッチ２３６（例えば、ＰＣＩｅスイッチ）、一つ以上のドライブ（２４２Ａ～２４２Ｂ、単に２４２とも称する。）、及び一つ以上のＵ．２グラフィク処理装置（２５０Ａ～２５０Ｂ、単に２５０とも称する。）（Ｕ．２　ＧＰＵ）を含む。Ｕ．２　ＧＰＵの用語を本明細書に亘って使用するが、ＧＰＵはＵ．２連結以外の連結で作動し、このような連結は本発明の範囲に含まれる。この用語は明確性のため単純に使用される。内部スイッチ２３６は、ＢＭＣ２３４、ドライブ（２４２Ａ～２４２Ｂ）、及びＵ．２　ＧＰＵ（２５０Ａ～２５０Ｂ）を、シャシーミッドプレーン２４０を通じて連結する。外部スイッチ２３２は、ドライブ（２４２Ａ～２４２Ｂ）、Ｕ．２　ＧＰＵ（２５０Ａ～２５０Ｂ）、ＢＭＣ２３４、及びネットワーク２２０を連結する。本実施形態によると、シャシー２３１はドライブ及び／又はＵ．２　ＧＰＵを受容するための複数のスロットを含む。

【００２１】

例えば、いくつかの実施形態において、ストレージドライバー２１６はＮＶＭｅ－オーバーファブリック（ＮＶＭｅ－ｏＦ：ＮＶＭｅ－ｏｖｅｒＦａｂｒｉｃｓ）ドライバーのような遠隔直接アクセスストレージ（ｒＤＡＳ）ドライバーであり、ネットワーク２２０はイーサネットネットワークであり、外部スイッチ２３２はイーサネットスイッチであり、ドライブはイーサネットソリッドステートドライブ（ｅＳＳＤ）である。ホストサーバ２１０は、イーサネットネットワークを通じて、一つ以上のストレージ部２３０内のｅＳＳＤとデータ通信するためにｒＤＡＳドライバー利用する。いくつかの実施形態において、内部スイッチ２３６はＰＣＩｅスイッチである。

【００２２】

図３は、本発明の一実施形態によるグラフィック処理装置（Ｕ．２　ＧＰＵ３５０）のブロック図である。いくつかの実施形態において、図２のＵ．２　ＧＰＵ（２５０Ａ～２５０Ｂ）は図３のＵ．２　ＧＰＵ３５０で具現される。図３を参照すると、Ｕ．２　ＧＰＵ３５０は、プロセッサ３６０、ＤＲＡＭ３６２、不揮発性メモリー（ＮＶＭ）３６４、電源供給器／電源調節器３６６、及びコネクター３５２を含む。コネクター３５２はプロセッサ３６０を連結するためのインターフェースを提供する。コネクター３５２はＰＣＩｅスイッチのような内部スイッチ及びイーサネットスイッチのような外部スイッチとインターフェースするためのインターフェースを提供する。例えば、コネクターは、イーサネットインターフェース３５４、ＰＣＩｅインターフェース３５６、及びシステム管理バス（ＳＭＢｕｓ）インターフェース３５８を提供する。一実施形態において、コネクター３５２はＵ．２コネクター／ＳＦＦ－８６３９コネクターである。

【００２３】

Ｕ．２　ＧＰＵ３５０は、ドライブ（例えば、ｅＳＳＤ）と並んでストレージ部のシャシーに連結されてストレージ部のＢＭＣと通信するように構成される。いくつかの実施形態において、コネクター３５２は、ｅＳＳＤのようなドライブと同様に、シャシーで同一のスロットに連結されるように構成される。いくつか実施形態において、Ｕ．２　ＧＰＵ３５０は、シャシーに連結された場合、イーサネットスイッチのような外部スイッチと通信する。いくつかの実施形態において、Ｕ．２　ＧＰＵ３５０は、シャシーに挿入されるか又はシャシーから除去され、ストレージ部の動作に対して自動的に収容又は除去される現場交換装置として具現される。例示は、以下の図６で説明される。この方法において、Ｕ．２　ＧＰＵ３５０の形態の処理リソースは、効率的に、ストレージ部に追加されるか、ストレージ部から除去されるか、又はストレージ部内で交換される。また、Ｕ．２　Ｇ

ＰＵ３５０は、与えられたスロットで利用可能な最大電力（コネクターが支援する最大電力）の提供を受けて、処理動作のための電力量を利用する。反面、スロット内のドライブは格納動作のために同一の電力量を利用する。例えば、一実施形態において、コネクターはＵ．２コネクターであり、Ｕ．２　ＧＰＵには２５Ｗの電力が提供される。

