

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7017546号
(P7017546)

(45)発行日 令和4年2月8日(2022.2.8)

(24)登録日 令和4年1月31日(2022.1.31)

(51)国際特許分類	F I			
G 0 6 F 13/14 (2006.01)	G 0 6 F	13/14	3 1 0 H	
G 0 6 F 13/10 (2006.01)	G 0 6 F	13/10	3 4 0 A	
	G 0 6 F	13/14	3 3 0 A	

請求項の数 4 (全17頁)

(21)出願番号	特願2019-176446(P2019-176446)	(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	令和1年9月27日(2019.9.27)	(74)代理人	110000279 特許業務法人ウィルフォート国際特許事務所
(65)公開番号	特開2021-56567(P2021-56567A)	(72)発明者	井上 真理 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
(43)公開日	令和3年4月8日(2021.4.8)	(72)発明者	岩満 幸治 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
審査請求日	令和2年2月6日(2020.2.6)	(72)発明者	戸塚 崇夫 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

最終頁に続く

(54)【発明の名称】 ストレージシステム、パス管理方法、及びパス管理プログラム

(57)【特許請求の範囲】

【請求項1】

1以上の記憶デバイスを有するストレージノードを複数備えるストレージシステムであって、

前記ストレージノードは、プロセッサ部を備え、

前記プロセッサ部は、上位装置から前記記憶デバイスの記憶領域が提供される所定の論理ユニットにアクセス可能なパスの中から、前記上位装置に対して使用可能なパスとして通知する優先パスを選択し、

前記上位装置からの前記所定の論理ユニットへのパスの問い合わせに対して前記優先パスを応答するようにし、

前記プロセッサ部は、

前記上位装置の設定可能なパスの数を特定可能なパス数特定情報を取得し、

前記パス数特定情報により特定される設定可能なパスの数だけ前記優先パスを決定し、

前記ストレージノードは、

前記論理ユニットに対するデータのアクセスを制御するストレージ制御部を備え、

同一の論理ユニットへのアクセスを担当する2つのストレージノードのストレージ制御部のペアが構成され、一方のストレージ制御部は、通常用いられる現用系に設定され、他方のストレージ制御部は、待機系に設定されており、

前記プロセッサ部は、

前記論理ユニットにアクセス可能なパスの中から、前記論理ユニットに対する排他アク

セス用のキーが設定されているパス、前記論理ユニットの現用系に設定されているストレージ制御部へのパス、前記論理ユニットの待機系に設定されているストレージ制御部へのパス、および前記論理ユニットの待機系に設定されているストレージ制御部を備えるストレージノードにおける現用系に設定されているストレージ制御部に対応する待機系が設定されているストレージ制御部へのパスを前記優先パスとして選択する
ストレージシステム。

【請求項 2】

前記パス数特定情報は、前記上位装置で実行されている OS の種類の情報である
請求項 1 に記載のストレージシステム。

【請求項 3】

1 以上の記憶デバイスを有するストレージノードを複数備えるストレージシステムによる
パス管理方法であって、

前記ストレージノードは、

上位装置から前記記憶デバイスの記憶領域が提供される所定の論理ユニットにアクセス可能なパスの中から、前記上位装置に対して使用可能なパスとして通知する優先パスを選択し、

前記上位装置からの前記所定の論理ユニットへのパスの問い合わせに対して前記優先パスを応答するようにし、

前記ストレージノードは、

前記上位装置の設定可能なパスの数を特定可能なパス数特定情報を取得し、

前記パス数特定情報により特定される設定可能なパスの数だけ前記優先パスを決定し、

前記ストレージノードは、

前記論理ユニットに対するデータのアクセスを制御するストレージ制御部を備え、

同一の論理ユニットへのアクセスを担当する 2 つのストレージノードのストレージ制御部のペアが構成され、一方のストレージ制御部は、通常用いられる現用系に設定され、他方のストレージ制御部は、待機系に設定されており、

前記ストレージノードは、

前記論理ユニットにアクセス可能なパスの中から、前記論理ユニットに対する排他アクセス用のキーが設定されているパス、前記論理ユニットの現用系に設定されているストレージ制御部へのパス、前記論理ユニットの待機系に設定されているストレージ制御部へのパス、および前記論理ユニットの待機系に設定されているストレージ制御部を備えるストレージノードにおける現用系に設定されているストレージ制御部に対応する待機系が設定されているストレージ制御部へのパスを前記優先パスとして選択する

パス管理方法。

【請求項 4】

1 以上の記憶デバイスを有するストレージノードを複数備えるストレージシステムにおけるストレージノードを構成するコンピュータに実行させるパス管理プログラムであって、前記コンピュータに、

上位装置から前記記憶デバイスの記憶領域が提供される所定の論理ユニットにアクセス可能なパスの中から、前記上位装置に対して使用可能なパスとして通知する優先パスを選択させ、

前記上位装置からの前記所定の論理ユニットへのパスの問い合わせに対して前記優先パスを応答するようにさせ、

前記上位装置の設定可能なパスの数を特定可能なパス数特定情報を取得させ、

前記パス数特定情報により特定される設定可能なパスの数だけ前記優先パスを決定させ、

前記ストレージノードは、

前記論理ユニットに対するデータのアクセスを制御するストレージ制御部を備え、

同一の論理ユニットへのアクセスを担当する 2 つのストレージノードのストレージ制御部のペアが構成され、一方のストレージ制御部は、通常用いられる現用系に設定され、他方のストレージ制御部は、待機系に設定されており、

10

20

30

40

50

前記コンピュータに、

前記論理ユニットにアクセス可能なパスの中から、前記論理ユニットに対する排他アクセス用のキーが設定されているパス、前記論理ユニットの現用系に設定されているストレージ制御部へのパス、前記論理ユニットの待機系に設定されているストレージ制御部へのパス、および前記論理ユニットの待機系に設定されているストレージ制御部を備えるストレージノードにおける現用系に設定されているストレージ制御部に対応する待機系が設定されているストレージ制御部へのパスを前記優先パスとして選択させる

