

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5706531号
(P5706531)

(45) 発行日 平成27年4月22日(2015.4.22)

(24) 登録日 平成27年3月6日(2015.3.6)

(51) Int. Cl. F I
G06F 12/00 (2006.01) G06F 12/00 501B

請求項の数 15 (全 39 頁)

(21) 出願番号	特願2013-535687 (P2013-535687)	(73) 特許権者	000005108
(86) (22) 出願日	平成23年9月27日 (2011.9.27)		株式会社日立製作所
(86) 国際出願番号	PCT/JP2011/072058		東京都千代田区丸の内一丁目6番6号
(87) 国際公開番号	W02013/046331	(74) 代理人	100091096
(87) 国際公開日	平成25年4月4日 (2013.4.4)		弁理士 平木 祐輔
審査請求日	平成25年11月15日 (2013.11.15)	(74) 代理人	100105463
			弁理士 関谷 三男
		(74) 代理人	100102576
			弁理士 渡辺 敏章
		(72) 発明者	森 宣仁
			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
		(72) 発明者	岩崎 正明
			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】 計算機システム、及び情報管理方法

(57) 【特許請求の範囲】

【請求項1】

少なくとも1つのストレージサブシステムと、前記ストレージサブシステムに接続される複数の計算機と、を含む計算機システムであって、

前記ストレージサブシステムは、

レスポンス性能が異なる複数の物理記憶デバイスと、

前記物理記憶デバイスの実記憶領域を含むプールと、当該プールから構成される1つ以上の仮想論理ボリュームとを対応付ける仮想論理ボリューム管理情報とを格納するメモリと、

前記複数の計算機の何れかに、前記仮想論理ボリュームに含まれる実記憶領域を割り当てる処理を実行するプロセッサと、を有し、

前記複数の計算機の少なくとも1つの計算機は、

前記複数の計算機上で実行されるタスクから発行されるアクセスを監視するか否かを設定し、

前記タスクの前記実記憶領域に対するアクセス特性と前記監視に関する設定に基づいて、前記レスポンス性能が異なる物理記憶デバイスの実記憶領域間で移動するデータの移動元と移動先を含む再配置基準情報を設定し、

前記ストレージサブシステムのプロセッサは、

前記再配置基準情報を参照し、前記複数の計算機に割り当てられた前記実記憶領域に対する前記複数の計算機からのアクセス特性に基づいて、前記実記憶領域に格納されたデ

10

20

ータを、前記レスポンス性能が異なる物理記憶デバイスの別の実記憶領域に移動する再配置処理を実行する、
ことを特徴とする計算機システム。

【請求項 2】

請求項 1 において、

前記複数の計算機の少なくとも 1 つの計算機は、前記複数の計算機のいくつかで実行される前記タスクが使用する記憶領域毎のデータの再配置発生量を示す再配置発生情報を生成し、これを出力することを特徴とする計算機システム。

【請求項 3】

請求項 2 において、

前記少なくとも 1 つの計算機は、前記再配置発生情報に基づいて、前記再配置処理を要するタスクであることを示す再配置要否情報を決定し、当該再配置要否情報に基づいて、前記再配置基準情報を設定することを特徴とする計算機システム。

【請求項 4】

請求項 3 において、

前記少なくとも 1 つの計算機は、前記再配置発生情報の出力に応じたユーザによる指示入力に応答して、前記再配置要否情報を決定することを特徴とする計算機システム。

【請求項 5】

請求項 3 において、

前記少なくとも 1 つの計算機は、
複数のタスクについて前記再配置処理における移動先決定の判断基準から外すべき候補となりうるタスクであることを示す判断除外候補情報を保持し、
前記計算機毎の各タスクのデータの再配置発生量の割合を示す第 1 の再配置発生割合情報を算出し、
前記判断除外候補情報に含まれるタスクであって、前記第 1 の再配置発生割合情報が第 1 の閾値を超えるタスクを特定し、
前記特定されたタスクを前記再配置要否情報として出力する、
ことを特徴とする計算機システム。

【請求項 6】

請求項 3 において、

前記少なくとも 1 つの計算機は、
前記再配置処理を実行する複数の時間帯に亘って前記再配置発生情報を生成し、
連続する少なくとも 2 つの時間帯で、タスクを実行する全ての計算機に関する前記データの再配置発生量の合計値が第 2 の閾値を超過する連続する時間帯を特定し、
前記特定された連続する時間帯のそれぞれにおける、前記タスクを実行する全ての計算機に関する前記データの再配置発生量の割合を示す第 2 の再配置発生割合情報を算出し、
前記特定された連続する時間帯のそれぞれの前記第 2 の再配置発生割合情報の 1 つが第 3 の閾値を超過し、他の前記第 2 の配置発生割合情報が前記第 3 の閾値を下回るタスクを特定し、
前記特定されたタスクを前記再配置要否情報として出力する、
ことを特徴とする計算機システム。

【請求項 7】

請求項 1 において、

前記複数の計算機は、複数のホスト計算機と、少なくとも 1 つの管理計算機を含み、
前記再配置基準情報を設定する前記少なくとも 1 つの計算機は、管理計算機であることを特徴とする計算機システム。

【請求項 8】

請求項 1 において、

前記複数の計算機は、複数のホスト計算機と、少なくとも 1 つの管理計算機を含み、

10

20

30

40

50

前記複数のホスト計算機のそれぞれが、自身のホスト計算機に関する前記再配置基準情報を設定することを特徴とする計算機システム。

【請求項 9】

請求項 1 において、

前記少なくとも 1 つの計算機は、前記再配置基準情報を、前記仮想論理ボリューム単位、前記タスクを実行する計算機による前記ストレージサブシステムに対するアクセス単位、或いは、前記仮想論理ボリュームの前記実記憶領域に対するアクセス単位で設定することを特徴とする計算機システム。

【請求項 10】

請求項 2 において、

前記少なくとも 1 つの計算機は、前記レスポンス性能が所定値を示す物理記憶デバイスで構成される第 1 の階層から、当該第 1 の階層よりもレスポンス性能が高い物理記憶デバイスで構成される第 2 の階層に再配置するデータの量を、前記データの再配置発生量とすることを特徴とする計算機システム。

【請求項 11】

少なくとも 1 つのストレージサブシステムと、前記ストレージサブシステムに接続される複数の計算機と、を含む計算機システムにおける情報管理方法であって、

前記ストレージサブシステムは、レスポンス性能が異なる複数の物理記憶デバイスと、前記物理記憶デバイスの実記憶領域を含むプールと、当該プールから構成される 1 つ以上の仮想論理ボリュームとを対応付ける仮想論理ボリューム管理情報とを格納するメモリと、前記複数の計算機の何れかに、前記仮想論理ボリュームに含まれる実記憶領域を割り当てる処理を実行するプロセッサと、を有し、