【００２４】

図２を再び参照すると、ドライブ（２４２Ａ～２４２Ｂ）はプロセッサ２４４及びフラッシュメモリー又は他の不揮発性メモリーのようなメモリー２４６を含む。いくつかの実施形態において、ドライブ（２４２Ａ～２４２Ｂ）はプロセッサ２４４によって利用されるダイナミックランダムアクセスメモリー（ＤＲＡＭ）を含む。一つ以上のドライブ（２４２Ａ～２４２Ｂ）において、プロセッサ２４４は、対応するドライブのメモリー２４６に格納されたデータに対し、マシンラーニングアルゴリズムのようなデータ処理アルゴリズムを実施する命令を実行する。このような命令を実行する過程として、プロセッサ２４４は、データを処理する命令だけでなく、メモリー２４６からのデータをＵ．２　ＧＰＵ２５０に伝送する。いくつかの実施形態において、データ及び命令は、内部スイッチ２３６を経由してシャシーミッドプレーン２４０を通じ、ドライブ２４２とＵ．２　ＧＰＵ２５０との間に伝送される。いくつかの実施形態において、データ及び命令は外部スイッチ２３２を通じてドライブ２４２とＵ．２　ＧＰＵ２５０との間に伝送される。ドライブ（２４２Ａ～２４２Ｂ）及びＵ．２　ＧＰＵ（２５０Ａ～２５０Ｂ）は両方共にＵ．２コネクターのようなコネクターを利用する。

【００２５】

図４は、本発明の一実施形態によるグラフィック処理装置（Ｕ．２　ＧＰＵ２５０）と通信するドライブ２４２（即ち、図２のドライブ２４２）に対する手順を示すフローチャートである。図４の手順はドライブ２４２に内蔵されたプロセッサ２４４により実行される。６１０ステップで、プロセッサ２４４はトリガー（ｔｒｉｇｇｅｒ）を感知する。トリガーは、プロセッサ２４４がマシンラーニングアルゴリズムのようなアルゴリズムによりドライブ２４２のメモリー２４６に含まれるデータを処理しなければならないことを示す。例えば、いくつかの実施形態において、トリガーは特定の時間の経路である。即ち、アルゴリズムはバックグラウンドで周期的に自動的に実行されるようにトリガーされる。いくつかの実施形態において、トリガーはホストサーバ２１０又は他の外部のソースから受信された命令又はクエリー（ｑｕｅｒｙ）である。いくつかの実施形態において、トリガーは、ドライブ２４２により受信されるデータの新しいブロック、又はホスト２１０から受信される他のデータの作業（例えば、リード（ｒｅａｄ）、ライト（ｗｒｉｔｅ）、削除又、はアップデート）である。いくつかの実施形態において、トリガーは、追加分析、処理、又は他の類型の処理を必要とするマシンラーニングアルゴリズムの結果である。いくつかの実施形態において、トリガー、は臨界値を超過して格納されたデータ量のような、ドライブのいくつかの内部状態又は動作により生成される。いくつかの実施形態において、トリガーは他のドライブ２４２から受信され、ドライブ２４２は他のドライブ２４２にトリガーを伝達する。

【００２６】

いくつかの実施形態において、アルゴリズムはドライブ２４２のメモリー２４６又はその外の位置（例えば、別途のＤＲＡＭ）に格納され、トリガーは単純に格納されたアルゴリズムが実行されるべきであることを示す。いくつかの実施形態において、トリガーは実行されなければならないアルゴリズムを含む。いくつかの実施形態において、ドライブ２４２は格納された複数のアルゴリズムを有し、トリガーは実行するアルゴリズムを識別するタグ（ｔａｇ）を含む。６２０ステップで、プロセッサ２４４はアルゴリズムの実行を開始する。いくつかの実施形態において、アルゴリズムはマシンラーニングアルゴリズムである。６２２ステップで、プロセッサ２４４は、例えばアルゴルズムにおける命令又は現在のプロセッサ２４４の利用に基づいて、メモリー２４６上のデータの一つ以上のデータチャンクで一つ以上の変換、関数、又は他の形態の処理が遂行されなければならないことを識別する。

