

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5193801号
(P5193801)

(45) 発行日 平成25年5月8日(2013.5.8)

(24) 登録日 平成25年2月8日(2013.2.8)

(51) Int. Cl. F I
G06F 13/10 (2006.01) G O 6 F 13/10 3 4 O A
G06F 3/06 (2006.01) G O 6 F 3/06 3 O 2 Z

請求項の数 8 (全 16 頁)

(21) 出願番号 特願2008-277600 (P2008-277600)
 (22) 出願日 平成20年10月29日(2008.10.29)
 (65) 公開番号 特開2010-108114 (P2010-108114A)
 (43) 公開日 平成22年5月13日(2010.5.13)
 審査請求日 平成23年3月3日(2011.3.3)

(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
 (74) 代理人 100100310
 弁理士 井上 学
 (74) 代理人 100098660
 弁理士 戸田 裕二
 (72) 発明者 西川 記史
 神奈川県川崎市麻生区王禅寺1099番地
 株式会社日立製作所システム開発研究所
 内
 審査官 横山 佳弘

最終頁に続く

(54) 【発明の名称】 ストレージシステムの性能向上又は管理方法、システム、装置及びプログラム

(57) 【特許請求の範囲】

【請求項1】

複数の仮想サーバを提供する計算機と、管理計算機と、複数のLUを提供する複数のストレージコントローラを備えたストレージシステムと、から構成する計算機システムであって、

前記計算機は、前記複数の仮想サーバ間でI/O処理性能の調整を行うための指標である、前記複数の仮想サーバの複数のI/O優先度を管理するI/O優先度管理情報を有し、

前記複数のLUは前記複数の仮想サーバの複数の仮想ディスクデータを格納し、
 前記ストレージシステムは、前記複数のLUに対するI/O要求を処理するストレージコントローラを前記複数のストレージコントローラから特定するLU関連設定情報を有し、

前記管理計算機は、前記計算機が有する前記I/O優先度管理情報に対応した仮想サーバ設定情報と、前記ストレージシステムが有するLU関連設定情報に対応するストレージシステム構成情報と、を有し、前記仮想サーバ設定情報と前記ストレージシステム構成情報とを参照することで、前記I/O優先度管理情報が示す複数の仮想サーバの複数のI/O優先度を元にして前記複数のLUの一つのI/O要求を処理するストレージコントローラを決定し、前記決定を元にしたストレージ設定要求を送信し、

前記ストレージシステムは、前記ストレージ設定要求を受信したら、LU関連設定情報を変更することで、前記ストレージ設定要求で指定された前記複数のLUの一つのI/O

10

20

要求を処理するストレージコントローラを変更する、
ことを特徴とした計算機システム。

【請求項 2】

請求項 1 記載の計算機システムであって、前記管理計算機は、前記複数の仮想ディスクデータと前記複数の L U との格納関係を示す仮想ディスクデータ管理情報を有することを特徴とした計算機システム。

【請求項 3】

請求項 2 記載の計算機システムであって、

前記管理計算機は、前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報とを参照することで前記複数の仮想サーバの一部に対応するストレージコントローラを前記複数のストレージコントローラから特定し、前記複数の仮想サーバの一部に対応するストレージコントローラに関する情報を表示することを特徴とした計算機システム。

10

【請求項 4】

請求項 2 または請求項 3 に記載の計算機システムであって、

前記管理サーバは、前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報と仮想サーバ設定情報とを参照することで、前記複数のコントローラの各々について対応する一つ以上の仮想サーバの I / O 優先度の合計値を計算し、当該計算結果を表示する、

ことを特徴とした計算機システム。

【請求項 5】

20

複数の仮想サーバを提供する計算機と、前記複数の仮想サーバの複数の仮想ディスクデータを格納する複数の L U を提供し、複数のストレージコントローラを備えたストレージシステムと、に接続された管理計算機の性能管理方法であって、

前記複数の仮想サーバ間で I / O 処理性能の調整を行うための指標である、前記複数の仮想サーバの複数の I / O 優先度を管理する前記計算機が有する I / O 優先度管理情報に対応する仮想サーバ設定情報を格納するステップと、

前記複数の L U に対する I / O 要求を処理するストレージコントローラを前記複数のストレージコントローラから特定する、前記ストレージシステムが有する L U 関連設定情報に対応するストレージシステム構成情報を格納するステップと、

前記仮想サーバ設定情報と前記ストレージシステム構成情報とを参照することで、前記 I / O 優先度管理情報が示す複数の仮想サーバの複数の I / O 優先度を元にして前記複数の L U の一つの I / O 要求を処理するストレージコントローラを決定し、前記決定を元にしたストレージ設定要求を送信するステップと、

30

を有することを特徴とした管理計算機の性能管理方法。

【請求項 6】

請求項 5 記載の性能管理方法であって、

前記複数の仮想ディスクデータと前記複数の L U との格納関係を示す仮想ディスクデータ管理情報を格納するステップを有することを特徴とした管理計算機の性能管理方法。

【請求項 7】

請求項 6 記載の性能管理方法であって、

前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報とを参照することで前記複数の仮想サーバの一部に対応するストレージコントローラを前記複数のストレージコントローラから特定するステップと、

40

前記複数の仮想サーバの一部に対応するストレージコントローラに関する情報を表示するステップと、

を有することを特徴とした管理計算機の性能管理方法。

【請求項 8】

請求項 6 または請求項 7 に記載の性能管理方法であって、

前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報と仮想サーバ設定情報とを参照することで、前記複数のコントローラの各々について対応する一つ以上の仮

50

想サーバの I / O 優先度の合計値を計算し、当該計算結果を表示するステップを有することを特徴とした管理計算機の性能管理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ストレージシステムの性能向上又は管理方法、システム、装置及びプログラムに関する。

【背景技術】

【0002】

Virtual Machine (以後 VM と省略することがある) 技術は、サーバ計算機の CPU、主記憶、ネットワーク I / F、及びストレージ I / F を用い、仮想的に複数の計算機を提供する技術である。当該技術によって提供される仮想的な計算機を Virtual machine または仮想サーバと呼ぶ。