前記情報管理方法は、

前記複数の計算機の少なくとも 1 つの計算機が、前記複数の計算機上で実行されるタスクから発行されるアクセスを監視するか否かを設定し、さらに、前記タスクの前記実記憶領域に対するアクセス特性と前記監視に関する設定に基づいて、前記レスポンス性能が異なる物理記憶デバイスの実記憶領域間で移動するデータの移動元と移動先を含む再配置基準情報を設定し、

前記ストレージサブシステムのプロセッサが、前記再配置基準情報を参照し、前記複数の計算機に割り当てられた前記実記憶領域に対する前記複数の計算機からのアクセス特性に基づいて、前記実記憶領域に格納されたデータを、前記レスポンス性能が異なる物理記憶デバイスの別の実記憶領域に移動する再配置処理を実行する、ことを特徴とする情報管理方法。

【請求項 12】

請求項 11 において、

さらに、前記複数の計算機の少なくとも 1 つの計算機が、前記複数の計算機のいくつかで実行される前記タスクが使用する記憶領域毎のデータの再配置発生量を示す再配置発生情報を生成し、これを出力することを特徴とする情報管理方法。

【請求項 13】

請求項 12 において、

さらに、前記少なくとも 1 つの計算機が、前記再配置発生情報に基づいて、前記再配置処理を要するタスクであることを示す再配置要否情報を決定し、当該再配置要否情報に基づいて、前記再配置基準情報を設定することを特徴とする情報管理方法。

【請求項 14】

請求項 13 において、

前記少なくとも 1 つの計算機は、複数のタスクについて前記再配置処理における移動先決定の判断基準から外すべき候補となりうるタスクであることを示す判断除外候補情報を保持し、

前記少なくとも 1 つの計算機が、さらに、

前記計算機毎の各タスクのデータの再配置発生量の割合を示す第 1 の再配置発生割合

10

20

30

40

50

情報を算出し、

前記判断除外候補情報に含まれるタスクであって、前記第1の再配置発生割合情報が第1の閾値を超えるタスクを特定し、

前記特定されたタスクを前記再配置要否情報として出力する、
ことを特徴とする情報管理方法。

【請求項15】

請求項13において、

前記少なくとも1つの計算機が、さらに、

前記再配置処理を実行する複数の時間帯に亘って前記再配置発生情報を生成し、

連続する少なくとも2つの時間帯で、タスクを実行する全ての計算機に関する前記データの再配置発生量の合計値が第2の閾値を超過する連続する時間帯を特定し、

前記特定された連続する時間帯のそれぞれにおける、前記タスクを実行する全ての計算機に関する前記データの再配置発生量の割合を示す第2の再配置発生割合情報を算出し、

前記特定された連続する時間帯のそれぞれの前記第2の再配置発生割合情報の1つが第3の閾値を超過し、他の前記第2の再配置発生割合情報が前記第3の閾値を下回るタスクを特定し、

前記特定されたタスクを前記再配置要否情報として出力する、
ことを特徴とする情報管理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、計算機システム、及び情報管理方法に関し、例えば、ストレージシステムにおけるストレージボリュームプールの階層を管理するための技術に関する。

【背景技術】

【0002】

近年、シンプロビジョニングと呼ばれる技術（例えば、特許文献1参照）がストレージシステムにおいて盛んに用いられてきている。ここで、シンプロビジョニングとは、仮想ボリューム（仮想論理ボリュームともいう）をホスト計算機に割り当てる技術である。また、仮想ボリュームとは、ホスト計算機から仮想ボリュームへデータの書き込みがあって初めて、プール化された論理ボリューム（以降、プールボリュームと呼ぶ）から、物理ディスク上の実データ保存領域（以降、ページと呼ぶ）が割り当てられるボリュームである。

【0003】

また、特許文献2に開示されているように、仮想ボリュームに割り当てられたページに対するホスト計算機からのI/O量に応じて、当該ページを、当該ページの格納されているプールボリュームから、性能やコストの異なる別のプールボリュームへ移行させる（再配置処理）ように運用されている。つまり、ストレージ装置は仮想ボリュームを構成する各ページに対するI/O数を記録し（以降、この記録の実行をI/Oモニタと呼ぶ）、I/O数に応じてページの配置先階層（性能およびコストで分類したプールボリューム）を決定する。例えば高I/O数のページから順に上位階層（高性能かつ高コストのプールボリューム）に配置する。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2003-15915号公報

【特許文献2】特開2007-66259号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、ストレージ装置は、実際にI/Oを発行しているホスト計算機上のタスクが何であるかを認識しないため、仮想ボリュームを構成する各ページに対する全I/Oを記録してしまう。このため、高レスポンス不要のタスクのI/O数が一時的に多くなる場合等に、高レスポンス要のタスクのデータを格納するページ群に対するI/O数が、高レスポンス不要のタスクのデータを格納するページ群に対するI/O数に比べて相対的に低くなり、結果として高レスポンス要のタスクのデータを格納するページ群が下位階層に配置されてしまい、性能が劣化する可能性があった。

【0006】

本発明はこのような状況に鑑みてなされたものであり、タスクの種類および稼働状態に応じて再配置処理における移動先階層を決定するための判断基準を適切に設定し、高レスポンス要のタスクのデータが下位階層に配置されて性能が劣化することを防ぐための技術を提供するものである。

10

【課題を解決するための手段】**【0007】**

上記課題を解決するために、本発明による計算機システムは、少なくとも1つのストレージサブシステム(ストレージ装置)と、ストレージサブシステムに接続される複数の計算機と、を有する。ここで、ストレージサブシステムは、レスポンス性能が異なる複数の物理記憶デバイスと、物理記憶デバイスの実記憶領域を含むプールと、当該プールから構成される1つ以上の仮想論理ボリュームとを対応付ける仮想ボリューム管理情報とを格納するメモリと、複数の計算機の何れかに、仮想論理ボリュームに含まれる実記憶領域を割り当てる処理を実行するプロセッサと、を有する。そして、複数の計算機の少なくとも1つの計算機(管理計算機或いは各ホスト計算機)は、複数の計算機上で実行されるタスクに関するアクセス特性を、レスポンス性能が異なる物理記憶デバイスの実記憶領域間でデータを移動する再配置処理における移動先決定の判断基準として考慮するか否かを示す再配置基準情報を設定する。また、ストレージサブシステムは、再配置基準情報を参照し、複数の計算機に割り当てられた実記憶領域に対する複数の計算機からのアクセス特性に基づいて、実記憶領域に格納されたデータを、レスポンス性能が異なる物理記憶デバイスの別の実記憶領域に移動する再配置処理を実行する。

20

【0008】

本発明に関連する更なる特徴は、本明細書の記述、添付図面から明らかになるものである。また、本発明の態様は、要素及び多様な要素の組み合わせ及び以降の詳細な記述と添付される特許請求の範囲の様態により達成され実現される。