【００２７】
６２４ステップで、プロセッサ２４２は、必要な変換、関数、又はその他の処理作業を遂行するＵ．２　ＧＰＵ２５０のアドレスを識別するための能力テーブル（ＣＡＰ　ｔａｂｌｅ：ｃａｐａｂｉｌｉｔｙ　ｔａｂｌｅ）を調査する。いつかの実施形態において、能力テーブルはドライブ２４２に格納される。図５は、本発明の一実施形態による能力テーブル５００の一例を示す図である。能力テーブル５００は、プロセッサ２４２がＵ．２　ＧＰＵ２５０を利用して遂行される変換、関数、又はその他の処理作業に対応する一連の変換１～変換ｎに対する項目を含む。各項目はプロセッサ２４２が与えられた変換を遂行するように利用するＵ．２　ＧＰＵ２５０のアドレスで満たされる。例えば、能力テーブル５００を格納するドライブに対するプロセッサは、変換１を遂行するためにスロット－３内のＵ．２　ＧＰＵ又はスロット－４内のＵ．２ＧＰＵを利用する。プロセッサ２４２は、アルゴリズムにより識別された変換を遂行するため、能力テーブルから識別されたＵ．２　ＧＰＵを選択する。
【００２８】
６３０ステップで、プロセッサ２４４はＵ．２　ＧＰＵ２５０により処理されるデータチャンクを検索する。データチャンクはメモリー２４６から取り出されるか、又はドライブ２４２のリード／ライトのキャッシュバッファーから取り出される。６３２ステップで、プロセッサ２４４は、例えばメッセージのヘッダーに命令を置き、本文にデータチャンクを置くことで、データチャンク及びデータチャンクを処理するための命令を含むメッセージを生成する。６３４ステップで、プロセッサ２４４は、例えばＰＣＩｅプロトコルを利用する内部スイッチ２３６を通じて通信することで、Ｕ．２　ＧＰＵ２５０にメッセージを伝達する。
【００２９】
６２８ステップで、プロセッサ２４４は、Ｕ．２　ＧＰＵ２５０により処理される全てのデータチャンクがＵ．２　ＧＰＵに伝送されたか否かを判別する。全てのデータチャンクが伝送されていない場合、プロセッサ２４４は、残っているデータチャンクに対して６３０、６３２、及び６３４ステップを反復する。プロセッサ２４４が、処理される全てのデータチャンクがＵ．２　ＧＰＵ２５０に伝送されたと判別した場合、手順は６３６ステップに進行する。
【００３０】
ドライブ２４２のプロセッサ２４４からメッセージを受信すると、直ちにＵ．２　ＧＰＵ２５０は、結果データを生成するため、メッセージから受信されたデータチャンクに対して、メッセージで識別される変換を遂行する。Ｕ．２　ＧＰＵ２５０が結果データを生成すると、Ｕ．２　ＧＰＵ２５０はプロセッサ２４４に変換が完了したことを示す処理応答のメッセージを伝送する。いくつかの実施形態において、処理応答のメッセージは結果データを含み、プロセッサ２４４はドライブ２４２のメモリー２４６に結果データを格納する。いくつかの実施形態において、Ｕ．２　ＧＰＵ２５０は追加的に又は代替的に結果データを、Ｕ．２　ＧＰＵ２５０を含むストレージ部１３０内の他の位置又はストレージ部１３０の外部に伝送する。例えば、Ｕ．２　ＧＰＵ２５０は追加的に又は代替的に結果データをホストサーバ２１０に伝送するか、或いはプロセッサ２４４がＵ．２　ＧＰＵ２５０に対するメッセージにアドレスを含ませ、Ｕ．２　ＧＰＵ２５０は結果データを指定されたアドレスに伝達する。
【００３１】
６３６ステップで、プロセッサ２４４は６３４ステップで伝送されたメッセージを受信した各Ｕ．２　ＧＰＵ２５０から処理応答のメッセージを受信したか否かを判別する。全ての処理応答のメッセージを受信していない場合、６３８ステップで、プロセッサ２４は、残っている処理応答のメッセージを受信するために待機する。全ての処理応答のメッセージが受信された場合、手順は６２０ステップに戻り、プロセッサ２４４はアルゴリズムの実行を継続する。或いは、例えばアルゴリズムが完了した場合、手順は６１０ステップに戻り、プロセッサ２４４は続けて進行するために他のトリガーが感知されるのを待つ。

【００３２】
図６は、本発明の一実施形態によるグラフィック処理装置（Ｕ．２　ＧＰＵ２５０Ａ～２５０Ｂ）を管理するベースボード管理制御器２３４（ＢＭＣ）に対する手順を示すフローチャートである。５１０ステップで、ＢＭＣ２３４は、パワーオン（ｐｏｗｅｒ－ｏｎ）、プラグイン（ｐｌｕｇ－ｉｎ）、又はプラグアウト（ｐｌｕｇ－ｏｕｔ）のイベントがあったことを感知する。例えば、ＢＭＣ２３４は、シャシー２３１のポート（ミッドプレーンスロット）で“現在の”ピン（ｐｉｎ）をモニターし、新しいＵ．２　ＧＰＵ２５０が現在のピンに連結（ミッドプレーンスロットに挿入）された場合に、プラグインのイベントが発生したことを感知するか、又はＵ．２　ＧＰＵ２５０が現在のピン（ミッドプレーンスロット）から除去された場合に、プラグアウトのイベントが発生したことを感知する。一実施形態において、ＢＭＣ２３４はＮＶＭｅ管理インターフェース（ＮＶＭｅ－ＭＩ：ＮＶＭｅ　Ｍａｎａｇｅｍｅｎｔ　Ｉｎｔｅｒｆａｃｅ）プロトコルを用いてＵ．２　ＧＰＵ２５０と通信するためにＰＣＩｅ又はＳＭＢｕｓインターフェースを利用する。