パス管理プログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、1以上の記憶デバイスを有するストレージノードを複数備えるストレージシステム等に関し、上位装置から論理デバイスにアクセスするパスを管理する技術に関する。

【背景技術】

【0002】

近年、企業や官庁等において蓄積されるデータ量は増加の一途を辿っており、データを格納するストレージシステムにおいては、ストレージ装置を容易にスケールアウトできる構成であることが望ましい。

【0003】

ストレージシステムに格納されるデータは論理ユニットとして管理されている。ストレージシステムに格納されるデータを利用する上位装置（ホスト装置）では、ストレージシステムに管理されている論理ユニットに対する接続パスを管理している。ホスト装置は、ストレージシステムに対して論理ユニットに対する接続パスを問い合わせることで把握することができる。

20

【0004】

ホスト装置においては、実行されるOS（オペレーティングシステム）の種類によって、設定できる接続パスの上限が異なっている。このため、スケールアウト等によりストレージ装置の数が増加すると、ストレージシステムから得られる接続パスの数が、OSに設定できる接続パスの上限を超える場合がある。

【0005】

このように、接続パスの数が上限を超えている場合には、OS側の再設定等を手動で行う必要があり、手間がかかる。

30

【0006】

これに対して、特許文献1には、削除させる論理パスを選定して、ホスト装置に通知することにより、ホスト装置におけるOS上の論理パスを自動更新する技術が開示されている。

【先行技術文献】

【特許文献】

【0007】

【文献】特開2018-156144号公報

【発明の概要】

40

【発明が解決しようとする課題】

【0008】

しかし、特許文献1に開示された技術においては、ホスト装置側で、ストレージ装置からのパスの削除の通知に対応してパスを削除する機能を備える必要がある。すなわち、ホスト装置をストレージ装置に合わせて機能を追加しておかなければならないという問題がある。

【0009】

また、ホスト装置の他の構成や制限等により、またはホスト装置のユーザのポリシー等により、ホスト装置側で使用するパスを選択することが要請される場合がある。

【0010】

本発明は、上記事情に鑑みなされたものであり、その目的は、上位装置に適切なパスを容

50

易に認識させることのできる技術を提供することにある。

【課題を解決するための手段】

【0011】

上記目的を達成するため、一観点に係るストレージシステムは、1以上の記憶デバイスを有するストレージノードを複数備えるストレージシステムであって、ストレージノードは、プロセッサ部を備え、プロセッサ部は、上位装置から記憶デバイスの記憶領域が提供される所定の論理ユニットにアクセス可能なパスの中から、上位装置に対して使用可能なパスとして通知する優先パスを選択し、上位装置からの所定の論理ユニットへのパスの問い合わせに対して優先パスを応答するようにする。

【発明の効果】

10

【0012】

本発明によれば、上位装置に適切なパスを容易に認識させることができる。

【図面の簡単な説明】

【0013】

【図1】図1は、一実施形態に係る計算機システムの全体構成図である。

【図2】図2は、一実施形態に係るストレージノードの構成図である。

【図3】図3は、一実施形態に係る計算機システムの機能構成図である。

【図4】図4は、一実施形態に係るCTLMテーブルの構成図である。

【図5】図5は、一実施形態に係るパス優先度管理テーブルの構成図である。

【図6】図6は、一実施形態に係るCLMテーブルの構成図である。

20

【図7】図7は、一実施形態に係るノードペア管理テーブルの構成図である。

【図8】図8は、一実施形態に係るSCチェーンテーブルの構成図である。

【図9】図9は、一実施形態に係るパス優先度選択処理のフローチャートである。

【図10】図10は、一実施形態に係るボリュームパス作成処理のフローチャートである。

【図11】図11は、一実施形態に係るボリュームパス追加処理のフローチャートである。

【図12】図12は、一実施形態に係るCTLMテーブル生成処理のフローチャートである。

【発明を実施するための形態】

【0014】

実施形態について、図面を参照して説明する。なお、以下に説明する実施形態は特許請求の範囲に係る発明を限定するものではなく、また実施形態の中で説明されている諸要素及びその組み合わせの全てが発明の解決手段に必須であるとは限らない。

30

【0015】

以下の説明では、「AAAテーブル」の表現にて情報を説明することがあるが、情報は、どのようなデータ構造で表現されていてもよい。すなわち、情報がデータ構造に依存しないことを示すために、「AAAテーブル」を「AAA情報」と呼ぶことができる。

【0016】

また、以下の説明では、「プロセッサ部」は、1以上のプロセッサを含む。少なくとも1つのプロセッサは、典型的には、CPU (Central Processing Unit) のようなマイクロプロセッサである。1以上のプロセッサの各々は、シングルコアでもよいしマルチコアでもよい。プロセッサは、処理の一部または全部を行うハードウェア回路を含んでもよい。

40

【0017】

図1は、一実施形態に係る計算機システムの全体構成図である。

【0018】

計算機システム1は、1以上のコンピュータノード2と、ストレージシステム3とを備える。ストレージシステム3は、複数のストレージノード4を備える。コンピュータノード2と、ストレージシステム3の各ストレージノード4とは、ネットワーク6を介して接続されている。

【0019】

ネットワーク6は、例えばファイバーチャネル (FC: Fibre Channel)、

50

イーサネット（登録商標）、InfiniBand又は無線LAN（Local Area Network）などから構成されている。

【0020】

また、ストレージノード4は、バックエンドネットワーク5を介して接続されている。バックエンドネットワーク5は、例えば、イーサネット（登録商標）、InfiniBand又は無線LANなどから構成されている。

【0021】

なお、ネットワーク6及びバックエンドネットワーク5は、同一のネットワークにより構成されていてもよい。また、各コンピュータノード2及び各ストレージノード4は、ネットワーク6やバックエンドネットワーク5以外の管理用ネットワークに接続されていてもよい。

10

【0022】

コンピュータノード2は、例えば、汎用のコンピュータにより構成され、ストレージノード3に対してホスト（上位装置）として機能する。なお、コンピュータノード2は、仮想マシンのような仮想的なコンピュータ装置であってもよい。