30

【0009】

本明細書の記述は典型的な例示に過ぎず、本発明の特許請求の範囲又は適用例を如何なる意味に於いても限定するものではないことを理解する必要がある。

【発明の効果】**【0010】**

本発明の一実施形態によれば、ホスト上のタスクの種類および稼働状態に応じて、再配置処理の対象とするか否か(I/OモニタのON/OFF)を制御することで、高レスポンス要のタスクのデータが下位階層に配置されて性能が劣化することを防ぐことが可能となる。また、ユーザは、どのタスクを再配置処理の対象とすれば良いか(どのタスクのI/OモニタのON/OFFを制御すればよいか)を容易に判断できるようになる。

40

【図面の簡単な説明】**【0011】**

【図1】本発明の実施形態による計算機システム(ストレージシステム)の概略構成を示す図である。

【図2】本発明の実施形態におけるホスト計算機および管理サーバのハードウェアシステム構成例を示す図である。

【図3】本発明の実施形態におけるストレージ装置のハードウェアシステム構成例を示す図である。

50

【図4】本発明の実施形態における計算機システムのソフトウェアシステム構成例を示す図である。

【図5】本発明の実施形態におけるストレージ装置のソフトウェアシステム構成例を示す図である。

【図6】本発明の実施形態における論理ボリュームと物理ディスクの関係例を示す概念図である。

【図7】本発明の実施形態における論理ボリューム管理テーブルの構成例を示す図である。

【図8】本発明の実施形態における仮想ボリュームと論理ボリュームの関係例を示す概念図である。

10

【図9】本発明の実施形態における仮想ボリューム管理テーブルの構成例を示す図である。

【図10】本発明の実施形態におけるボリュームプール管理テーブルの構成例を示す図である。

【図11】本発明の実施形態におけるプール階層管理テーブルの構成例を示す図である。

【図12】本発明の実施形態における未使用領域管理テーブルの構成例を示す図である。

【図13】本発明の第1の実施形態における仮想ボリュームI/OモニタON/OFF管理テーブルの構成例を示す図である。

【図14】本発明の第1の実施形態におけるI/Oモニタ処理を説明するためのフローチャートである。

20

【図15】本発明の実施形態におけるページ階層再配置処理を説明するためのフローチャートである。

【図16】本発明の実施形態における再配置ページ管理テーブルの構成例を示す図である。

【図17】本発明の実施形態におけるホストボリューム管理テーブルの構成例を示す図である。

【図18】本発明の実施形態におけるI/Oモニタ要否管理テーブルの構成例を示す図である。

【図19】本発明の第1の実施形態におけるI/OモニタのON/OFF制御処理を説明するためのフローチャートである。

30

【図20】本発明の実施形態におけるタスク稼働情報の構成例を示す図である。

【図21】本発明の実施形態におけるI/OモニタON/OFF判断用情報表示画面の構成例を示す図である。

【図22】本発明の実施形態における再配置容量管理テーブルの構成例を示す図である。

【図23】本発明の実施形態におけるタスク稼働情報管理テーブルの構成例を示す図である。

【図24】本発明の実施形態におけるI/OモニタOFF推奨表示処理を説明するためのフローチャートである。

【図25】本発明の実施形態におけるタスク種別毎I/OモニタOFF候補管理テーブルの構成例を示す図である。

40

【図26】本発明の実施形態におけるI/Oモニタ制御設定推奨表示処理を説明するためのフローチャートである。

【図27】本発明の第2の実施形態におけるページI/OモニタON/OFF管理テーブルの構成例を示す図である。

【図28】本発明の第2の実施形態におけるI/Oモニタ処理を説明するためのフローチャートである。

【図29】本発明の第2の実施形態におけるI/OモニタのON/OFF制御処理を説明するためのフローチャートである。

【図30】本発明の第3の実施形態におけるI/OモニタのON/OFF制御処理を説明するためのフローチャートである。

50

【図31】本発明の第4の実施形態におけるI/OモニタのON/OFF制御処理を説明するためのフローチャートである。

【図32】本発明の実施形態におけるI/Oモニタ制御設定推奨表示画面の構成例を示す図である。

【発明を実施するための形態】

【0012】

以下、添付図面を参照して本発明の実施形態について説明する。添付図面では、機能的に同じ要素は同じ番号で表示される場合もある。なお、添付図面は本発明の原理に則った具体的な実施形態と実装例を示しているが、これらは本発明の理解のためのものであり、決して本発明を限定的に解釈するために用いられるものではない。

10

【0013】

本実施形態では、当業者が本発明を実施するのに十分詳細にその説明がなされているが、他の実装・形態も可能で、本発明の技術的思想の範囲と精神を逸脱することなく構成・構造の変更や多様な要素の置き換えが可能であることを理解する必要がある。従って、以降の記述をこれに限定して解釈してはならない。

【0014】

更に、本発明の実施形態は、後述されるように、汎用コンピュータ上で稼動するソフトウェアで実装しても良いし専用ハードウェア又はソフトウェアとハードウェアの組み合わせで実装しても良い。

【0015】

なお、以後の説明では「テーブル」形式によって本発明の各情報について説明するが、これら情報は必ずしもテーブルによるデータ構造で表現されていなくても良く、リスト、DB、キュー等のデータ構造やそれ以外で表現されていても良い。そのため、データ構造に依存しないことを示すために「テーブル」、「リスト」、「DB」、「キュー」等について単に「情報」と呼ぶことがある。

20

【0016】

また、各情報の内容を説明する際に、「識別情報」、「識別子」、「名」、「名前」、「ID」という表現を用いることが可能であり、これらについてはお互いに置換が可能である。

【0017】

以下では「プログラム」を主語（動作主体）として本発明の実施形態における各処理について説明を行うが、プログラムはプロセッサによって実行されることで定められた処理をメモリ及び通信ポート（通信制御装置）を用いながら行うため、プロセッサを主語とした説明としてもよい。また、プログラムを主語として開示された処理は管理サーバ等の計算機、情報処理装置が行う処理としてもよい。プログラムの一部または全ては専用ハードウェアで実現してもよく、また、モジュール化されていても良い。各種プログラムはプログラム配布サーバや記憶メディアによって各計算機にインストールされてもよい。

30

【0018】

本発明の各実施形態は、タスクの種類および稼働状態に応じてI/OモニタのON/OFFを制御し、高レスポンス要のタスクのデータが下位階層に配置されて性能が劣化することを防ぐことができる技術を提供する。また、各実施形態は、I/OモニタのON/OFF制御を実行する場合、ユーザがどのタスクについてI/OモニタのON/OFFを制御することを要するかについて、判断するための有効な情報を提示する。