【0003】

Virtual machine 技術を適用したサーバ計算機の場合、同計算機の有するストレージ I / F (以後、I / O アダプタと呼ぶことがある) を複数の仮想サーバで共有するため、I / O 性能が問題となっていた。

【0004】

特許技術 1 に開示の技術では仮想サーバ毎に仮想的なストレージ I / F とデータキューを割り当てることで、仮想サーバで消費するデータ帯域を制御する技術が開示されている。

【0005】

【特許文献 1】米国特許公開明細書 2008 / 002704

【発明の開示】

【発明が解決しようとする課題】

【0006】

特許文献 1 に記載の技術をストレージ I / O に適用した場合、ストレージシステムは仮想サーバの設定とは無関係にストレージ I / O の処理を行うため、計算機システム全体としての性能管理が困難である。

【課題を解決するための手段】

【0007】

本発明は、複数の仮想サーバを提供する計算機と、管理計算機と、複数の LU を提供する複数のストレージコントローラを備えたストレージシステムと、から構成する計算機システムの性能管理に関する方法、装置、システムプログラム、記憶メディアを提供する。

【0008】

本発明の一実施例によると、仮想サーバ毎に仮想サーバに関する I / O 優先度に従って仮想サーバの I / O 処理を制御することで、仮想サーバ間の I / O 処理性能の調整を行う計算機と、ストレージシステム 2 の設定を連携させ、又は連携設定させるための仮想サーバに対応したストレージシステムの情報を表示する。

【0009】

また、本発明の他の実施例によると、前記計算機は前記複数の仮想サーバの複数の I / O 優先度を管理する I / O 優先度管理情報を有し、複数の LU は前記複数の仮想サーバの複数の仮想ディスクデータを格納し、前記ストレージシステムは、前記複数の LU に対する I / O 要求を処理するストレージコントローラを前記複数のストレージコントローラから特定する LU 関連設定情報を有し、前記管理計算機は、前記計算機が有する前記 I / O 優先度管理情報に対応した仮想サーバ設定情報と、前記ストレージシステムが有する LU 関連設定情報に対応するストレージシステム構成情報と、を有し、前記仮想サーバ設定情報と前記ストレージシステム構成情報とを参照することで、前記 I / O 優先度管理情報が示す複数の仮想サーバの複数の I / O 優先度を元にしたストレージシステムの構成を計算し、前記構成を元にしたストレージ設定要求を送信し、前記ストレージシステムは、前記

10

20

30

40

50

ストレージ設定要求を受信したら、LU関連設定情報を変更することで、前記ストレージ設定要求で指定された前記複数のLUの一つのI/O要求を処理するストレージコントローラを変更する。

【0010】

なお、前記管理計算機は、前記複数の仮想ディスクデータと前記複数のLUとの格納関係を示す仮想ディスクデータ管理情報を有してもよく、前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報とを参照することで前記複数の仮想サーバの一部に対応するストレージコントローラを前記複数のストレージコントローラから特定し、前記複数の仮想サーバの一部に対応するストレージコントローラに関する情報を表示してもよく、前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報と仮想サーバ設定情報とを参照することで、前記複数のコントローラの各々について対応する一つ以上の仮想サーバのI/O優先度の合計値を計算し、当該計算結果を表示してもよい。

また、本発明の他の実施例によると、前記複数の仮想サーバの複数のI/O優先度を管理する前記計算機が有するI/O優先度管理情報に対応する仮想サーバ設定情報を格納するステップと、前記複数のLUに対するI/O要求を処理するストレージコントローラを前記複数のストレージコントローラから特定する、前記ストレージシステムが有するLU関連設定情報に対応するストレージシステム構成情報を格納するステップと、前記仮想サーバ設定情報と前記ストレージシステム構成情報とを参照することで、前記I/O優先度管理情報が示す複数の仮想サーバの複数のI/O優先度を元にしたストレージシステムの構成を計算し、前記構成を元にしたストレージ設定要求を送信するステップとを有する。

【0011】

なお、前記複数の仮想ディスクデータと前記複数のLUとの格納関係を示す仮想ディスクデータ管理情報を格納するステップを有してもよく、前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報とを参照することで前記複数の仮想サーバの一部に対応するストレージコントローラを前記複数のストレージコントローラから特定するステップと、前記複数の仮想サーバの一部に対応するストレージコントローラに関する情報を表示するステップとを有してもよく、前記仮想ディスクデータ管理情報と前記ストレージシステム構成情報と仮想サーバ設定情報とを参照することで、前記複数のコントローラの各々について対応する一つ以上の仮想サーバのI/O優先度の合計値を計算し、当該計算結果を表示するステップを有してもよい。

【発明の効果】

【0012】

本発明により、仮想サーバを提供するサーバ計算機とストレージシステムの両方を考慮した性能管理が可能となる。

【発明を実施するための最良の形態】

【0013】

以下に、本発明の実施の形態を説明する。

【0014】

図1は、本発明が実施される計算機システムの構成例を示すものである。該計算機システムは、サーバ計算機（以後、サーバと省略することがある）1、ストレージシステム2、管理サーバ計算機（以後、管理サーバと省略することがある）28から構成される。また、サーバ1とストレージシステム2は、サーバ1のデータインタフェース19とストレージシステムのポート21を接続する第一のネットワーク3によって、接続される。また、サーバ1のネットワークインタフェース13とストレージシステム2のネットワークインタフェース27、管理サーバ28のネットワークインタフェース13は第二のネットワーク4に接続することで、サーバ1と管理サーバ28とストレージシステム2との通信を実現している。

【0015】

サーバ1は、CPU11a、主記憶12a、ネットワークインタフェース13a、入力装置14a、記憶装置18a、ストレージインターフェース（I/F）19aから構成

10

20

30

40

50

されている。なお、サーバ1 aはCD-ROM 16 aを有してもよく、これ以外の部品を含んでも良い。なお、入出力装置14 aの例としてはディスプレイとキーボードとポインタデバイスが考えられるが、これ以外の装置であってもよい。また、入出力装置の代替としてシリアルインターフェースやイーサネットインターフェースを入出力装置とし、当該インターフェースにディスプレイ又はキーボード又はポインタデバイスを有する表示用計算機を接続し、表示用情報を表示用計算機に送信したり、入力用情報を表示用計算機から受信することで、表示用計算機で表示を行ったり、入力を受け付けることで入出力装置14での入力及び表示を代替してもよい。