【００３３】
５２０ステップで、ＢＭＣ２３４がイベントをパワーオンのイベントと判別した場合、ＢＭＣ２３４はＢＭＣ２３４に連結されたＵ．２　ＧＰＵ２５０の能力を判別する。例えば、図６に示すように、５２４ステップで、ＢＭＣ２３４はＢＭＣ２３４に連結されたＵ．２　ＧＰＵ２５０に対する必須製品のデータ（ＶＰＤ：Ｖｉｔａｌ　Ｐｒｏｄｕｃｔ　Ｄａｔａ）をリード（ｒｅａｄ）する。５２６ステップで、ＢＭＣ２３４は、Ｕ．２　ＧＰＵ２５０の能力に連関する情報を集めるために、追加的に又は代替的にＮＶＭｅ“識別（Ｉｄｅｎｔｉｆｙ）”コマンドを利用する。５２２ステップで、ＢＭＣ２３４がＵ．２　ＧＰＵ２５０の能力を判別するために、スキャンされないＢＭＣ２３４に連結された一つ以上のＵ．２　ＧＰＵ２５０が有ると判別した場合、ＢＭＣ２３４は、残っているＵ．２　ＧＰＵ又はＵ．２　ＧＰＵに対して５２４ステップ及び５２６ステップを反復する。ＢＭＣ２３４がＢＭＣ２３４に連結された各Ｕ．２　ＧＰＵ２５０がその能力を判別するためにスキャンされたと判別した場合、手順は５３０ステップに進行する。

【００３４】
５３０ステップで、ＢＭＣ２３４がＵ．２　ＧＰＵの判別された能力に基づいてドライブ２４２で能力テーブルをアップデートする。いくつかの実施形態において、判別された能力はＵ．２　ＧＰＵの現在の使用効率、Ｕ．２　ＧＰＵの特徴、Ｕ．２　ＧＰＵゼネレーション（ｇｅｎｅｒａｔｉｏｎ）、Ｕ．２　ＧＰＵの処理能力、スレッド（ｔｈｒｅａｄ）プロセッサの個数、Ｕ．２　ＧＰＵのＤＲＡＭサイズ、帯域幅、遅延時間、精密度、入出力サイズ、及びＭＨｚ動作速度の中の一つ以上を含む。例えば、ＢＭＣ２３４がスロット－３でＵ．２　ＧＰＵ２５０がドライブ２４２に対する変換－１又は変換－２を遂行すると判別した場合、ドライブ２４２に対する能力テーブルは、変換－１及び変換－２に対応するものとしてスロット－３のアドレスを含むようにアップデートされる。５２８ステップで、ＢＭＣ２３４がＢＭＣ２３４に連結される一つ以上のドライブ２４２が能力テーブルをアップデートしなかったと判別した場合、ＢＭＣ２３４は、５３０ステップに戻り、次のドライブ２４２をアップデートする。ＢＭＣ２３４がＢＭＣ２３４に連結された各ドライブ２４２が能力テーブルをアップデートしたと判別した場合、手順は５１０ステップに戻り、ＢＭＣ２３４は、他のパワーオン、プラグイン、プラグアウトのイベントの感知の待機を再開する。

【００３５】
ＢＭＣ２３４が５１０ステップで感知されたイベントが、パワーオンのイベントではないと判別した場合、手順は５３２ステップに進行する。５３２ステップで、ＢＭＣ２３４がイベントをプラグインのイベントであると判別した場合、ＢＭＣ２３４はプラグインのイベントを誘発したＢＭＣ２３４に連結されたＵ．２　ＧＰＵ２５０の能力を判別する。このプラグインのイベントは、Ｕ．２　ＧＰＵが現場交換可能な装置として具現され、現場交換可能なＵ．２　ＧＰＵが既に一つ以上のドライブ２４２を含むシャシーに連結され

る場合に対応する。Ｕ．２　ＧＰＵ２５０の能力を判別する例として、図６に示すように、５３４ステップで、ＢＭＣ２３４はＵ．２　ＧＰＵ２５０に対する必須製品のデータ（ＧＰＵ．ＶＰＤ）をリード（ｒｅａｄ）する。５２６ステップで、ＢＭＣ２３４は追加的に又は代替的にＵ．２　ＧＰＵ２５０の能力に関連する情報を集めるためにＮＶＭｅ“識別”コマンドを利用する。プラグインのイベントを誘発したＵ．２　ＧＰＵ２５０の能力が判別された場合、手順は５４０及び５３８ステップに進行する。５３８及び５４０ステップで、５３０及び５２８ステップに対して上述したように、ＢＭＣ２３４は、新たなＵ．２　ＧＰＵ２５０の能力に基づいてＢＭＣ２３４に連結されたドライブ２４２の能力テーブルをアップデートする。