【0023】

コンピュータノード2は、例えば、ストレージシステム3へのログイン時に、ストレージシステム3の各パスを介して、所定の論理ユニット（LU）にアクセス可能なパスか否かを問い合わせ、ストレージシステム3からアクセス可能なパスであるとの応答を受け付けた場合には、そのパスをその論理ユニットに接続可能なパスとして設定する。なお、コンピュータノード2において設定可能なパスの数は、コンピュータノード2のOS（Operating System）によって所定数に制限される場合がある。

20

【0024】

また、コンピュータノード2は、ユーザ操作や実装されたアプリケーションプログラムからの要求に応じて、ネットワーク6を介してストレージシステム3にリード要求又はライト要求（アクセス要求、又はI/O（Input/Output）要求という）を送信する。

【0025】

ストレージノード4は、例えば、汎用の物理サーバで構成され、コンピュータノード2に対してデータを読み書きするための記憶領域を提供する記憶デバイス12を有する。

30

【0026】

図2は、一実施形態に係るストレージノードの構成図である。

【0027】

ストレージノード4は、例えば、PC（Personal Computer）によって構成されている。ストレージノード4は、1以上のCPU（Central Processing Unit）10と、1以上のメモリ11と、複数の記憶デバイス12と、1以上のネットワークインターフェース（I/F）13と、1以上のネットワークI/F14と、を備えている。CPU10と、メモリ11と、記憶デバイス12と、ネットワークI/F13と、ネットワークI/F14とは、内部バス15を介して接続されている。

【0028】

ネットワークI/F13は、例えば、FCカード、有線LANカード、無線LANカードなどのインターフェースであり、ネットワーク6を介しての他の装置（例えば、コンピュータノード2）との通信時におけるプロトコル制御を行う。

40

【0029】

ネットワークI/F14は、例えば、NIC（Network Interface Card）カード、有線LANカード、無線LANカードなどのインターフェースであり、バックエンドネットワーク5を介しての他の装置（ストレージノード4）との通信時におけるプロトコル制御を行う。

【0030】

CPU10は、メモリ11に格納されているプログラムに従って各種処理を実行する。

50

【0031】

メモリ11は、例えば、SRAM(Static RAM(Random Access Memory))やDRAM(Dynamic RAM)などの揮発性の半導体メモリであり、CPU10で実行されるプログラムや、必要な情報を記憶する。

【0032】

メモリ11は、パス優先度管理テーブル20と、ノードペア管理テーブル30と、CLM(Compute Target Mapping)テーブル40と、SC(Storage Controller)チェンテーブル50と、CTLM(Compute Target Lu Mapping)テーブル60と、パス管理プログラム80とを格納する。パス管理プログラム80は、CPU10に実行されることにより、パスを管理する処理、例えば、後述するパス優先度選択処理、ボリュームパス作成処理、ボリュームパス追加処理、CTLMテーブル生成処理を実行する。メモリ11の各テーブルの詳細については後述する。

10

【0033】

記憶デバイス12は、例えば、NVMe(Non-Volatile Memory)ドライブやSAS(Serial Attached SCSI(Small Computer System Interface))ドライブ、SATA(Serial ATA(Advanced Technology Attachment))、SSD(Solid State Drive)又はSCM(Storage Class Memory)などの不揮発性の記憶デバイスなどであり、コンピュータノード2に対してデータを読み書きするための記憶領域を有する論理ユニット(ボリューム)を提供する。

20

【0034】

図3は、一実施形態に係る計算機システムの機能構成図である。

【0035】

本実施形態では、ストレージシステム3は、複数のストレージノード4をグループとしたクラスタ7を単位として管理している。なお、図3においては、クラスタ7を1つとしているが、クラスタを複数備えるようにしてもよい。

【0036】

ストレージノード4は、コントロールプレーン71と、フロントエンド部(FE部)72と、1以上のストレージ制御部(SC部)74とを備えている。

30

【0037】

コントロールプレーン71は、クラスタ7全体や自ノードを制御する。また、コントロールプレーン71は、コンピュータノード2からLUのパスに対する問合せがあった場合には、CTLMテーブル60が存在する場合には、CTLMテーブル60を参照し、問い合わせ元のコンピュータノード2のコンピュータノード番号と、問合せ対象のLUのLUNとが登録されているか否かを判定し、登録されている場合には、アクセス可能として、コンピュータノード2に応答する。一方、CTLMテーブル60が無い場合には、コントロールプレーン71は、CLMテーブル40を参照し、問い合わせ元のコンピュータノード2のコンピュータノード番号と、問合せ対象のLUのLUNとが登録されているか否かを判定し、登録されている場合には、アクセス可能として、コンピュータノード2に応答する。

40

【0038】

FE部72は、ストレージノード4におけるI/O処理のフロントエンドとして機能する。例えば、FE部72は、コンピュータノード2からI/O要求が与えられた場合に、そのI/O要求を、自ノード内のストレージ制御部74や、他ノード内のストレージ制御部74に振り分ける。

【0039】

ストレージ制御部74は、SDS(Software Defined Storage)のコントローラとして機能する。本実施形態では、ストレージ制御部74は、図示しないストレージ制御ソフトウェアをCPU10が実行することにより構成される。ストレージ

50

制御部 74 は、FE 部 72 から与えられるコンピュータノード 2 からの I/O 要求を受け付け、受け付けた I/O 要求に応じて、LU 73 に対する I/O 処理を実行する。

【0040】

本実施形態では、ストレージノード 4 の各ストレージ制御部 74 は、別のストレージノード 4 のストレージ制御部 74 とで、冗長化構成を構成するペア（ストレージ制御部ペア）75 として設定される。なお、3 つ以上のストレージ制御部 74 により冗長化構成を形成するようにしてもよい。

【0041】

このストレージ制御部ペア 75 においては、一方のストレージ制御部 74 がコンピュータノード 2 からの I/O 要求を受け付けることができる状態（現用系の状態、アクティブモードという）に設定され、他方のストレージ制御部 74 がコンピュータノード 2 からの I/O 要求を受け付けない状態（待機系の状態、スタンバイモードという）に設定される。