【0016】

ストレージシステム2は、ポート21、ストレージコントローラ22、制御メモリ23、プロセッサ24、キャッシュメモリ25、記憶装置26、ネットワークインタフェース27から構成されている。

【0017】

なお、ストレージコントローラ22は複数の記憶装置26の記憶領域を用いて複数のLU29をサーバ1 a(またはサーバ1 aが提供する仮想サーバ)に提供する。なお、LU29とはサーバ1 aからアクセス可能な記憶領域であり、Fibre ChannelやiSCSIなどでLUNと呼ばれる識別子が割り当てられている。サーバ1 aはLUNによってLU29を指定することでリード要求やライト要求に代表されるI/O要求をストレージシステム2へ送信する。

【0018】

LU29の一部は仮想サーバに対応した仮想ディスクデータ104を格納する。仮想ディスクデータ104とは、仮想サーバが仮想的に提供するハードウェア構成(例えば、CPUの数やネットワークI/Fの数やメモリ量)の情報を記憶したり、仮想サーバによって起動されるOSのファイルや仮想サーバのユーザが用いるファイルが格納される。なお、仮想ディスクにはこれら以外の情報、例えば、仮想サーバが提供するメモリ空間の内容を格納してもよい。さらには、仮想ディスクは一つ以上のファイルから構成されていてもよい。

【0019】

管理サーバ28は、CPU11 b、主記憶12 b、ネットワークインタフェース13 b、入出力装置14 b、記憶装置18 bから構成されている。なお、管理サーバ28はCD-ROM 16 bを有してもよく、これ以外の部品を含んでも良い。なお、入出力装置14 bの例としてはディスプレイとキーボードとポインタデバイスが考えられるが、これ以外の装置であってもよい。また、入出力装置の代替としてシリアルインターフェースやイーサネットインターフェースを入出力装置とし、当該インターフェースにディスプレイ又はキーボード又はポインタデバイスを有する表示用計算機を接続し、表示用情報を表示用計算機に送信したり、入力用情報を表示用計算機から受信することで、表示用計算機で表示を行ったり、入力を受け付けることで入出力装置14での入力及び表示を代替してもよい。

サーバ1 aの記憶装置18には、仮想サーバ管理プログラム103、及び仮想ディスクデータI/O優先度管理情報105が格納されている。仮想サーバ管理プログラム103はCPU11により主記憶12に読み込まれて実行される。なお、仮想サーバ管理プログラム103はCD-ROM 16を経由して記憶装置18に格納されてもよい。

【0020】

記憶装置サブシステム2の制御メモリ23には、LU-コントローラ対応関係変更プログラム102、性能情報収集プログラム108、LU-コントローラ-記憶装置対応関係情報101が格納される。

【0021】

管理サーバ28の記憶装置18には、LU-コントローラ対応関係変更制御プログラム106、LU-記憶装置対応関係変更制御プログラム108、仮想ディスク情報管理情報107が記憶される。LU-コントローラ対応関係変更制御プログラム106及びLU-

10

20

30

40

50

記憶装置対応関係変更制御プログラム108は、管理サーバ28のCPU11により管理サーバ28の主記憶12に読み込まれて実行される。また、LU-コントローラ対応関係変更制御プログラム106及びLU-記憶装置対応関係変更制御プログラム108はCD-ROM16を経由して管理サーバ28の記憶装置18に格納される。

【0022】

また、記憶装置26には仮想サーバ管理プログラム103により管理され、サーバ1のCPU11により主記憶12に読み込まれて実行される仮想サーバが使用する仮想ディスクデータ104を格納するLU29が配置される。

【0023】

なお、図1中のCD-ROM16はプログラムを媒体から読み込むための手段であり、フロッピー（登録商標）ディスクやテープ等他の媒体であってもかまわない。また、仮想サーバ管理プログラム103は、ネットワークインタフェース13経由で記憶装置18に格納されても良い。

【0024】

さらに、仮想サーバ管理プログラム103や仮想ディスクデータI/O優先度管理情報105は、ストレージシステム2の記憶装置26上のLU29に配置されてもよい。

【0025】

なお、サーバ1aは一つでも良いが、複数であっても良い。さらに、ストレージシステム2は仮想サーバを提供しない計算機が接続され、LU29に対してリード又はライトを行っても良い。このときのアクセス対象のLU29は仮想ディスクデータ104を含むものであってもよく、含まないLUであってもよい。

【0026】

次に本発明の概要を説明する。なお、概要に説明の無い事項について権利を放棄するものではない。

【0027】

本発明はサーバ1aが提供する仮想サーバ毎に、I/O優先度に従って仮想サーバのI/O処理の処理を制御することで、仮想サーバ間のI/O処理性能の調整を行う仮想サーバ管理プログラム103を有するサーバ1aと、ストレージシステム2とが連携することで、計算機システムとして性能管理を容易にする。管理サーバ28はサーバ1aとストレージシステム2とを連携させる計算機である。

【0028】

即ち、サーバ1aが提供する仮想サーバのI/O優先度についての情報を取得し、当該取得した情報を基に、ストレージシステム2の性能に関連する設定要求（ストレージ設定要求）を生成し、当該設定要求をストレージシステム2に送信し、当該設定要求を受信したストレージシステムは当該設定要求に従って、ストレージコントローラ22のLUへのI/O要求の性能に関する所定の設定情報（以後、LU関連設定情報と呼ぶ）を変更する。以後、ストレージシステム2は変更後のLU関連設定情報（その一例がLU-コントローラ-記憶装置対応関係管理情報）を元に、仮想サーバから受信したI/O要求を処理する。このように仮想サーバのI/O優先度に従って、ストレージシステム2のLU関連設定を変更するの処理設定を変更することで、計算機システムの性能管理を容易にする。