【００３６】

ＢＭＣ２３４がイベントをプラグインのイベントではないと判別した場合、例えばＢＭＣ２３４がイベントをプラグアウトのイベントであると判別した場合、手順は５４０ステップに進行する。このプラグアウトのイベントは、Ｕ．２　ＧＰＵが現場交換可能な装置として具現され、Ｕ．２　ＧＰＵがシャシー２３１から除去された場合に対応する。５４０及び５３８ステップで、ＢＭＣ２３４はＵ．２　ＧＰＵ２５０に対応するプラグアウトのイベントに基づいてドライブ２４２の能力テーブルをアップデートする。例えば、ＢＭＣ２３４は能力テーブルからＵ．２　ＧＰＵ２５０の以前のアドレスを除去する。

【００３７】

いくつかの実施形態において、パワーオン、プラグイン、及びプラグアウト以外に、ＢＭＣ２３４は、イベントがストレージ管理イベントであると判別する。これはストレージ管理者（例えば、ホストサーバ２１０のアプリケーション２１２）が、どのドライブ２４２がどのＵ．２　ＧＰＵ２５０を利用するかを変更したことを示す。イベントがストレージ管理のイベントの場合、ＢＭＣ２３４は、これによって能力テーブルをアップデートする。例えば、ストレージ管理のイベントは特定のＵ．２　ＧＰＵ２５０が特定のドライブ２４２を支援できないことを示し、ＢＭＣ２３４は特定のドライブ２４２の能力テーブルから特定のＵ．２　ＧＰＵ２５０を除去する。

【００３８】

いくつかの実施形態において、ＢＭＣ２３４がドライブ２４２の能力テーブルをアップデートした場合、ＢＭＣ２３４は、どのドライブに対して及びどの変換のために利用可能なものとしてリストするためのＵ．２　ＧＰＵを判別するに当たって、各Ｕ．２　ＧＰＵの能力以外の考慮事項を考慮する。例えば、一実施形態において、ＢＭＣ２３４は、特定のドライブ２４２を特定のＵ．２　ＧＰＵに割り当てることで、Ｕ．２　ＧＰＵ２５０に対する負荷の均衡をとる。このような実施形態において、ＢＭＣ２３４は、割り当てられたＵ．２　ＧＰＵ２５０のアドレスを含むように、与えられたドライブ２４２の能力テーブルのみをアップデートする。他の実施形態において、ＢＭＣ２３４は特定のＵ．２　ＧＰＵ２５０の能力に基づいて特定の変換を処理するために特定のＵ．２　ＧＰＵ２５０を割り当てる。このような実施形態において、ＢＭＣ２３４は割り当てられたＵ．２　ＧＰＵ又は特定の変換のためのＵ．２　ＧＰＵのアドレスを含むように能力テーブルをアップデートするが、Ｕ．２　ＧＰＵ能力が他の変換を処理するのに十分であるとしても、他の変換のためのアドレスを含まないこともある。例えば、ＢＭＣ２３４は変換－３に対してスロット－２に位置する最も強力なＵ．２　ＧＰＵ２５０を予約する。スロット－２でＵ．２　ＧＰＵは、変換－１～変換－９を処理するために適切な能力を有する。ＢＭＣ２３４は変換－３に対応するものとしてスロット－２を含むように能力テーブルをアップデートするが、他の変換に対してスロット－２を含まないことも有り、従って変換－３以外の他の変換に対してドライブ２４２がスロット－２でＵ．２　ＧＰＵを利用できないようにする。いくつかの実施形態において、ＢＭＣ２３４は、利用可能なＵ．２　ＧＰＵの特徴、現在利用可能なＵ．２　ＧＰＵの特徴（例えば、専有されないか又は完全に割り当てられない）、及び／又は同時に実行する並列アルゴリズムの個数と比較して、変換に関与するアルゴリズムの類型に基づいてロードバランシング（ｌｏａｄ　ｂａｌａｎｃｉｎｇ）を遂行する。

【００３９】
本明細書で使用した用語は、特定の実施形態を説明するための目的であり、本発明の思想を制限しようとするものではない。本明細書で使用したように、文脈が明確に違うことを意味しない限り、単数形は複数形を含むものとして意図する。“含む”、“含み”の用語は、本明細書で使用した場合、明示された特徴、整数、ステップ、動作、要素及び／又は部品の存在を明示するが、一つ以上の他の特徴、整数、ステップ、動作、要素、部品及び／又はグループの存在又は追加を排除しない。本明細書で使用したように、“及び／又は”の用語は一つ以上の関連する列挙項目の全ての組合せ及び特定の組合せを含む。“少なくても一つ”のような表現は、要素のリストの前にある場合、全体要素のリストを修正し、リストの個別要素を修正しない。