10

【0042】

ストレージ制御部ペア 75 では、アクティブモードに設定されたストレージ制御部 74（アクティブのストレージ制御部という）やこのストレージ制御部 74 が配置されたストレージノード 4 に障害が発生した場合などに、スタンバイモードに設定されていたストレージ制御部 74（スタンバイのストレージ制御部という）がアクティブモードに切り替えられる。これにより、アクティブのストレージ制御部 74 が稼働し得なくなった場合に、このストレージ制御部 74 が実行していた I/O 処理を、ストレージ制御部ペア 75 を構成するスタンバイのストレージ制御部 74 が引き継ぐことになる。

20

【0043】

ストレージノード 4 においては、コンピュータノード 2 に対して、データを格納する記憶領域として LU 73 を提供している。LU 73 は、ストレージノード 4 内におけるボリュームと対応付けられて管理されている。コンピュータノード 2 は、LU 73 にアクセスする際には、ストレージノード 4 に定義されたターゲット 76 に対して I/O 要求を送信する。これを実現するために、コンピュータノード 2 は、LU 73 にアクセスするための、自ノードの物理ポート（図示せず）から、ストレージノード 4 のターゲット 76 までのパスを管理している。なお、ターゲット 76 は、ストレージノード 4 のネットワーク I/F 13 の物理ポートと 1 対 1 に対応付けられている。

【0044】

図 4 は、一実施形態に係る C T L M テーブルの構成図である。

30

【0045】

C T L M テーブル 60 は、コンピュータノード 2 に提供される論理ユニット（Logical Unit）と、ストレージシステム 3 におけるボリュームとの対応関係を格納する。C T L M テーブル 60 のエントリは、コンピュータノード #（番号）61 と、ターゲット #（番号）62 と、L U N 64 と、V o l u m e #（番号）65 とのフィールドを含む。なお、L U N 64 と、V o l u m e #（番号）65 とのフィールドによって、L U N 管理テーブル 63 が構成される。なお、本実施形態においては、各ストレージノード 4 は、図 4 に示す C T L M テーブル 60 における自身に存在するターゲット番号に対応するエントリのみを格納するようにしている。

40

【0046】

コンピュータノード # 61 には、エントリに対応するコンピュータノード 2 の識別番号（コンピュータノード番号）が格納される。ターゲット # 62 には、エントリに対応するボリュームにアクセスする際のストレージノード 4 のターゲット 76 の識別番号（ターゲット番号）が格納される。L U N 64 には、エントリに対応する論理ユニットの識別番号（L U N）が格納される。L U N は、論理ユニットについてのコンピュータノード 2 での一意の番号である。V o l u m e #（番号）65 には、エントリに対応するボリュームの番号が格納される。ボリュームの番号は、ストレージシステム 3 内で一意の番号である。

【0047】

図 5 は、一実施形態に係るパス優先度管理テーブルの構成図である。

50

【 0 0 4 8 】

パス優先度管理テーブル 2 0 は、コンピュータノード 2 毎に設けられている。パス優先度管理テーブル 2 0 は、ボリューム毎の優先するターゲット（優先パスに対応）を管理するテーブルであり、ボリューム毎のエントリを有する。パス優先度管理テーブル 2 0 のエントリは、ボリューム番号 2 1 と、ストレージノード番号 2 2 と、優先ターゲット番号 2 3 と、第 2 優先ターゲット番号 2 4 と、第 3 優先ターゲット番号 2 5 と、第 4 優先ターゲット番号 2 6 とのフィールドを含む。

【 0 0 4 9 】

ボリューム番号 2 1 には、エントリに対応するボリュームの識別番号が格納される。ストレージノード番号 2 2 には、エントリに対応するボリュームを格納しているストレージノード 4 の識別番号が格納される。優先ターゲット番号 2 3 には、エントリに対応するボリュームにアクセスする際に優先して使用することが好ましい 1 番目のターゲットの番号が格納される。第 2 優先ターゲット番号 2 4 には、エントリに対応するボリュームにアクセスする際に優先して使用することが好ましい 2 番目のターゲットの番号が格納される。第 3 優先ターゲット番号 2 5 には、エントリに対応するボリュームにアクセスする際に優先して使用することが好ましい 3 番目のターゲットの番号が格納される。第 4 優先ターゲット番号 2 6 には、エントリに対応するボリュームにアクセスする際に優先して使用することが好ましい 4 番目のターゲットの番号が格納される。

【 0 0 5 0 】

図 6 は、一実施形態に係る C L M テーブルの構成図である。

【 0 0 5 1 】

C L M テーブル 4 0 は、ストレージノード 4 においてアクセス可能な各ボリュームについての、コンピュータノード 2 における論理ユニットと、ストレージシステム 3 におけるボリュームとの対応関係を格納する。C L M テーブル 4 0 のエントリは、コンピュータノード # (番号) 4 1 と、L U N 4 3 と、V o l u m e # (番号) 4 4 とのフィールドを含む。なお、L U N 4 3 と、V o l u m e # (番号) 4 4 とのフィールドによって、L U N 管理テーブル 4 2 が構成される。

【 0 0 5 2 】

コンピュータノード # 4 1 には、エントリに対応するコンピュータノード 2 の識別番号（コンピュータノード番号）が格納される。L U N 4 3 には、エントリに対応する論理ユニットの識別番号（L U N）が格納される。V o l u m e # (番号) 4 4 には、エントリに対応するボリュームの番号が格納される。

【 0 0 5 3 】

図 7 は、一実施形態に係るノードペア管理テーブルの構成図である。

【 0 0 5 4 】

ノードペア管理テーブル 3 0 は、ストレージ制御部ペア 7 5 を管理するためのテーブルであり、ストレージ制御部ペア 7 5 毎のエントリを格納する。ノードペア管理テーブル 3 0 のエントリは、ストレージ制御部ペア # (番号) 3 1 と、アクティブストレージノード # (番号) 3 2 と、スタンバイストレージノード # (番号) 3 3 とのフィールドを含む。