40

【0019】

(1) 第1の実施形態

図1から図26、及び図32に基づいて、本発明による第1の実施形態について説明する。

【0020】

< 計算機システムの構成 >

図1は、本発明の実施形態による計算機システム（ストレージシステム、ITシステム

50

とも言う)の全体概要を示すブロック図である。

【0021】

計算機システム10は、少なくとも1つのホスト計算機100(単に、「ホスト」ということもある)と、少なくとも1つの管理サーバ(「管理計算機」とも言う)200と、1つ以上のストレージ装置(「ストレージサブシステム」とも言う)400と、を有し、互いにLAN301で接続されている。また、ストレージ装置400とホスト計算機100は互いにSAN300で接続されている。なお、ホスト計算機は、複数存在してもよく、その場合は管理用のLAN301に加えてデータ転送用のLANでホスト計算機が互いに接続されていてもよい。

【0022】

<ホスト計算機の内部構成>

図2Aは、ホスト計算機100の内部構成を示す図である。ホスト計算機100は、1つ以上のCPU110と、1つ以上のメモリ111と、1つ以上のSANアダプタ112と、1つ以上のLANアダプタ113と、1つ以上のストレージデバイス114と、を有し、それらは互いに内部バス115で接続されている。

【0023】

ホスト計算機100は、SANアダプタ112を介してストレージ装置400と接続されている。また、ホスト計算機100は、LANアダプタ113を介して管理サーバ200に接続されている。なお、ストレージデバイス114は必ずしも備えなくてもよい。備えない場合は、ストレージ装置400内のボリュームをソフトウェアの記憶領域として用いるようにすると良い。

【0024】

<管理サーバの内部構成>

図2Bは、管理サーバ200の内部構成を示す図である。管理サーバ200の内部構成はホスト計算機100と同様だが、必ずしもSANアダプタを備えていなくても良い。

【0025】

管理サーバ200は、LANアダプタ212を介して、ストレージ装置400およびホスト計算機100と接続されている。

【0026】

<ストレージ装置の内部構成>

図3は、ストレージ装置400の内部構成を示す図である。ストレージ装置400は、1つ以上のコントローラ410および1つ以上の物理ディスク411を有する。

【0027】

コントローラ410は、1つ以上のCPU412と、1つ以上のメモリ413と、1つ以上のNVRAM414と、1つ以上のキャッシュメモリ415と、1つ以上のバックエンドインターフェース416および417と、1つ以上のLANアダプタ418と、1つ以上のSANアダプタ419と、を有し、それらは相互に内部バス420で接続されている。

【0028】

コントローラ410は、バックエンドインターフェース416および417を介して物理ディスク411と接続されている。また、ストレージ装置400はLANアダプタ418を介して管理サーバ200と接続されている。また、ストレージ装置400はSANアダプタ419を介してホスト計算機100と接続されている。また、物理ディスク411はSAS(Serial Attached SCSI(Small Computer System Interface))ディスク、SATA(Serial Advanced Technology Attachment)ディスク、SSD(Solid State Drive)といった複数の種別のディスクから構成されていてもよい。

【0029】

<計算機システムのソフトウェア構成>

図4は、本実施形態の計算機システム10のソフトウェア構成の概要を示すブロック図である。ホスト計算機100上では、あるアプリケーションによる1つ以上のタスク50

10

20

30

40

50

0 と、1 つ以上の OS (Operating System) 5 0 1 が動作する。これらソフトウェアはストレージデバイス 1 1 4 またはストレージ装置 4 0 0 に格納されており、メモリ 1 1 1 にロードされ、CPU 1 1 0 を用いて実行される。

【 0 0 3 0 】

また、管理サーバ 2 0 0 上では、管理ソフトウェア 6 0 0 が動作する。管理ソフトウェア 6 0 0 は、メモリ 2 1 1 にロードされ、CPU 2 1 0 を用いて実行される。

【 0 0 3 1 】

さらに、ストレージ装置 4 0 0 上には 1 つ以上のボリュームプール 7 0 0 があり、これらから作成された 1 つ以上の仮想ボリューム 7 0 1 (後述) がある。仮想ボリューム 7 0 1 は、SAN 3 0 0 を介してホスト計算機 1 0 0 に割り当てられている。なお、本システム構成として仮想ボリューム 7 0 1 だけでなく、後述の論理ボリューム (実ボリューム) が混在していてもよい。

10

【 0 0 3 2 】

OS 5 0 1 は、複数のタスク 5 0 0 を管理し、実行すると共に、仮想ボリューム 7 0 1 および論理ボリューム 7 0 2 を認識し、管理する。

【 0 0 3 3 】

管理ソフトウェア 6 0 0 は、ホスト計算機 1 0 0 およびストレージ装置 4 0 0 から構成情報を取得し、管理テーブルに格納する。管理ソフトウェア 6 0 0 の詳細については後述する。

【 0 0 3 4 】

20

< ストレージ装置のソフトウェア構成 >

図 5 は、ストレージ装置 4 0 0 内のソフトウェア構成の概要を示す図である。ストレージ装置 4 0 0 上では、ストレージエージェント 7 0 3、仮想ボリュームマネージャ 7 0 4、及び論理ボリュームマネージャ 7 0 5 が動作している。これらは物理ディスク 4 1 1 または NVRAM 4 1 4 に格納されており、メモリ 4 1 3 にロードされ、CPU 4 1 2 を用いて実行される。

【 0 0 3 5 】

論理ボリュームマネージャ 7 0 5 は、物理ディスク 4 1 1 から 1 つ以上の論理ボリューム 7 0 2 を作成し、論理ボリューム 7 0 2 と物理ディスク 4 1 1 の間のマッピングを管理するソフトウェアである。

30

【 0 0 3 6 】

仮想ボリュームマネージャ 7 0 4 は、ボリュームプール 7 0 0 に登録された論理ボリューム 7 0 2 から 1 つ以上の仮想ボリューム 7 0 1 を作成し、仮想ボリューム 7 0 1 と、論理ボリューム 7 0 2 の間のマッピングを管理するソフトウェアである。

【 0 0 3 7 】

仮想ボリュームマネージャ 7 0 4 は、仮想ボリューム 7 0 1 にデータの書き込みがあった際、論理ボリューム 7 0 2 から未使用の領域を仮想ボリューム 7 0 1 の書き込みがあった場所に割り当てるソフトウェアである。

【 0 0 3 8 】

< 論理ボリュームと物理ディスクの関係 >

40

図 6 は、論理ボリューム 7 0 2 と物理ディスク 4 1 1 の関係を示す概念図である。図 6 の例では、論理ボリューム 7 0 2 は 4 つの物理ディスク 8 0 0、8 0 1、8 0 2、8 0 3 から構成されているものとする。1 - 1、1 - 2、1 - 3 ... とラベル付けされた物理ディスク内の領域は、あらかじめ決められたサイズに区切られた領域であり、ストライプと呼ばれる。また、P 1、P 2 ... とラベル付けされた領域は、対応するストライプのパリティ情報を格納する領域であり、パリティストライプと呼ばれる。