【0029】

特に、仮想サーバを提供するサーバ1aとストレージシステム2の連携は下記ケースを含む状況においてより性能管理が容易になる。

【0030】

（ケース1）ある仮想サーバ（区別のため、VS-Aと命名）のI/O優先度Paよりも低いI/O優先度Pbに対応付けられた仮想サーバ（区別のため、VS-Bと命名）より、ライト要求を受信した場合。管理サーバの処理によって当該I/O優先度に伴ったストレージ資源（例えば、ストレージコントローラ22、キャッシュメモリ125、プロセッサ25やプロセッサ25でI/O要求を処理する割合を指す。）にて実行するため、過度なストレージ資源の割り当てが不要になり、仮想サーバVS-AのI/O要求をより多

10

20

30

40

50

くの資源で処理することができる。

【 0 0 3 1 】

(ケース 2) 仮想サーバ V S - B の I / O 優先度 P b よりも高い I / O 優先度 P a に対応付けられた仮想サーバ V S - A よりリード要求を受信した場合。当該リード要求に応じたリードデータの送信処理に過小なストレージ資源を用いてしまうことで V S - B の I / O 要求の性能が良いという I / O 優先度とは逆の I / O 性能となってしまうことを回避できる。

【 0 0 3 2 】

(ケース 3) 仮想サーバを提供しない非仮想化サーバ計算機がストレージシステム 2 に接続され、ストレージシステム 2 が当該計算機からの I / O 要求を処理する場合。仮想サーバの I / O 優先度に従って、非仮想化サーバ計算機に対応する I / O 要求の処理を別なストレージコントローラや記憶装置で実行させることで、仮想サーバの I / O 性能低下を回避することができる。

【 0 0 3 3 】

(ケース 4) ストレージシステム 2 ではサーバ計算機からのリード要求に対応するリード処理及びライト要求に対応するライト処理とは別な処理が動作する場合。この場合は、ストレージシステム 2 は当該別な処理をどのストレージコントローラで実行するかを示した管理情報を有し、当該情報は仮想サーバの I / O 優先度に従って変更される。なお、この管理情報は直接的には L U に関係していないものの、当該処理を I / O 優先度に従って制御する場合は広義の L U 関連設定情報と見なしても良い。このように、前記別な処理を別なストレージコントローラや記憶装置で実行させることで、仮想サーバの I / O 性能低下を回避することができる。なお前記別な処理の一例としては、R A I D 技術によるパリティ計算処理や R A I D 技術適用時に記憶装置 2 6 に障害が発生した場合のリカバリ処理や、ストレージシステム 2 による記憶装置 2 6 に格納したデータの複製処理 (リモートコピーやローカルコピー) や、記憶装置 2 6 のフォーマットやデータ削除処理、その他リード処理やライト処理とは非同期に実行されるストレージコントローラの処理が考えられる。

【 0 0 3 4 】

または、管理サーバ 2 8 は、仮想サーバの I / O 優先度と当該設定用要求による L U 関連設定情報を管理サーバ 2 8 のユーザに表示する。表示のため、管理サーバ 2 8 は、仮想ディスクデータ I / O 優先度管理情報 1 0 5 と仮想ディスク情報管理情報とを参照することで表示対象の仮想サーバに対応した L U (仮想サーバに対応した仮想ディスクデータ 1 0 4 が格納されている) を特定し、特定した L U が関係するストレージシステムの処理の設定を表示することが考えられる。これによって仮想サーバ 2 8 のユーザは容易に仮想サーバとストレージシステムの設定を確認することが出来る。

【 0 0 3 5 】

なお、以後の実施例では L U 関連情報がされた場合の追従処理として、以下が説明されている。

(追従処理 1) 選択された L U の I / O 処理を行うストレージコントローラを別なストレージコントローラに変更し、以後は当該別なストレージコントローラで選択された L U の I / O 処理を行う。これによって I / O 優先度が高くなった仮想ディスクデータに対応する L U に対する I / O 要求がストレージコントローラが原因で性能低下 (特に I O P S が高い場合) することを回避できる。

(追従処理 2) 選択された L U のデータを現在格納されている記憶装置から別な記憶装置へ移動する。これによって I / O 優先度が高くなった仮想ディスクデータに対応する L U に対する I / O 要求が記憶装置が原因の性能低下 (特にスループットが高い場合や、キャッシュ効果が強くない環境下での I O P S が高い場合) を回避できる。

【 0 0 3 6 】

なお、前記追従処理は I / O 優先度が高くなった仮想ディスクデータに対応する L U に対する I / O 要求を調節可能な処理であれば上記処理以外でもよい。

10

20

30

40

50

【 0 0 3 7 】

また、本発明では説明の容易化として複数の記憶装置上にLUを定義するものとした。しかし、RAID技術を導入した場合、複数の記憶装置をまとめてパリティグループとし、当該パリティグループに対してLUを定義することも考えられる。この場合、記憶装置をパリティグループと読み替えて各種処理または情報を管理してもよい。

【 0 0 3 8 】

さらに、本発明では仮想ディスクデータに対してI/O優先度が対応しているが、仮想サーバに対してI/O優先度を割り当てる場合にも対応可能である。その場合は仮想ディスクデータに対するI/O優先度と説明されている部分を仮想サーバに対するI/O優先度と読み替えればよい。

10

【 0 0 3 9 】

以上で、本発明の概要説明を終わる。

【 0 0 4 0 】

図2は、仮想ディスクデータI/O優先度管理情報105の例である。該管理情報105は、仮想ディスクデータの名前(または識別子)を記憶する仮想ディスクデータ名フィールド201、及び仮想ディスクデータへのI/Oの優先度であるI/O優先度を記憶するI/O優先度フィールド202から構成されるテーブルである。なお、仮想ディスクデータI/O優先度管理情報105は仮想ディスクデータに対するI/O優先度を格納できればテーブル以外のデータ構造であってもよい。

【 0 0 4 1 】

20

当該情報を元に、仮想サーバ管理プログラム103は、仮想サーバのOSが発行するリード要求またはライト要求を、優先度に従い送信または受信や要求に関する処理を行う。

【 0 0 4 2 】

図3は、仮想ディスク情報管理情報107の構成例である。該情報は、仮想ディスクデータの名前(または識別子)を記憶する仮想ディスクデータ名フィールド301、当該仮想ディスクデータを記憶するストレージシステムの名前(または識別子)を記憶するストレージ装置名フィールド302、当該仮想ディスクデータが記憶されるLUの番号(または識別子)を記憶するLUNフィールド303から構成されるテーブルである。なお、仮想ディスク情報管理情報107は、仮想ディスクデータが格納されたストレージシステム2とLU29が特定できればテーブル以外のデータ構造であってもよい。