【 0 0 5 5 】

ストレージ制御部ペア番号 3 1 には、ストレージ制御部ペア 7 5 の識別番号（ストレージ制御部ペア番号）が格納される。アクティブストレージノード番号 3 2 には、エントリに対応するストレージ制御部ペア 7 5 の内のアクティブ状態のストレージ制御部 7 4 を格納するストレージノード 4（アクティブストレージノード）の識別番号が格納される。スタンバイストレージノード番号 3 3 には、エントリに対応するストレージ制御部ペア 7 5 の内のスタンバイ状態のストレージ制御部 7 4 を格納するストレージノード 4（スタンバイストレージノード）の識別番号が格納される。

【 0 0 5 6 】

図 8 は、一実施形態に係る S C チェーンテーブルの構成図である。

【 0 0 5 7 】

10

20

30

40

50

SC (ストレージ制御部) チェーンテーブル 50 は、アクティブのストレージ制御部からストレージ制御部ペア 75 の関係に基づいて迎れるストレージ制御部 74 を管理するテーブルであり、ストレージノード 4 毎のエントリを有する。SC チェーンテーブル 50 のエントリは、ストレージノード # 51 と、アクティブストレージ制御部 # 52 と、スタンバイストレージ制御部 # 53 と、第 2 スタンバイストレージ制御部 # 54 と、第 3 スタンバイストレージ制御部 # 55 とのフィールドを含む。

【 0 0 5 8 】

ストレージノード # 51 には、エントリに対応するストレージノード 4 の番号が格納される。アクティブストレージ制御部 # 52 には、エントリに対応するストレージノード 4 におけるボリュームの I/O 処理を実行するアクティブのストレージ制御部 74 の番号が格納される。スタンバイストレージ制御部 # 53 には、エントリに対応するアクティブのストレージ制御部 74 とストレージ制御部ペア 75 となっているスタンバイのストレージ制御部 74 の番号が格納される。第 2 スタンバイストレージ制御部 # 54 には、スタンバイストレージ制御部 # 53 の番号のスタンバイのストレージ制御部 74 があるストレージノード 4 におけるアクティブのストレージ制御部 74 とストレージ制御部ペア 75 となっているスタンバイのストレージ制御部 74 の番号が格納される。第 3 スタンバイストレージ制御部 # 55 には、第 2 スタンバイストレージ制御部 # 54 の番号のスタンバイのストレージ制御部 74 があるストレージノード 4 におけるアクティブのストレージ制御部 74 とストレージ制御部ペア 75 となっているスタンバイのストレージ制御部 74 の番号が格納される。

【 0 0 5 9 】

次に、計算機システム 1 における処理動作について説明する。まず、パス優先度選択処理について説明する。

【 0 0 6 0 】

図 9 は、一実施形態に係るパス優先度選択処理のフローチャートである。

【 0 0 6 1 】

パス優先度選択処理は、例えば、ストレージシステム 3 を起動する際に、ストレージノード 4 のコントロールプレーン 71 により実行される処理である。なお、コントロールプレーン 71 によるこの処理は、CPU 10 がパス管理プログラム 80 を実行することにより実現される。

【 0 0 6 2 】

まず、CPU 10 は、ストレージノード 4 で管理している LU (ボリュームと対応) の内の未処理の 1 つの LU を処理対象とし、処理対象の LU (対象 LU) のパスとして定義済みのターゲット情報を各ストレージノード 4 から取得する (ステップ S 11)。ここで、ターゲット情報としては、ターゲット番号と、LU 番号と、対応するパスが排他管理されている場合における排他管理用のキー (例えば、SCSI Reserve key) とが含まれていてもよい。

【 0 0 6 3 】

次いで、CPU 10 は、対象 LU のパスのターゲットにおける未処理のターゲットのパス (対象パス) を対象として以下の処理 (S 12 ~ S 17) を行う。まず、CPU 10 は、処理対象のターゲット (対象ターゲット) のターゲット情報として、排他管理用のキーが登録されているか否かを判定し (ステップ S 12)、登録されている場合 (ステップ S 12: Y) には、排他管理のために必要なパスであることを示しているため、このパスをパス優先度管理テーブル 20 に登録するために、処理をステップ S 17 に進める。

【 0 0 6 4 】

一方、排他管理用のキーが登録されていない場合 (ステップ S 12: N) には、CPU 10 は、このパスがアクティブのストレージ制御部 74 へのパスであるか否かを判定する (ステップ S 13)。

【 0 0 6 5 】

この結果、このパスがアクティブのストレージ制御部 74 へのパスである場合 (ステップ

10

20

30

40

50

S 1 3 : Y) には、対象 L U に高速にアクセスできるパスであることを示しているので、処理をステップ S 1 6 に進める。

【 0 0 6 6 】

一方、対象パスがアクティブのストレージ制御部 7 4 へのパスでない場合 (ステップ S 1 3 : N) には、CPU 1 0 は、対象パスがスタンバイのストレージ制御部 7 4 へのパスであるか否かを判定する (ステップ S 1 4) 。

【 0 0 6 7 】

この結果、対象パスがスタンバイのストレージ制御部 7 4 へのパスである場合 (ステップ S 1 4 : Y) には、アクティブのストレージ制御部 7 4 への障害時において、迅速なアクセスが可能なパスであることを示しているので、処理をステップ S 1 6 に進める。

10

【 0 0 6 8 】

一方、対象パスがスタンバイのストレージ制御部 7 4 へのパスでない場合 (ステップ S 1 4 : N) には、CPU 1 0 は、対象パスが第 2 のスタンバイのストレージ制御部 7 4 へのパスであるか否かを判定する (ステップ S 1 5) 。

ここで、第 2 のスタンバイのストレージ制御部 7 4 とは、スタンバイのストレージ制御部 7 4 と同一のストレージノード 4 に存在するアクティブのストレージ制御部 7 4 とストレージ制御部ペア 7 5 を構成している他のストレージノード 4 のスタンバイのストレージ制御部 7 4 である。