【 0 0 3 9 】

論理ボリュームマネージャ 7 0 5 が、論理ボリューム 7 0 2 と物理ディスク 4 1 1 のマッピング関係を管理するために、ボリューム管理テーブルがストレージ装置 4 0 0 内に保持されている。ボリューム管理テーブルの詳細については後述する。

50

【 0 0 4 0 】

< ボリューム管理テーブル >

図7は、ストレージ装置400が保持するボリューム管理テーブルの構成例を示す図である。ボリューム管理テーブルは、論理ボリュームID列900と、ディスク列901と、RAIDレベル列902と、ストライプサイズ列903と、ディスク種別列904とによって構成される。

【 0 0 4 1 】

論理ボリュームID列900には、論理ボリュームマネージャ705が各ボリュームに割り当てられ、各ボリュームを特定・識別するための識別子(識別情報)が登録される。

【 0 0 4 2 】

ディスク列901には、論理ボリュームを構成する物理ディスクを特定・識別するための識別子(識別情報)が登録される。

【 0 0 4 3 】

RAIDレベル列902には、論理ボリュームの構成に使用されているRAIDレベルが登録される。

【 0 0 4 4 】

ストライプサイズ列には、論理ボリュームの構成に使用されているストライプのサイズが登録される。

【 0 0 4 5 】

ディスク種別列904には論理ボリュームを構成する物理ディスクの種別が登録される。

【 0 0 4 6 】

< 仮想ボリュームと論理ボリュームの関係 >

図8は、仮想ボリューム701と論理ボリューム702の関係を示す概念図である。多くの場合、仮想ボリュームへの領域割り当ては物理ディスクの領域を直接割り当てるのではなく、図8のように論理ボリュームを構成した後、論理ボリューム内の一部領域を割り当てる形態を取る。

【 0 0 4 7 】

また、仮想ボリュームマネージャ704(図5)は、仮想ボリューム701とボリュームプール700に含まれる論理ボリューム702との関係や、ボリュームプール内における論理ボリューム702の階層および使用状況を管理する。このため、仮想ボリューム管理テーブル1100、ボリュームプール管理テーブル1200、プール階層管理テーブル1300、未使用領域管理テーブル1400がストレージ装置400内に保持される。なお、仮想ボリューム管理テーブル1100、ボリュームプール管理テーブル1200、プール階層管理テーブル1300、未使用領域管理テーブル1400の詳細については後述する。

【 0 0 4 8 】

< 仮想ボリューム管理テーブルの構成例 >

図9は、ストレージ装置400が保持する仮想ボリューム管理テーブル1100の構成例を示す図である。仮想ボリューム管理テーブル1100は、仮想ボリュームID列1101と、開始LBA列1102と、終了LBA列1103と、プールID列1104と、ページID列1105と、によって構成される。

【 0 0 4 9 】

仮想ボリュームID列1101には、仮想ボリュームに付与され、それを特定・識別するための識別子(識別情報)が登録される。

【 0 0 5 0 】

開始LBA列1102には、仮想ボリューム内の領域の開始LBAが登録される。終了LBA列1103には、仮想ボリューム内の領域の終了LBAが登録される。

【 0 0 5 1 】

プールID列1104には、仮想ボリュームに対して領域を割り当てているボリューム

10

20

30

40

50

プールを特定・識別するための識別子（識別情報）が登録される。

【0052】

ページID列1105には、仮想ボリューム内の各領域（開始LBAと終了LBAで指定された領域）に対応するページ（詳細後述）を特定・識別するための識別子（識別情報）が登録される。

【0053】

<ボリュームプール管理テーブル>

図10は、ストレージ装置400が保持するボリュームプール管理テーブル1200の構成例を示す図である。

【0054】

ボリュームプール管理テーブル1200は、プールID列1201と、ページID列1202と、論理ボリュームID列1203と、開始LBA列1204と、終了LBA列1205と、ストレージI/O数列1206と、によって構成される。

【0055】

プールID列1201には、ボリュームプールに付与され、それを特定・識別するための識別子（識別情報）が登録される。

【0056】

ページID列1202には、ページに付与され、それを特定・識別するための識別子（識別情報）が登録される。なお、ここでページとは、仮想ボリューム701に対してボリュームプール700から割り当てられる領域の単位である。各ページは、ボリュームプール内のいずれかの論理ボリュームに配置される。

【0057】

論理ボリュームID列1203には、ページの配置先の論理ボリュームを特定・識別するための識別子（識別情報）が登録される。

【0058】

開始LBA列1204には、論理ボリューム内におけるページの開始LBAが登録される。また、終了LBA列1205には、論理ボリューム内におけるページの終了LBAが登録される。開始LBA及び終了LBAによって各ページのサイズを知ることができる。一般的に、ページサイズは固定であるが、開始・終了LBAを適宜設定することにより、各ページを異なるサイズとして登録することもできる。

【0059】

ストレージI/O数列1206には、各ページに対して発行されたI/O数が記録される。図10では、プール2についてはI/O数カウント不要と設定されているため、プール2のストレージI/O数1206は全て「0」となっている。ストレージI/O数1206の詳細については後述する。

【0060】

<プール階層管理テーブル>

図11は、ストレージ装置400が保持するプール階層管理テーブル1300の構成例を示す図である。

【0061】

プール階層管理テーブル1300は、プールID列1301と、階層列1302と、論理ボリュームID列1303と、によって構成される。

【0062】

プールID列1301には、ボリュームプールを特定・識別するための識別子（識別情報）が登録される。

【0063】

階層列1302には、ボリュームプール内に設定されている階層が登録される。なお、階層は、論理ボリュームのディスク種別やRAIDレベルに基づいて決定されるが、ユーザが定義してもよい。

【0064】

10

20

30

40

50

論理ボリュームID列1303には、各階層に所属する論理ボリュームを特定・識別するための識別子（識別情報）が登録される。

【0065】

図11からは、例えば、プール1が、Tier0とTier1の2つの階層から構成され、Tier0は論理ボリューム1及び2、Tier1は論理ボリューム0及び3から構成されていることが分かる。

【0066】

<未使用領域管理テーブル>

図12は、ストレージ装置400が保持する未使用領域管理テーブル1400の構成例を示す図である。

【0067】

未使用領域管理テーブル1400は、各論理ボリュームにおいて使用されていない領域を管理するテーブルであり、論理ボリュームID列1401と、開始LBA列1402と、終了LBA列1403と、によって構成される。

【0068】

論理ボリュームID列1401には、ボリュームプール700に登録されている論理ボリュームを特定・識別するための識別子（識別情報）が登録される。

【0069】

開始LBA列1402および終了LBA列1403には、論理ボリューム702の中で未使用の領域1004の開始LBAおよび終了LBA列がそれぞれ登録される。

【0070】

通常、各論理ボリュームの領域はアドレス順に論理ボリュームに割り当てられて使用されていくが、使用中のボリュームが解放されて未使用領域になる場合もある。このため、図12に示されるように、未使用領域が不連続になる場合がある。