【００４０】
本明細書で使用したように、“できる”の使用は、本発明の実施形態を説明する場合、“本発明の一つ以上の実施形態”を指し示す。本明細書で使用したように、“使用”、“使用する”、“使用された”の用語は“利用”、“利用する”、“利用された”の用語に各々同義語と見なされ得る。なお、“例示的な”の用語は例示又は図面を指し示すものとして意図する。

【００４１】
本明細書で説明した本発明の実施形態による電子又は電気装置及び／又は他の関連する装置又は部品は、任意の適切なハードウェア、ファームウェア（例えば、応用注文型集積回路（ＡＳＩＣ：ａｐｐｌｉｃａｔｉｏｎ－ｓｐｅｃｉｆｉｃ　ｉｎｔｅｇｒａｔｅｄ　ｃｉｒｃｕｉｔ））、ソフトウェア、又はソフトウェア、ファームウェア、及びハードウェアの組合せを利用して具現される。例えば、このような装置の多様な部品は、一つの集積された回路（ＩＣ）チップ又は別途のＩＣチップに形成される。なお、このような装置の多様な部品は、フレキシブル印刷回路フィルム、テープキャリアパッケージ（ＴＣＰ：Ｔａｐｅ　Ｃａｒｒｉｅｒ　Ｐａｃｋａｇｅ）、印刷回路基板（ＰＣＢ：Ｐｒｉｎｔｅｄ　Ｃｉｒｃｕｉｔ　Ｂｏａｒｄ）上に具現されるか又は一つの基板に形成される。なお、このような装置の多様な部品は、一つ以上のコンピューティング装置で、本明細書で説明した多様な機能を遂行するためコンピュータプログラムの命令を実行し、他のシステム部品と相互作用する一つ以上のプロセッサで実行されるプロセス又はスレッドである。コンピュータプログラムの命令は、例えばランダムアクセスメモリー（ＲＡＭ）のような標準メモリー装置を用いるコンピューティング装置で具現されるメモリーに格納される。また、コンピュータプログラムの命令は、例えばＣＤ－ＲＯＭ、フラッシュドライブなどのような他の非一時的なコンピュータ読み取り可能なメディアに格納される。また、当業者（通常の技術者）は多様なコンピューティング装置の機能性が一つのコンピューティング装置に結合されるか又は統合されることを認識しなければならない。また、本発明の例示的な実施形態の思想及び範囲を逸脱せずに特定のコンピューティング装置の機能性が一つ以上の他のコンピューティング装置に分散されることを認識しなければならない。

【００４２】
別途に定義しない場合、本明細書で使用した全ての用語（技術的及び科学的用語を含む）は、本発明が属する技術における当業者により、一般的に理解されるものと同一の意味を有する。用語は、一般的に利用される辞典で定義されるように、関連する技術及び／又は本明細書の文脈上の意味と一致する意味を有するものとして解釈されるべきであり、本明細書で明確に定義しない限り、理想的であるか形式的な感覚により解釈してはならない。

【００４３】
以上、本発明の実施形態について図面を参照しながら詳細に説明したが、本発明は、上述の実施形態に限定されるものではなく、本発明の技術的範囲から逸脱しない範囲内で多様に変更実施することが可能である。

【産業上の利用可能性】
【００４４】

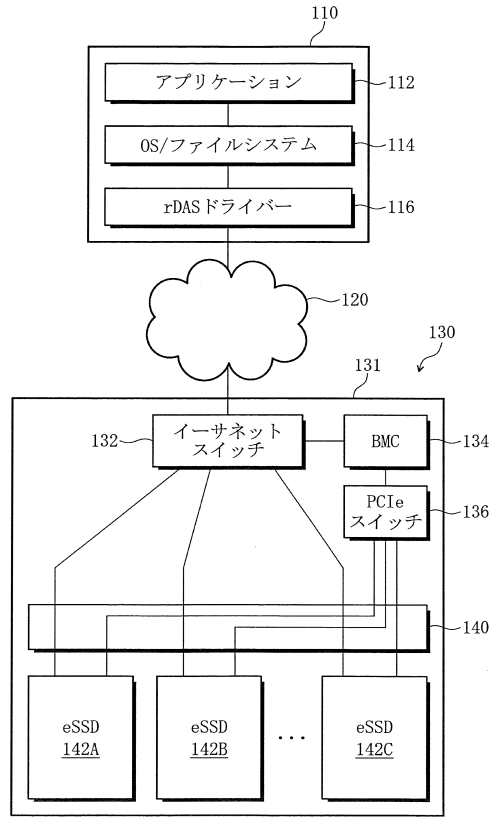
本発明は、ストレージシステムにおいて、複数のグラフィック処理装置を利用してマシンラーニングアルゴリズムを実行するデータストレージ及び処理システムとその動作方法に有用である。