【 0 0 6 9 】

この結果、対象パスが第 2 のスタンバイのストレージ制御部 7 4 へのパスである場合 (ステップ S 1 5 : Y) には、アクティブのストレージ制御部 7 4 及びスタンバイのストレージ制御部 7 4 への障害時において、次に迅速なアクセスが可能なパスであることを示しているので、処理をステップ S 1 6 に進める。

20

【 0 0 7 0 】

一方、対象パスが第 2 のスタンバイのストレージ制御部 7 4 へのパスでない場合 (ステップ S 1 5 : N) には、CPU 1 0 は、処理をステップ S 1 8 に進める。

【 0 0 7 1 】

ステップ S 1 6 では、CPU 1 0 は、パス優先度管理テーブル 2 0 が枯渇しているか、すなわち、対応するエントリのすべての優先ターゲットのフィールド 2 3 ~ 2 6 にターゲット番号が格納されているか否かを判定する。

【 0 0 7 2 】

この結果、パス優先度管理テーブル 2 0 が枯渇している場合 (ステップ S 1 6 : Y) には、この対象 L U についてのパスの候補が所定数揃っていることを意味しているので、CPU 1 0 は、処理をステップ S 1 9 に進める。

30

【 0 0 7 3 】

一方、パス優先度管理テーブル 2 0 が枯渇していない場合 (ステップ S 1 6 : N) 、または、排他管理用のキーが登録されている場合 (ステップ S 1 2 : Y) には、CPU 1 0 は、パス優先度管理テーブル 2 0 の対応するエントリの優先ターゲットのフィールド (フィールド 2 3 ~ 2 6) の空いているフィールドに、ステップ S 1 2 ~ ステップ S 1 5 で Y と判定されたターゲットの番号を設定し (ステップ S 1 7) 、処理をステップ S 1 8 に進める。

40

【 0 0 7 4 】

ステップ S 1 8 では、CPU 1 0 は、ステップ S 1 1 で取得したターゲット数分のターゲットに対して処理をしたか否かを判定し、ターゲット数分のターゲットに対して処理をしていない場合 (ステップ S 1 8 : N) には、処理をステップ S 1 2 に進めて、他のターゲットを対象にステップ S 1 2 以降の処理を実行する。一方、ターゲット数分のターゲットに対して処理を行った場合 (ステップ S 1 8 : Y) には、CPU 1 0 は、処理をステップ S 1 9 に進める。

【 0 0 7 5 】

ステップ S 1 9 では、CPU 1 0 は、ストレージノード 4 で管理しているすべての L U に対して処理をしたか否かを判定し、すべての L U に対する処理をしていない場合 (ステッ

50

プ S 1 9 : N) には、処理をステップ S 1 1 に進めて、他の L U を処理対象として処理を行う。一方、すべての L U に対して処理をした場合 (ステップ S 1 9 : Y) には、処理を終了する。

【 0 0 7 6 】

以上のパス優先度選択処理によると、L U (ボリューム) にアクセスするパスとして、所定の条件に合致する所定数以内のターゲットへのパスを適切に選択して、パス優先度管理テーブル 2 0 に設定することができる。

【 0 0 7 7 】

次に、ボリュームパス作成処理について説明する。

【 0 0 7 8 】

図 1 0 は、一実施形態に係るボリュームパス作成処理のフローチャートである。

【 0 0 7 9 】

ボリュームパス作成処理は、新たな L U が作成された際に、新たな L U が作成されたストレージノード 4 において実行される。

【 0 0 8 0 】

ストレージノード 4 の C P U 1 0 は、C L M テーブル 4 0 の L U N 管理テーブル 4 2 に対して新たな L U と、この L U に対して割り当てられたボリュームとの対応関係を追加する (ステップ S 2 1) 。次いで、C P U 1 0 は、この L U を利用するコンピュータノード 2 の O S 種別情報 (パス数特定情報の一例) を取得する (ステップ S 2 2) 。

【 0 0 8 1 】

次いで、C P U 1 0 は、O S 種別情報に基づいて、コンピュータノード 2 の O S が特定 O S (ここでは、設定可能な L U のパスが所定数 (例えば、4 つ) 以下である O S) であるか否かを判定する (ステップ S 2 3) 。この結果、特定 O S でない場合 (ステップ S 2 3 : N) には、L U のパスの数を制限しなくてもよいので、ボリュームパス作成処理を終了する。

【 0 0 8 2 】

一方、特定 O S である場合 (ステップ S 2 3 : Y) には、C P U 1 0 は、S C チェーンテーブル 5 0 を参照し (ステップ S 2 4) 、S C チェーンテーブル 5 0 から各ストレージノード 4 のつながりを示すチェーン情報を取得する (ステップ S 2 5) 。

【 0 0 8 3 】

次いで、C P U 1 0 は、ノードペア管理テーブル 3 0 を参照し (ステップ S 2 6) 、ノードペア管理テーブル 3 0 からノードペア情報、すなわち、ストレージ制御部ペア 7 5 を構成する他のストレージ制御部 7 4 を格納するストレージノード 4 の情報を取得する (ステップ S 2 7) 。

【 0 0 8 4 】

次いで、C P U 1 0 は、L U のパスとする所定数のストレージ制御部 7 4 のあるストレージノード 4 (対象ストレージノード) を特定する (ステップ S 2 8) 。L U のパスとする所定数のストレージ制御部 7 4 は、ステップ S 1 3 , S 1 4 , S 1 5 での判定対象に該当するストレージ制御部 7 4 であってもよい。

【 0 0 8 5 】

次いで、C P U 1 0 は、パス優先度管理テーブル 2 0 に対象ストレージノードのターゲット情報を登録する (ステップ S 2 9) 。すなわち、C P U 1 0 は、パス優先度管理テーブル 2 0 に、作成した L U に対応するボリュームのエントリを作成し、このエントリに対象ストレージノード 4 のターゲット情報 (ターゲット番号) を格納する。

【 0 0 8 6 】

次いで、C P U 1 0 は、対象ストレージノードに対して、パス更新情報を送信する (ステップ S 3 0) 。ここで、パス更新情報には、作成した L U の L U N と、対応するボリュームのボリューム番号と、ステップ S 2 9 で登録したターゲット情報とが含まれる。