30

【 0 0 4 3 】

図4は、LU-コントローラ-記憶装置対応関係管理情報101の構成例である。該情報は、LUの番号(または識別子)を記憶するLUNフィールド、当該LUに対するI/Oの制御を行うコントローラを示すコントローラ名フィールド402、及び当該LUが記憶される記憶装置を示す記憶装置フィールド403から構成されるテーブルである。なお、LU-コントローラ-記憶装置対応関係管理情報101は、LU毎にI/O処理を行うコントローラが特定できればテーブル以外のデータ構造であってもよい。

【 0 0 4 4 】

図9は、記憶装置及びLUの容量を管理する容量管理情報109の構成例である。該情報は、記憶領域の種別を示す領域種別フィールド901、記憶領域の名称を示す名称フィールド902、及び領域の容量を示す容量フィールド903から構成されるテーブルである。なお、当該情報はこれらフィールドが格納されていればテーブル以外のデータ構造であってもよい。

40

【 0 0 4 5 】

なお管理サーバは、ストレージシステム構成情報を主記憶12bまたは記憶装置18bに格納している。当該情報は、管理サーバが有してCPU11bによって実行されるストレージシステム情報取得プログラムが、LU関連設定情報取得要求をストレージシステム2へ送信し、当該要求を受信したストレージシステム2がLU-コントローラ-記憶装置対応関係管理情報101に格納したLUとコントローラの対応情報及びLUと記憶装置の対応情報を送信し、対応情報を受信したストレージシステム情報取得プログラムが、当該

50

対応情報を元にLUとコントローラの対応とLUと記憶装置との対応をストレージシステム構成情報に格納することで管理されている。

【0046】

なお管理サーバは、仮想サーバ設定情報を主記憶12bまたは記憶装置18bに格納している。当該情報は、管理サーバが有してCPU11bによって実行される仮想サーバ情報取得プログラムが、I/O優先度取得要求を一つ以上のサーバ1aへ送信し、当該要求を受信したサーバ1aが仮想ディスクデータI/O優先度管理情報105に格納した仮想ディスクデータに対するI/O優先度(仮想サーバに対するI/O優先度と読み替えても良い)を含むI/O優先度情報を送信し、当該情報を受信した仮想サーバ情報取得プログラムが、当該情報を元に仮想サーバに対するI/O優先度を仮想サーバ設定情報に格納す

10

【0047】

図5は、LU-コントローラ対応関係変更制御プログラム106で実施される処理の例である。以下、当該プログラムをCPU11bが実行することで行われる処理について記載する。なお、以下の処理は特に断らない限り、仮想サーバ設定情報と仮想ディスクデータ管理情報107と仮想ディスクデータ管理情報とを参照することで行われる。

【0048】

(ステップ501) 該プログラムは、まず仮想ディスク毎のI/O優先度を受信する。取得方法には、以下の2方法がある。

(方法1) 仮想サーバ管理プログラム103より取得する方法。サーバ1aにI/O優先度取得要求を送信し、当該要求を受信したサーバ1aは仮想サーバの識別子とI/O優先度を含むI/O優先度情報を送信することで最新の仮想サーバ設定情報を最新にし、当該情報を参照する。

20

(方法2) 管理者より受け取る方法。当該プログラムは入出力装置14bによって入力された仮想サーバに対するI/O優先度を受信する。また、管理者から仮想ディスク毎のI/O優先度を取得する場合には、該情報を仮想サーバ管理ソフト103に伝えるか否かも同時に受け取り、伝えると指定された場合は当該I/O優先度と当該I/O優先度に対応する仮想サーバの識別子を含むI/O優先度設定情報をサーバ1aに送信する。なお、I/O優先度設定情報を受信したサーバ1bは当該設定情報に基づいて仮想ディスクデータI/O優先度管理情報105を更新する。また、当該プログラムは仮想サーバ設定情報も

30

【0049】

なお、方法1や方法2に限らず、仮想サーバ管理プログラムはネットワークインターフェース13aまたは入出力装置14aより仮想サーバの識別子と対応するI/O優先度を含むI/O優先度設定情報を含むI/O優先度設定要求を受信し、I/O優先度設定要求を受信したサーバ1bは当該設定情報に基づいて仮想ディスクデータI/O優先度管理情報105が更新される。

【0050】

(ステップ502) 各仮想ディスクデータについて、仮想ディスクデータ管理情報107を参照することで、仮想ディスクデータが格納されているLUを特定する。

40

【0051】

(ステップ503、ステップ504) ステップ502で求めたLU毎に、特定元である仮想ディスクデータのI/O優先度を加算し、LU毎のI/O優先度の合計値を計算し計算結果をLU-I/O優先度対応関係管理情報(主記憶12bまたは記憶装置18bに保存されている)に格納し、当該情報をI/O優先度の降順にソートする。

【0052】

(ステップ505、ステップ506、507) ストレージコントローラ毎のI/O優先度の合計が他のストレージコントローラと比較して小さいストレージコントローラを選択し、情報の先頭のLUを当該ストレージコントローラに割り当てるものと判断し、情報の先頭行を削除する。これを当該情報からレコードがなくなるまで繰り返す。なお、判断結果

50

は判断結果情報として主記憶 1 2 b または記憶装置 1 8 b に格納される。

【 0 0 5 3 】

(ステップ 5 0 8) ステップ 5 0 5 の判断結果情報に従い、L U - コントローラ - 記憶装置対応関係管理情報 1 0 1 の L U とコントローラの対応関係を更新する。なお、仮想ディスク毎の I / O 優先度を管理者より受け取り、かつ仮想サーバ管理ソフトへの該情報の伝達も同時に受け取った場合には、仮想サーバ管理プログラム 1 0 3 に対して、仮想ディスク別 I / O 優先度管理情報 1 0 5 をステップ 5 0 1 で受け取った値に変更してもよい。