【符号の説明】

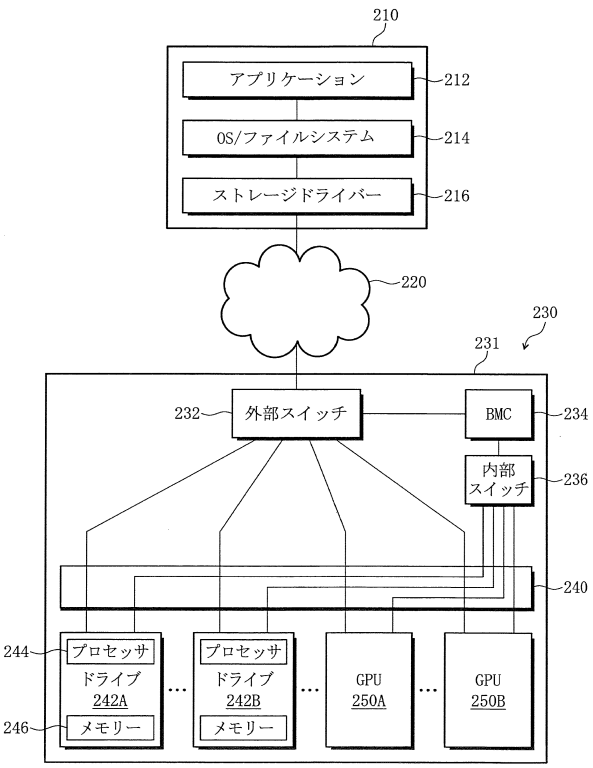
【００４５】

１１０、２１０　ホストサーバ
１１２、２１２　アプリケーション
１１４、２１４　ＯＳ／ファイルシステム
１１６　遠隔直接アクセスストレージ（ｒＤＡＳ：ｒｅｍｏｔｅ　Ｄｉｒｅｃｔ　Ａｃｃｅｓｓ　Ｓｔｏｒａｇｅ）ドライバー
１２０、２２０　ネットワーク
１３０、２３０　ストレージ部
１３１、２３１　シャシー（ｃｈａｓｓｉｓ）
１３２　イーサネット（ｅｔｈｅｒｎｅｔ）スイッチ
１３４、２３４　ベースボード管理制御器（ＢＭＣ：ｂａｓｅｂｏａｒｄ　ｍａｎａｇｅｍｅｎｔ　ｃｏｎｔｒｏｌｌｅｒ）
１３６　ＰＣＩｅスイッチ
１４０、２４０　シャシーミッドプレーン（ｃｈａｓｓｉｓ　ｍｉｄｐｌａｎｅ）
１４２Ａ～１４２Ｃ　ｅＳＳＤ（ｅｔｈｅｒｎｅｔ　Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅｓ）
２１６　ストレージドライバー（遠隔直接アクセスストレージドライバー）
２３２　外部スイッチ（イーサネットスイッチ）
２３６　内部スイッチ（ＰＣＩｅスイッチ）
２４２、２４２Ａ～２４２Ｂ　ドライブ
２４４、３６０　プロセッサ
２４６　メモリー
２５０、２５０Ａ～２５０Ｂ、３５０　Ｕ．２　グラフィック処理装置（ＧＰＵ）
３５２　コネクター
３５４　イーサネットインターフェース
３５６　ＰＣＩｅインターフェース
３５８　システム管理バス（ＳＭＢｕｓ）インターフェース
３６２　ＤＲＡＭ
３６４　不揮発性メモリー（ＮＶＭ）
３６６　電源供給器／電源調節器
５００　能力テーブル
６２０　アルゴリズム

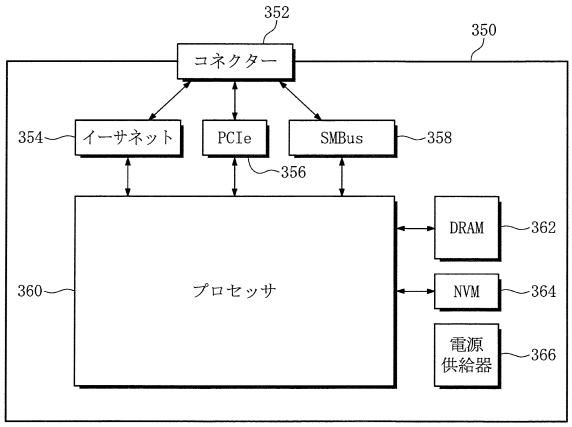
【図１】


110
アプリケーション
112
OS/ファイルシステム
114
rDASドライバー
116
120
130
131
132
イーサネット
スイッチ
BMC
134
PCIe
スイッチ
136
140
eSSD
142A
eSSD
142B
eSSD
142C