【 0 0 8 7 】

このボリュームパス作成処理によると、新たに作成した L U のパスの情報を、パスとする

10

20

30

40

50

ターゲットを持つストレージノード4に適切に通知することができ、後述するボリュームパス追加処理及びCTL Mテーブル生成処理を経て、ストレージノード4によりコンピュータノード2に対して問い合わせが来た場合に、適切にコンピュータノード2に対してパス優先度管理テーブル20に登録したパスを応答することができるようになる。

【0088】

次に、ボリュームパス追加処理について説明する。

【0089】

図11は、一実施形態に係るボリュームパス追加処理のフローチャートである。

【0090】

ボリュームパス追加処理は、ボリュームパス作成処理においてパス更新情報が送信される対象ストレージノードで実行される処理である。

10

【0091】

まず、対象ストレージノード4のCPU10は、パス更新情報を、ボリュームパス作成処理を実行したストレージ制御部74、すなわち、LUに対するアクティブのストレージ制御部74を有するストレージノード4から受信する(ステップS41)。

【0092】

次いで、ストレージノード4のCPU10は、パス更新情報に含まれている、LUNとボリューム番号とを、CL Mテーブル40のLUN管理テーブル42に追加する(ステップS42)。

【0093】

次いで、ストレージノード4のCPU10は、パス優先度管理テーブル20に、パス更新情報に基づくターゲット情報を設定する(ステップS43)。

20

【0094】

このボリュームパス追加処理によると、新たに追加されたLU(ボリューム)に対するパスとなるターゲットの情報を適切にパス優先度管理テーブル20に設定することができる。

【0095】

図12は、一実施形態に係るCTL Mテーブル生成処理のフローチャートである。

【0096】

CTL Mテーブル生成処理は、例えば、パス優先度選択処理、ボリュームパス作成処理、又はボリュームパス追加処理によってパス優先度管理テーブル20が作成または更新された後にストレージノード4において実行される処理である。

30

【0097】

ストレージノード4のCPU10は、コンピュータノード2の情報、例えば、OSの種類の情報を取得する(ステップS51)。なお、コンピュータノード2の情報の取得元としては、コンピュータノード2自体でもよく、コンピュータノード2の情報を収集している別の装置であってもよい。

【0098】

次いで、ストレージノード4のCPU10は、コンピュータノード2のOSが特定OS(設定可能なLUのパスが所定数(例えば、4つ)以下であるOS)であるか否かを判定する(ステップS52)。この結果、コンピュータノード2のOSが特定OSでない場合(ステップS52:N)には、CPU10は、処理を終了する。

40

【0099】

一方、コンピュータノード2のOSが特定OSである場合(ステップS52:Y)には、CPU10は、CL Mテーブル40を取得し(ステップS53)、パス優先度管理テーブル20を取得し(ステップS54)、CL Mテーブル40からLUN43に登録されているLUNに対応するボリューム番号を取得し、取得したボリューム番号に対応するパス優先度管理テーブル20のエントリを特定して、そのエントリのターゲット情報(フィールド23~26のターゲット番号)を取得する(ステップS55)。

【0100】

次いで、CPU10は、ターゲット情報に含まれているターゲットが、自身のストレージ

50

ノード４のターゲット７６であるか否かを判定し（ステップＳ５６）、自身のストレージノード４のターゲット７６でない場合（ステップＳ５６：Ｎ）には、処理を終了する。

【０１０１】

一方、自身のストレージノード４のターゲット７６である場合（ステップＳ５６：Ｙ）には、ＣＰＵ１０は、ＣＴＬＭテーブル６０に、コンピュータノードの番号と、ターゲット番号と、ＬＵＮと、ボリューム番号とを含むエントリを追加し（ステップＳ５７）、処理を終了する。なお、ステップＳ５３～Ｓ５７の処理は、ＣＬＭテーブル４０の各エントリを対象として実行される。

【０１０２】

以上説明したように、ＣＴＬＭテーブル生成処理によると、パス優先度管理テーブル２０に設定されたターゲットをターゲットとするパスのうち、各ストレージノード４に係するパスに対応するエントリを含むＣＴＬＭテーブル６０を作成することができる。すなわち、ＣＴＬＭテーブル６０には、パス優先度管理テーブル２０に設定されたパスのみが登録されることとなる。したがって、コンピュータノード２が各ストレージノード４に或るＬＵＮに対するパスを問い合わせると、パス優先度管理テーブル２０に登録されたターゲットのストレージノード４からのみ使用可能なパスとしての応答が得られることとなり、コンピュータノード２側では、優先度の高いパスを適切に把握して設定することが可能となる。

10

【０１０３】

なお、本発明は、上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲で、適宜変形して実施することが可能である。

20

【０１０４】

例えば、上記実施形態では、設定可能なＬＵのパスが４つ以下であるＯＳを例に、優先するパスを決定するようにしていたが、例えば、設定可能なＬＵのパスの数が４以外のＯＳにおいても同様に優先するパスを決定するようにしてもよい。また、設定可能なＬＵのパスの数が異なる複数のＯＳが混在する場合には、ＯＳごとに、優先するパスの数が異なるパス優先度管理テーブルを備え、それぞれに対応するＣＴＬＭテーブルを作成するようにしてもよい。

【０１０５】

また、上記実施形態において、ＣＰＵが行っていた処理の一部又は全部を、ハードウェア回路で行うようにしてもよい。また、上記実施形態におけるプログラムは、プログラムソースからインストールされてよい。プログラムソースは、プログラム配布サーバ又は記憶メディア（例えば可搬型の記憶メディア）であってもよい。