【 0 0 5 4 】

図 6 は、L U - 記憶装置対応関係変更制御プログラム 1 0 8 で実施される処理の例である。以下、当該プログラムを C P U 1 1 b が実行することで行われる処理について記載する。なお、以下の処理は特に断らない限り、仮想サーバ設定情報と仮想ディスクデータ管理情報 1 0 7 と仮想ディスクデータ管理情報とを参照することで行われる。

【 0 0 5 5 】

(ステップ 6 0 1) 当該プログラムは、まず仮想ディスクデータ毎の I / O 優先度を受信する。なお、受信方法はステップ 5 0 1 にて説明した方法と同じ方法が考えられる。

【 0 0 5 6 】

(ステップ 6 0 2) 当該プログラムは、各仮想ディスクデータについて、仮想ディスクデータ管理情報を参照することで、仮想ディスクデータが格納されている L U を特定する。

【 0 0 5 7 】

(ステップ 6 0 3) 当該プログラムは、ステップ 6 0 2 で特定した L U 毎に、特定元である仮想ディスクデータの I / O 優先度を加算し、L U 毎の I / O 優先度の合計値を計算する。

【 0 0 5 8 】

(ステップ 6 0 4) 当該プログラムは、ステップ 6 0 3 の計算結果を元に、L U - コントローラ - 記憶装置対応関係管理情報 1 0 1 を参照して L U が格納されている記憶装置を特定し、記憶装置毎に特定元となる L U の I / O 優先度を合計することで、記憶装置毎の I / O 優先度を計算する。

【 0 0 5 9 】

(ステップ 6 0 5) 当該プログラムは、各記憶装置の I / O 優先度のいずれかが閾値を超えているかどうかを判断し、超えている場合は、L U と記憶装置との対応を変えた場合についてステップ 6 0 4 の計算を行うことで、各記憶装置の I / O 優先度が閾値を超えない L U と記憶装置との対応を計算する。なお、ここでの計算では閾値を超えない対応関係が算出できる場合もあるが、算出できない場合もある。なお、当該計算には各 L U の容量、及び記憶装置の空き容量が必要となるが、これらは図 4 に示す L U - コントローラ - 記憶装置対応関係管理情報 1 0 1、及び容量管理情報 1 0 9 を参照して求める。なお、算出結果は算出結果情報として主記憶 1 2 b または記憶装置 1 8 b に格納する。

【 0 0 6 0 】

(ステップ 6 0 6) その後、現在の構成と算出結果情報とを元に、ステップ 6 0 5 にて新たな L U と記憶装置との対応を算出したと判断した場合は、その算出結果情報に従い、記憶装置との対応が変更された L U のデータを移動元の記憶装置から移動先の記憶装置へ移動する。そして、L U - コントローラ - 記憶装置対応関係管理情報 1 0 1 の L U と記憶装置の対応関係を更新する。なお、仮想ディスク毎の I / O 優先度を管理者より受け取り、かつ仮想サーバ管理ソフトへの該情報の伝達も同時に受け取った場合には、仮想サーバ管理プログラム 1 0 3 に対して、仮想ディスク別 I / O 優先度管理情報 1 0 5 をステップ 6 0 1 で受け取った値に変更してもよい。

【 0 0 6 1 】

以下、図 1 の構成において以下の状態であった場合の図 5 に説明したストレージコントローラ割当て変更時の動作について説明する。

(1) L U 2 9 a と L U 2 9 b とが記憶装置 2 6 a に対応付けられ、L U 2 9 c と L U 2

10

20

30

40

50

9 dとが記憶装置26 bに対応付けられている。

(2) サーバ1 aは仮想サーバ管理プログラム103を処理することで仮想サーバ1乃至4を提供する。

(3) 仮想ディスクデータ104 a(仮想ディスクデータ名としてVDK01が割り当てられている)を用いて仮想サーバ1は提供される。また、仮想サーバ1による仮想ディスクデータ104 aに対するI/O優先度は60である。仮想ディスクデータ104 aを格納したLU29 aのI/O処理はストレージコントローラ22 aにて行われる。

【0062】

(4) 仮想ディスクデータ104 b(仮想ディスクデータ名としてVDK02が割り当てられている)を用いて仮想サーバ2は提供される。また、仮想サーバ2による仮想ディスクデータ104 bに対するI/O優先度は50である。仮想ディスクデータ104 bを格納したLU29 bのI/O処理はストレージコントローラ22 aにて行われる。

10

(5) 仮想ディスクデータ104 c(仮想ディスクデータ名としてVDK03が割り当てられている)を用いて仮想サーバ3は提供される。また、仮想サーバ3による仮想ディスクデータ104 cに対するI/O優先度は30である。仮想ディスクデータ104 cを格納したLU29 cのI/O処理はストレージコントローラ22 bにて行われる。

(6) 仮想ディスクデータ104 d(仮想ディスクデータ名としてVDK04が割り当てられている)を用いて仮想サーバ4は提供される。また、仮想サーバ4による仮想ディスクデータ104 aに対するI/O優先度は60である。仮想ディスクデータ104 dを格納したLU29 dのI/O処理はストレージコントローラ22 bにて行われる。

20

【0063】

以上説明した状態において、サーバ1 aが受信したI/O優先度設定情報に基づいて仮想ディスクデータ104 aに対するI/O優先度が160に変更された場合、計算機システムは以下の動作を行う。なお、図7は変更後の仮想ディスクデータI/O優先度管理情報105である。

【0064】

(動作1) ステップ501とステップ502とステップ503を処理することで、LU-コントローラ対応関係変更制御プログラム106は、LU26 aのI/O優先度が160であることを算出し情報格納する。(一つのLUに複数の仮想ディスクデータが格納される場合は仮想ディスクデータそのままのI/O優先度にはならない)。

30

【0065】

(動作2) ステップ504とステップ505とステップ506を処理することで、LU-コントローラ対応関係変更制御プログラム106は、ストレージコントローラ22 aのI/O優先度を210と算出する(ストレージコントローラ22 bのI/O優先度は110のまま)。そして、ストレージコントローラ22 aに対応していたLUのうち、最もI/O優先度が低いLU29 bを特定する。