【0071】

<仮想ボリュームI/OモニタON/OFF管理テーブル>

仮想ボリュームマネージャ704は、ホスト計算機100から、仮想ボリューム701を構成するページに対して発行されたI/O数（ストレージI/O数）を記録し、記録したストレージI/O数に基づいて定期的にページの配置先の階層を決定しページを移動する。本明細書では、ストレージI/O数を記録することをI/Oモニタ、ページの配置先の階層を決定しページを移動することをページ階層再配置（再配置処理）と呼ぶこととする。

【0072】

仮想ボリュームマネージャ704は、I/Oモニタおよびページ階層再配置のため、前述の仮想ボリューム管理テーブル1100、ボリュームプール管理テーブル1200、プール階層管理テーブル1300、未使用領域管理テーブル1400に加え、ストレージ装置400に保持されている仮想ボリュームI/OモニタON/OFF管理テーブルを参照する。

【0073】

図13は、仮想ボリュームI/OモニタON/OFF管理テーブル1500の構成例を示す図である。

【0074】

仮想ボリュームI/OモニタON/OFF管理テーブル1500は、仮想ボリュームID列1501と、I/OモニタON/OFF列1502と、によって構成される。

【0075】

仮想ボリュームID列1501には、仮想ボリュームを特定・識別するための識別子（識別情報）が登録される。

【0076】

I/OモニタON/OFF列1502には、I/Oモニタ処理時に当該仮想ボリュームに対するI/O数を記録する（ON）か否（OFF）かが登録される。なお、このI/O

10

20

30

40

50

モニタON/OFF列1502の情報に関しては、初期値としてデフォルト値が設定されているが、システムを順次運用していくうちに、後述の図19の処理によって値(ON/OFF)は変化するようになっている。

【0077】

< I/Oモニタ処理 >

図14は、仮想ボリュームマネージャ704によって実行されるI/Oモニタ処理を説明するためのフローチャートである。本処理はホスト計算機100からストレージ装置400内の仮想ボリューム701に対してI/Oが発行されたことを契機として開始される(ステップ1600)。以下、ステップごとに説明を記述する：

ステップ1601：仮想ボリュームマネージャ704は、仮想ボリュームI/OモニタON/OFF管理テーブル1500を参照し、I/O発行先の仮想ボリュームのI/OモニタがONになっているかを確認する。ONであれば(ステップ1601でYesの場合)、処理はステップ1602に移行する。OFFであれば(ステップS1601でNoの場合)、処理は終了(ステップ1603)する。

【0078】

ステップ1602：仮想ボリュームマネージャ704は、ボリュームプール管理テーブル1200の、I/O発行先のページのページIDおよびボリュームプールのプールIDで決定される行のストレージI/O数列1206の数字を1つ増加させる。

【0079】

なお、ストレージI/O数の最新の傾向を得るため、ストレージI/O数列1206は一定時間ごとに0にリセットされるが、このときリセット前のストレージI/O数を履歴として保持しておいてもよい。この場合、ボリュームプール管理テーブル1200はそのストレージI/O数が記録された時間帯情報を保持する。つまり、この場合、ボリューム管理テーブル1200は、図23のようになる。

【0080】

また、I/OモニタのON/OFFは仮想ボリューム単位でなく、ホスト計算機100のSANアダプタ112の識別子であるWWN単位で実行してもよい。この場合、ステップ1601において、仮想ボリュームI/OモニタON/OFF管理テーブル1500の代わりに、仮想ボリュームI/OモニタON/OFF管理テーブル1500の仮想ボリュームID列1501をホストWWN列で置き換えたものを使用し、I/O発行元のWWNによってI/O記録の有無を判定する。つまり、I/O発行元のI/O数がカウントされることになる。

【0081】

< ページ階層再配置処理 >

図15は、仮想ボリュームマネージャ704によって実行されるページ階層再配置処理を説明するためのフローチャートである。本処理は一定時間の経過、あるいはユーザの手動実行を契機として、各ボリュームプールに対して実行開始される(ステップ1700)。以下、ステップごとに説明を記述する。

【0082】

ステップ1701：仮想ボリュームマネージャ704は、ボリュームプール管理テーブル1200のストレージI/O数列1206に記録された、各ページのストレージI/O数に基づいて、各ページの配置先階層を決定する。

【0083】

配置先階層の決定方法は、例えば、ストレージI/O数の大きい順に上位階層から配置先を割り振る方法でも良いし、ストレージI/O数の小さい順に下位階層から配置先を割り振る方法でもよい。また、ストレージI/O数としてストレージI/O数列1206に記録された最新の記録だけでなく、過去のストレージI/O数履歴も考慮して各ページの配置先の階層を決定しても良いし、またストレージI/O数だけでなくその他の指標を利用して決定しても良い。履歴を考慮する場合には、過去数時間分のI/O数の平均値を取るようにする。このように履歴を考慮すれば、瞬間的にI/O数が増加しているが、他の

10

20

30

40

50

期間（時間帯）ではほとんどI/O数がカウントされていない場合に再配置しないようにすることができる。

【0084】

さらに、例えば、ホスト計算機100からのリード/ライト回数の合計値（すなわちストレージI/O数）を考慮するのではなく、リード回数のみを考慮して配置先の階層を決定しても良い。また、ランダムアクセスの回数や、シーケンシャルリードの回数等の指標を考慮して、配置先の階層を決定しても良い。I/O数、ランダムアクセス数、シーケンシャルリード回数等を、まとめてアクセスの性質やアクセス特性とすることができる。

【0085】

なお、モニタOFFに設定されている仮想ボリュームのページであっても再配置処理の対象になる。当該ページについてはI/O数=0であるから、下位階層に移行される場合がある。

10

【0086】

ステップ1702：仮想ボリュームマネージャ704は、ステップ1701で決定した各ページの階層配置に基づき、ページの移動（再配置）を実行する。この際、決定した階層と、現在配置されている階層とが異なるページのみ、移動の対象となる。なお、各ページの現在配置されている階層は、ボリュームプール管理テーブル1200およびプール階層管理テーブル1300を組み合わせることで確認できる。

【0087】

ページの移動は、各ページに対する以下の3つのトランザクション処理からなる。つまり、(i)ページが現在配置されている論理ボリュームの領域（旧領域）のデータを、決定した階層に所属する論理ボリュームの未使用領域（新領域）にコピーすること、(ii)ボリュームプール管理テーブル1200の論理ボリュームID列1203、開始LBA列1204、終了LBA列1205を新領域の値に更新すること、(iii)再配置ページ管理テーブル（詳細後述）を更新すること、である。(i)～(iii)の処理が全ての移動対象のページに対して適用された後、処理は終了する（ステップ1703）。