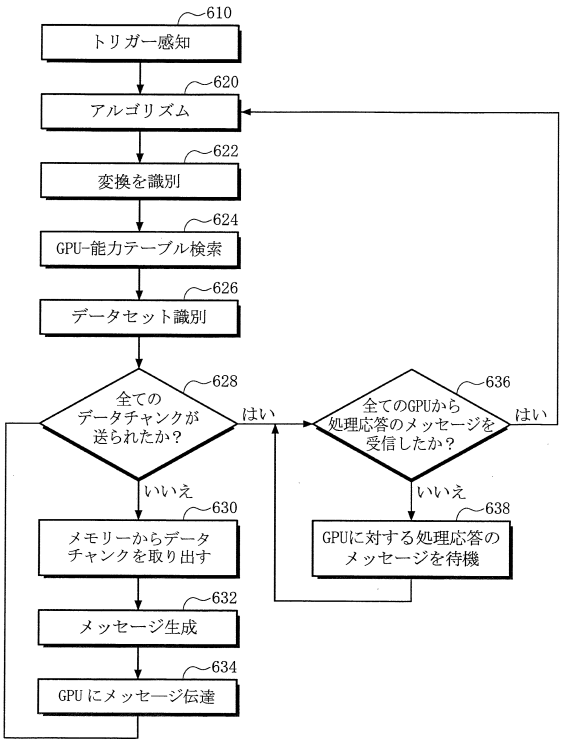
【図２】


210
アプリケーション
212
OS/ファイルシステム
214
ストレージドライバー
216
220
230
231
232
外部スイッチ
BMC
234
内部
スイッチ
236
240
244
プロセッサ
ドライブ
242A
メモリー
246
プロセッサ
ドライブ
242B
メモリー
GPU
250A
GPU
250B

【図３】


352
350
コネクター
354
イーサネット
PCIe
356
SMBus
358
360
プロセッサ
DRAM
362
NVM
364
電源
供給器
366

【図４】

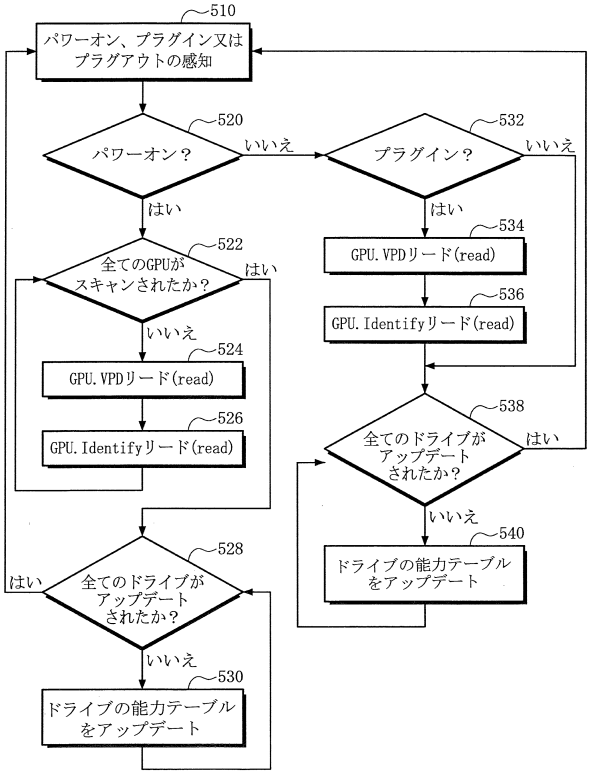

610
トリガー感知
620
アルゴリズム
622
変換を識別
624
GPU-能力テーブル検索
626
データセット識別
628
全ての
データチャンクが
送られたか？
はい
いいえ
630
メモリーからデータ
チャンクを取り出す
632
メッセージ生成
634
GPU にメッセージ伝達
636
全てのGPUから
処理応答のメッセージを
受信したか？
はい
いいえ
638
GPUに対する処理応答の
メッセージを待機

【図５】

500

変換	GPU 住所	GPU 住所	GPU 住所	GPU 住所	GPU 住所
変換 1	スロット-3	スロット-4			
変換 2	スロット-3	スロット-5	スロット-6	スロット-13	
変換 n	スロット-7				

【図６】



フロントページの続き

(72)発明者　オラリグ，　ソンポン　ポール
アメリカ合衆国，　９４５６６，　カリフォルニア州，　プレザントン，　パセオ　グラナダ　３０５０
(72)発明者　シュワドラー，　デイビッド
アメリカ合衆国，　９５０７０，　カリフォルニア州，　サラトガ，　パセオ　プレサーダ　１３１６５

審査官　松平　英

(56)参考文献　米国特許出願公開第２０１４／０１２９７５３（ＵＳ，Ａ１）
米国特許出願公開第２０１１／０２９２０５８（ＵＳ，Ａ１）
国際公開第２０１６／１８５５４２（WO，Ａ１）
米国特許出願公開第２０１１／０２９５９６７（ＵＳ，Ａ１）
米国特許出願公開第２０１４／０１７６５８３（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　３／０６
Ｇ０６Ｆ　　３／０８
Ｇ０６Ｆ　１３／１０
Ｇ０６Ｎ　２０／００