【0066】

(動作3) ステップ508の動作によって、管理サーバ28はストレージシステム2にLU-コントローラ対応変更要求を送信し、当該要求を受信したストレージシステム2はLU-コントローラ-記憶装置対応関係管理情報101を変更し、ストレージシステム2は当該情報101を参照することで、以後はLU29 bのI/O処理をコントローラ22 bで処理する。なお、図8は変更後のLU-コントローラ-記憶装置対応関係管理情報101である。また、LU-コントローラ対応変更要求はLUとストレージコントローラを指定するが、今回はLU29 bとストレージコントローラ22 bを指定した。

40

【0067】

以上示すように、本発明により仮想ディスクのI/O優先度に基づくコントローラの割当て変更が達成できることが確認された。

【0068】

以下、論理記憶割り当て変更時の動作について説明する。なお、例とする状況は上記コントローラ割り当て変更と同じである。なお、記憶装置割当て変更はデータ移動が伴うた

50

めに処理完了まで長時間を要する。そのため、本処理は記憶装置毎のI/O優先度の合計値の差が所定の値又は割合(例えば1.5倍等)生じた場合に行うようにしても良い。

【0069】

(動作1)ステップ601とステップ602とステップ603を処理することで、LU-コントローラ対応関係変更制御プログラム106は、LU26aのI/O優先度が160であることを算出し情報格納する。

【0070】

(動作2)ステップ604及びステップ605を処理することで、当該プログラムは記憶装置29aのI/O優先度を210と算出する(記憶装置29bのI/O優先度は110のまま)。そして、記憶装置29aに対応していたLUのうち、最もI/O優先度が低いLU29bを特定する。

10

【0071】

(動作3)ステップ606の動作によって、管理サーバ28はストレージシステム2にLU-記憶装置対応変更要求を送信する。なお、LU-コントローラ対応変更要求はLUとストレージコントローラを指定するが、今回はLU29bとストレージコントローラ22bを指定する。当該要求を受信したストレージシステム2は指定されたLU22bのデータを記憶装置29aから記憶装置29bへ移動し、LU-コントローラ-記憶装置対応関係管理情報101を変更する。なお、図11は変更後のLU-コントローラ-記憶装置対応関係管理情報101である。

【0072】

20

以上示すように、本発明により仮想ディスクのI/O優先度に基づく記憶装置の割当て変更が達成できることが確認された。

次に本発明での管理サーバ28での管理情報表示について述べる。これまで説明してきたように、本発明では管理サーバが仮想サーバの仮想ディスクデータのI/O優先度(または仮想サーバのI/O優先度)の変更に応じてストレージ装置の設定変更を行う例について説明してきた。しかし、コントローラや記憶装置の性能限界からI/O処理の負荷移動先がない場合や、信頼性やコスト上の問題から一部のLUについてI/O処理の負荷移動が出来ない場合がある。また、管理サーバによる変更が妥当であるかどうかを確認したい場合もありえる。

【0073】

30

本発明ではこうした場合に対応するために管理サーバにて以下の情報を表示してもよい。

(表示1)現在の仮想ディスクデータ(または仮想サーバ)と当該仮想ディスクデータを格納するLUについてのI/O要求を処理するストレージコントローラの識別子。当該情報を表示するために、管理サーバは仮想ディスクデータ管理情報107と、ストレージシステム構成情報を参照する。

(表示2)現在の仮想ディスクデータのI/O優先度(または仮想サーバのI/O優先度)と当該仮想ディスクデータを格納するLUに対するI/Oを処理するストレージコントローラのステップ503で求めたようなI/O優先度の合計(またはその他ストレージコントローラの処理性能に関連した設定情報や構成情報)。当該情報を表示するために、管理サーバは表示1で参照した情報に加えて仮想サーバ設定情報を参照する。

40

(表示3)現在の仮想ディスクデータ(または仮想サーバ)と当該仮想ディスクデータを格納するLUに対応する記憶装置の識別子。当該情報を表示するために管理サーバが参照する情報は表示1と同じである。

(表示4)現在の仮想ディスクデータのI/O優先度(または仮想サーバのI/O優先度)と当該仮想ディスクデータを格納するLUに対するI/Oを処理するストレージコントローラのステップ503で求めたようなI/O優先度の合計。当該情報を表示するために、管理サーバは表示3で参照した情報に加えて仮想サーバ設定情報を参照する。

【0074】

なお、これまで説明してきた管理サーバ28のプログラム及び情報はサーバ1aに存在

50

し、実行を行ってもよい。同様に当該プログラム及び情報はストレージシステム 2 の制御メモリまたは管理メモリに格納し、ストレージシステムのストレージコントローラ 2 2 が実行してもよい。なお、その場合の表示や入力管理サーバがイーサネットポートやシリアルインターフェースを持つ場合と同様の方法で実現する。

【0075】

また、ストレージシステム 2 は一つの装置を前提として説明したが、複数のストレージシステム 2 (互いが接続されている) を有するストレージクラスタについても同様の考えが適用できる。すなわち、仮にストレージシステム A とストレージシステム B がストレージクラスタに含まれているとして、ストレージシステム A が有するストレージコントローラ A と記憶装置 A とポート A がストレージコントローラ 2 2 a と記憶装置 2 6 a とポート 2 1 a と考え、ストレージシステム B のストレージコントローラ B と記憶装置 B とポート B がストレージコントローラ 2 2 b と記憶装置 2 6 b とポート 2 1 b と考える。このような状況で仮想ディスクデータの I/O 優先度変更に応じて、LU のデータをストレージシステム A からストレージシステム B へ移動し、それに伴って、LU に対応するポートとストレージコントローラをポート B とストレージコントローラ B とすることで、同様のメリットが得られる。

10

【0076】

なお、ポートを切り替える場合はサーバ 1 a に対して切り替え要求を送信することで、サーバ 1 a が送信する I/O 要求の送信先をポート A からポート B に切り替えることで実現できる。なお、当該切り替えは仮想サーバ管理プログラムが処理することで、仮想サーバに対応する OS の設定の変更を行わずに実現してもよい。