20

【0088】

<再配置ページ管理テーブル>

図16は、ストレージ装置400が保持する再配置ページ管理テーブル1800の構成例を示す図である。

30

【0089】

再配置ページ管理テーブル1800は、ページ階層再配置処理において階層を移動したページの移動履歴の情報を保持するテーブルであって、プールID列1801と、ページID列1802と、移動元階層1803と、移動先階層1804とによって構成される。

【0090】

プールID列には、ボリュームプールを特定・識別するための識別子（識別情報）が登録される。

【0091】

ページID列には、ページ階層再配置処理において階層を移動されたページを特定・識別するための識別子（識別情報）が登録される。

40

【0092】

移動元階層列1803および移動先階層列1804にはそれぞれ、ページ階層再配置処理前後のページ配置先階層が記録される。移動先階層1804は、次の再配置処理時には、移動元階層1803となる。

【0093】

なお、再配置ページ管理テーブル1800は、直前に実行したページ階層再配置処理の記録だけでなく、それより過去のページ階層再配置処理の履歴を保持してもよい。この場合、再配置ページ管理テーブル1800は、ページ階層再配置処理が実行された時刻を保持する。また、図16では、格上げ及び格下げのみ示されているが、階層が変化しない場合があっても良い。

50

【 0 0 9 4 】

< ホストボリューム管理テーブル >

ストレージエージェント 7 0 3 は、論理ボリューム 7 0 2 および仮想ボリューム 7 0 1 にストレージボリューム ID (後述) を付与し、管理する。論理ボリューム 7 0 2 および仮想ボリューム 7 0 1 を総称してストレージボリュームと呼ぶ。

【 0 0 9 5 】

図 1 7 は、管理サーバ 2 0 0 が保持するホストボリューム管理テーブル 1 9 0 0 の構成例を示す図である。

【 0 0 9 6 】

ホストボリューム管理テーブル 1 9 0 0 は、管理ソフトウェア 6 0 0 がホスト計算機 1 0 0 から情報を取得し作成する。ホストボリューム管理テーブル 1 9 0 0 は、ホスト ID 列 1 9 0 1 と、ホスト WWN 列 1 9 0 2 と、ホストボリューム ID 列 1 9 0 3 と、ストレージボリューム ID 列 1 9 0 4 と、によって構成される。

【 0 0 9 7 】

ホスト ID 列 1 9 0 1 には、管理ソフトウェア 6 0 0 が付与する、ホスト計算機 1 0 0 を特定・識別するための識別子 (識別情報) が登録される。

【 0 0 9 8 】

ホスト WWN 列 1 9 0 2 には、ホスト計算機 1 0 0 の SAN アダプタ 1 1 2 を特定・識別するための識別子 (識別情報) が登録される。この情報は、WWN 単位で I / O 数をモニタするとき用いることができる。

【 0 0 9 9 】

ホストボリューム ID 列 1 9 0 3 には、ホスト計算機 1 0 0 が認識し、付与するボリュームを特定・識別するための識別子 (識別情報) が登録される。以降、ホスト計算機 1 0 0 が認識するボリュームをホストボリュームと呼ぶ。

【 0 1 0 0 】

ストレージボリューム ID 列 1 9 0 4 には、ホストボリュームに対応するストレージボリュームを特定・識別するための識別子 (識別情報) が登録される。当該ストレージボリューム ID は、各ストレージボリュームに対して、ストレージエージェント 7 0 3 によって付与される情報である。

【 0 1 0 1 】

なお、ストレージボリューム ID は、ストレージエージェント 7 0 3 が付与・保持する情報であるが、OS 5 0 1 は、例えば、SCSI Inquiry を利用してストレージ装置 4 0 0 から取得し、保持する。

【 0 1 0 2 】

簡単のため、図 1 7 においてストレージボリューム ID は、「仮想ボリューム 0」「論理ボリューム 3」といった形で表現しているが、実際には、多くの場合、ストレージエージェント 7 0 3 が仮想ボリュームと論理ボリュームの区別なく、通し番号で付与し、管理する。この場合、管理ソフトウェア 6 0 0 は、ストレージボリューム ID と、仮想ボリューム ID および論理ボリューム ID との対応を、ストレージエージェント 7 0 3 に問い合わせることで取得できる。このように、ストレージボリューム ID と仮想ボリューム ID および論理ボリューム ID は容易に対応関係を特定できるため、以降「ストレージボリューム ID から特定した仮想ボリューム ID」または「ストレージボリューム ID から特定した論理ボリューム ID」と記載すべきところを、単に仮想ボリューム ID または論理ボリューム ID と表現することがある。

【 0 1 0 3 】

また、図 1 7 では、ホストボリュームとストレージボリュームが 1 対 1 に対応しているが、必ずしも 1 対 1 でなくても良く、1 対多、多対 1 であっても良い。

【 0 1 0 4 】

< I / O モニタ要否管理テーブル >

図 1 8 は、管理サーバ 2 0 0 が保持し、管理ソフトウェア 6 0 0 が参照する I / O モニ

10

20

30

40

50

タ要否管理テーブル 2000 の構成例を示す図である。

【0105】

I/O モニタ要否管理テーブル 2000 は、ホスト ID 列 2001 と、タスク名列 2002 と、タスク種別列 2003 と、I/O モニタ要否列 2004 とによって構成される。

【0106】

ホスト ID 列 2001 には、管理ソフトウェア 600 が付与し、ホスト計算機 100 を特定・識別するための識別子（識別情報）が登録される。

【0107】

タスク名列 2002 には、OS 501 上でタスク 500 を区別する名称が登録される。

【0108】

タスク種別列 2003 には、タスク 500 の一般的な分類が登録される。

【0109】

I/O モニタ要否列 2004 には、各タスクが発行する I/O 数を記録すべきか否かが登録される。この情報は、例えば、初期値としてデフォルトで全て「要」となっており（特定のタスクについてモニタの要否が最初から分かっている場合には、管理者が「要」或いは「否」に設定するようにしても良い）、計算機システム 10 を運用していくうちに、後述の図 21 の表示に基づいて、ユーザ（管理者）が決めるようにしても良いし、後述の図 24 や 26 の処理によって自動的にモニタ要否を「否」に決定するようにしても良い。

【0110】

なお、I/O モニタ要否管理テーブル 2000 は、ユーザが管理ソフトウェア 600 に入力して作成する方式でもよいし、管理ソフトウェア 600 が OS 501 上で稼働するタスク 500 を検出して自動的に作成する方式でもよい。また、I/O モニタ要否管理テーブル 2000 は I/O モニタの ON/OFF 制御処理の前に作成される。