30

【符号の説明】

【０１０６】

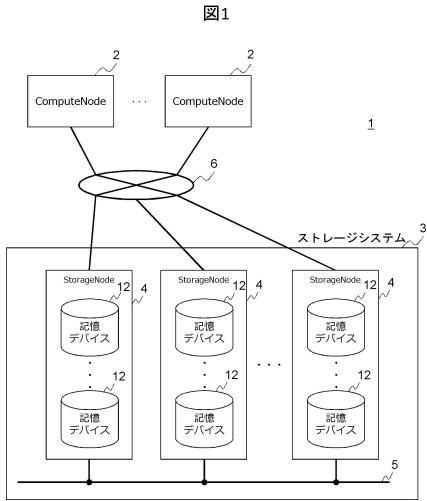
１…計算機システム、２…コンピュータノード、３…ストレージシステム、４…ストレージノード、５…バックエンドネットワーク、６…ネットワーク、１０…ＣＰＵ、１１…メモリ、８０…パス管理プログラム

40

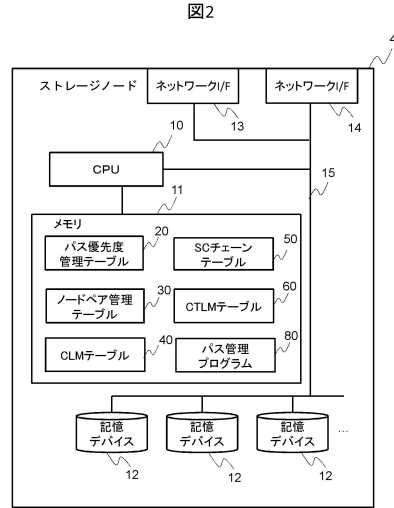
50

【図面】

【図1】

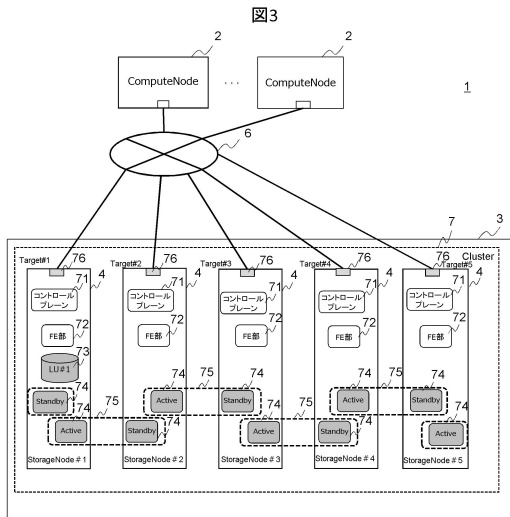


【図2】



10

【図3】



【図4】

図4

CTLMTテーブル

Computenode#	Target#	LUN管理テーブル	
		LUN	Volume#
1	0	0	0
		1	1
		2	2
		3	3
		4	4
		5	5
		6	6
3	8	8	8
	
		N	N
2	4	0	0
		1	1
		2	2

20

30

40

50

【 図 5 】

図5

バス優先度管理テーブル

Volume#0	StorageNode#	Priority target#0	Second priority target#0	Third priority target#0	Fourth priority target#0
⋮	⋮	⋮	⋮	⋮	⋮
Volume#M	StorageNode#	Priority target#M	Second priority target#M	Third priority target#M	Fourth priority target#M

【 図 6 】

図6

CLMテーブル

Computenode#	LUN管理テーブル	
	LUN	Volume#
1	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
	7	7
	8	8

N	N	
2	0	0
	1	1
	2	2

10

【 図 7 】

図7

ノードペア管理テーブル

StorageControllerPair#1	ActiveStorageNode#	StandbyStorageNode#
...
StorageControllerPair#M	ActiveStorageNode#	StandbyStorageNode#

【 図 8 】

図8

Storage制御部(SC)チェーンテーブル

StorageNode#	Active SC#	StandbySC# ※SecondActiveSC#	SecondStandbySC# ※ThirdActiveSC#	ThirdStandbySC#
...
StorageNode#	ActiveSC#	StandbySC# ※SecondActiveSC#	SecondStandbySC# ※ThirdActiveSC#	ThirdStandbySC#

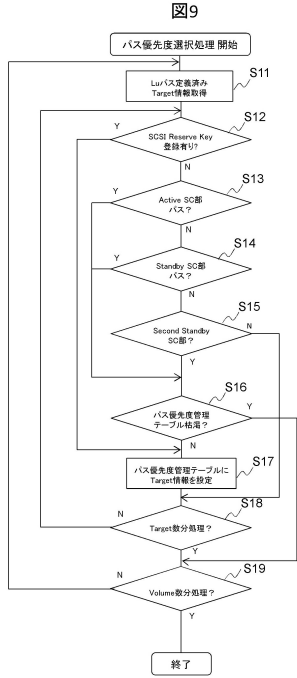
20

30

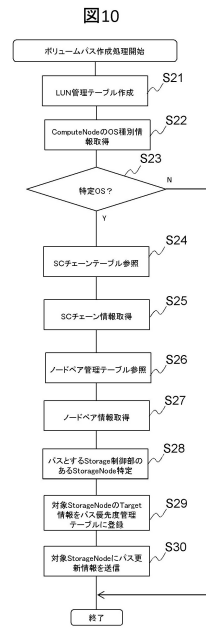
40

50

【 図 9 】



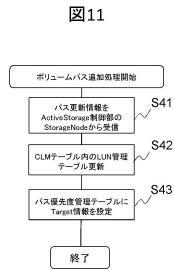
【 図 10 】



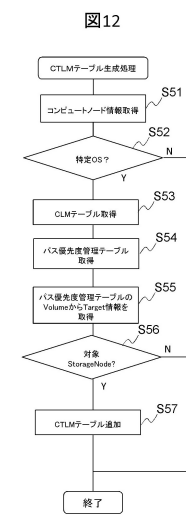
10

20

【 図 11 】



【 図 12 】



30

40

50

フロントページの続き

審査官 松平 英

- (56)参考文献 特表2012-531654(JP,A)
特開2018-156144(JP,A)
特開2019-101700(JP,A)
特開2007-034438(JP,A)
特開2012-208896(JP,A)
特開2019-125075(JP,A)
特開2016-157177(JP,A)
特開2014-146180(JP,A)
特開2005-310046(JP,A)
特開2015-158702(JP,A)
- (58)調査した分野 (Int.Cl., DB名)
G06F 13/14
G06F 13/10