20

【図面の簡単な説明】

【0077】

【図 1】本発明が実施される計算機システム及び記憶装置サブシステムの第一の構成例である。

【図 2】仮想ディスク別 I/O 優先度管理情報の構成例である。

【図 3】仮想ディスク情報管理情報の構成例である。

【図 4】LU - コントローラ - 記憶装置対応関係管理情報の構成例である。

【図 5】LU - コントローラ対応関係変更制御プログラムで実施される処理の例である。

【図 6】LU - 記憶装置対応関係変更制御プログラムで実施される処理の例である。

30

【図 7】仮想ディスク別 I/O 優先度管理情報の更新後の構成例である。

【図 8】LU - コントローラ - 記憶装置対応関係管理情報の第一の変更後の構成例である。

【図 9】容量管理情報の構成例である。

【図 10】LU - I/O 優先度対応関係管理情報の構成例である。

【図 11】LU - コントローラ - 記憶装置対応関係管理情報の第二の変更後の構成例である。

【符号の説明】

【0078】

1・・・サーバ

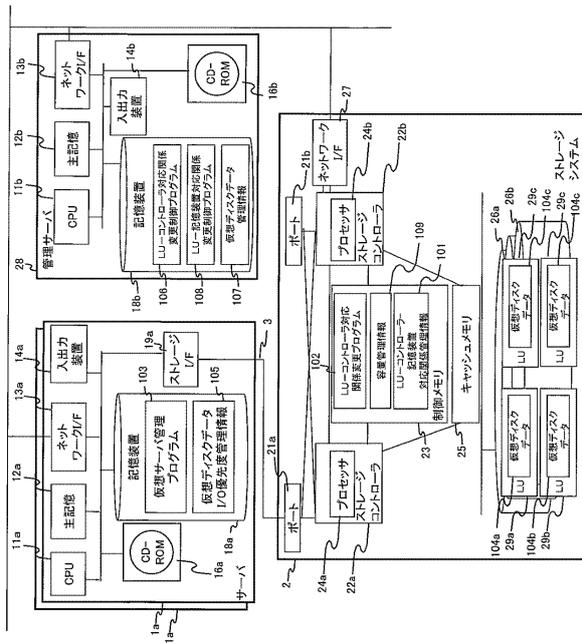
40

2・・・記憶装置サブシステム

28・・・管理サーバ

29・・・LU

【図1】



【図1】

【図2】

【図2】 仮想ディスクデータI/O優先度管理情報105

仮想ディスクデータ名	I/O優先度
VDK01	60
VDK02	50
VDK03	30
VDK04	80

【図3】

【図3】 仮想ディスクデータ管理情報

仮想ディスクデータ名	ストレージ装置名	LUN
VDK01	STORAGE01	1
VDK02	STORAGE01	2
VDK03	STORAGE01	3
VDK04	STORAGE01	4

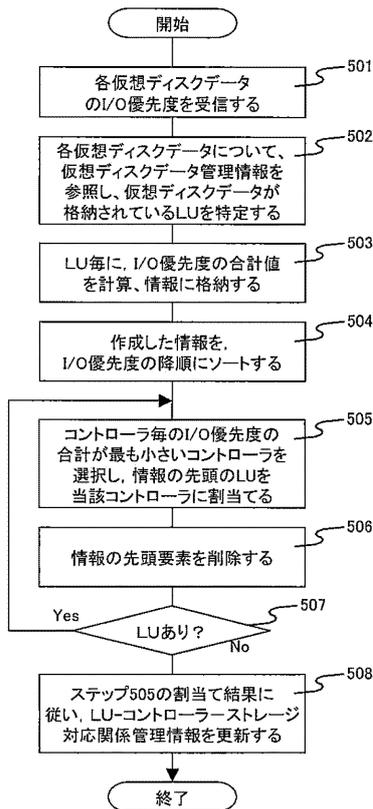
【図4】

【図4】 LU-コントローラ-記憶装置対応関係管理情報

LUN	コントローラ名	記憶装置
1	22a	26a
2	22a	26a
3	22b	26b
4	22b	26b

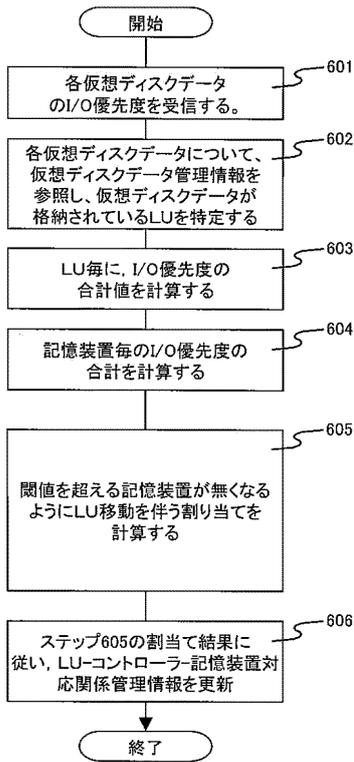
【図5】

【図5】 LU-コントローラ対応関係変更制御プログラム106で実施されるフローチャート



【図6】

【図6】 LU-記憶装置対応関係変更制御プログラム108で実施されるフローチャート



【図7】

【図7】 仮想ディスクデータI/O優先度管理情報105 (変更後)

仮想ディスクデータ名	I/O優先度
VDK01	160
VDK02	50
VDK03	30
VDK04	80

【図8】

【図8】 LU-コントローラ-記憶装置対応関係管理情報(変更後1)

LUN	コントローラ名	記憶装置
1	22a	26a
2	22b	26a
3	22b	26b
4	22b	26b

【図9】

【図9】 容量管理情報109

領域種別	名称	容量
記憶装置	26a	60GB
記憶装置	26b	60GB
LU	1	20GB
LU	2	20GB
LU	3	20GB
LU	4	20GB

【図11】

【図11】 LU-コントローラ-記憶装置対応関係管理情報(変更後2)

LUN	コントローラ名	記憶装置
1	22a	26a
2	22b	26b
3	22b	26b
4	22b	26b

【図10】

【図10】 LU-I/O優先度対応関係管理情報

LUN	I/O優先度
1	160
4	80
2	50
3	30

フロントページの続き

- (56)参考文献 特開2006-079495(JP,A)
特開2007-115019(JP,A)
特開2006-067401(JP,A)
特開2004-152060(JP,A)
特開2007-097222(JP,A)
特開2004-086512(JP,A)
特開2005-222123(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 13/10
G06F 3/06