【0111】

一般的に、夜間集計バッチ処理はモニタ不要になりやすいが、バッチ処理も時間通りに完了するとは限らないため、モニタを要する場合もある。例えば、バッチ処理で、ボリューム対するアクセス数が多く、終了予定時間に間に合わない可能性がある場合に、アクセススピードが遅い階層に各ページが配置されたままであると、不都合である。この場合には、モニタして再配置処理を可能にした方が効率的である。このように、I/O モニタしなくても良いタスクは必ずしも表面上明らかでない場合もあり、本発明の処理が非常に有効となってくるのである。

【0112】

< I/O モニタ ON/OFF 制御処理 >

図 19 は、管理サーバ 200 において、管理ソフトウェア 600 が実行する I/O モニタの ON/OFF 制御処理（例）を説明するためのフローチャートである。本処理は、定期的に、ホスト毎に実行開始される（ステップ 2100）。ここで、「定期的」とは、周期的に実行する場合、各タスクの実行起動時、ユーザ設定された時間に実行する場合等を含み、等間隔で実行する場合だけでなく、任意のタイミングで実行する場合を含むものである。なお、ストレージ I/O 数だけでなくその他の指標のモニタ ON/OFF を制御するようにしても良い。例えば、ホスト計算機 100 からのリード/ライト回数の合計値（すなわちストレージ I/O 数）のモニタ ON/OFF を制御するのではなく、リード回数だけのモニタ ON/OFF を制御する。また、ランダムアクセスの回数や、シーケンシャルリードの回数等の指標の ON/OFF であっても良い。つまり、本実施形態では、アクセスの性質或いはアクセス特性のモニタ ON/OFF を制御すると言うことが可能である。以下、I/O モニタの ON/OFF の制御について、ステップごとに説明を記述する。

【0113】

ステップ 2101：管理ソフトウェア 600 は、各ホスト計算機 100 と通信し、ホスト計算機 100 上の OS 501 から、ホスト計算機 100 毎のタスク稼働情報 2200（図 20 参照）を取得する。

【0114】

10

20

30

40

50

ステップ2102：管理ソフトウェア600は、取得したタスク稼働情報において、「稼働状態2202がON」かつ「I/Oモニタ要否管理テーブル2000においてI/Oモニタ要否列2004の値が「否」である」ようなタスクが1つでも存在するかを判定する。これは、I/Oモニタすべきでないタスクが動いている仮想ボリュームについてはI/OモニタをOFFにする趣旨である。なお、モニタOFFの条件として「稼働状態2202がON」の代わりに「ホストI/O発行量2203が閾値以上であるもの」などとしてもよい。また「1つでも存在するかを判定する」としたが「閾値（例えば、「否」であるタスクが3つ）以上存在する」などとしても良く、また、全てのタスクが「否」でないとモニタOFFとしないようにしても良い。ステップ2102で判定結果がyesであった場合、処理はステップ2103に進み、判定結果がnoであった場合、処理はステップ2104に進む。

10

【0115】

ステップ2103：管理ソフトウェア600は、ホスト計算機100に割り当てられた仮想ボリュームのI/OモニタをOFFにする。具体的には、まず、管理ソフトウェア600は、ホストボリューム管理テーブル1900（図17）からホスト計算機100に割り当てられた仮想ボリュームの仮想ボリュームIDを特定し、ストレージ装置400に対して仮想ボリュームIDおよびI/OモニタのOFF命令を送信する。そして、そのOFF命令を受信したストレージ装置400は、仮想ボリュームI/OモニタON/OFF管理テーブル1500（図13）の該当する仮想ボリュームIDのI/OモニタON/OFF列1502の値をOFFにする。

20

【0116】

ステップ2104：管理ソフトウェア600は、ホスト計算機100に割り当てられた仮想ボリュームのI/Oモニタ1502をONにする。具体的には、まず、管理ソフトウェア600は、ホストボリューム管理テーブル1900からホスト計算機100に割り当てられた仮想ボリュームの仮想ボリュームIDを特定し、ストレージ装置400に対して仮想ボリュームIDを含むI/OモニタのON命令を送信する。そして、そのON命令を受信したストレージ装置400は、仮想ボリュームI/OモニタON/OFF管理テーブル1500の該当する仮想ボリュームIDのI/OモニタON/OFF列1502の値をONにする。

【0117】

<タスク稼働情報>

図20は、ホスト計算機毎のタスク稼働情報2200の例を示す図である。タスク稼働情報2200は、タスク名2201と、稼働状態2202と、ホストI/O発行量2203とによって構成される。

30

【0118】

タスク名2201は、OS501上でタスク500を区別する名称を特定するための情報である。

【0119】

稼働状態2202は、各タスクが稼働している（ON）か否（OFF）かを示す情報である。

40

【0120】

ホストI/O発行量2203は、各タスクがOS501上で発行している時間当たりI/O数を示す情報である。

【0121】

なお、タスク稼働情報2200は、メモリ使用量等、他の情報を含んでもよい。

【0122】

<I/OモニタON/OFF判断用情報表示画面（GUI）>

図21は、管理サーバ200において、管理ソフトウェア600が表示する、I/OモニタON/OFF判断用情報表示画面2300の構成例を示す図である。I/OモニタON/OFF判断用情報表示画面2300は、ユーザがI/Oモニタ要否管理テーブル20

50

00を入力する際に、どのタスクのI/OモニタをOFFにすべきかを判断するために使用する画面である。図21は、仮想ボリューム701のプール毎に再配置発生量を管理するためのユーザインタフェースとなっている。なお、プール単位で再配置発生量をモニタするのは、第1の実施形態だけでなく、第2乃至第4の実施形態でも同様である。

【0123】

I/OモニタON/OFF判断用情報表示画面2300は大きく分けて3つの構成要素からなる。すなわち、プール内再配置発生量グラフ2301と、ホスト内タスクI/O比率グラフ2306と、全ホスト階層格上げページI/O寄与率グラフ2307である。なお、以下の記述においてホスト内タスクI/O比率グラフ2306、及び全ホスト階層格上げページI/O寄与率グラフ2307は、特に階層格上げ（下位階層から上位階層への移動）に注目しているが、階層格下げ（上位階層から下位階層への移動）に関しても、以降で述べる計算により同様のグラフを表示してもよい。また、以降で述べる計算は一例であり、ユーザがプール内再配置発生量グラフ2301、ホスト内タスクI/O比率グラフ2306、全ホスト階層格上げページI/O寄与率グラフ2307の各グラフから、以降で述べるのと同様の判断をできるのであれば、別の計算方法でもよい。

【0124】

(i) プール内再配置発生量グラフ2301

プール内再配置発生量グラフ2301は、ボリュームプール700の各時間帯でのページ階層再配置処理ごとに、階層格上げページ群（すなわち下位階層から上位階層に移動された一つ以上のページ）の合計容量（階層格上げ容量2304）およびそのページ群が割り当てられたホスト計算機の内訳2302を、ページ階層再配置処理が実行された時間帯ごとに表示する。また、階層格下げページ群（すなわち上位階層から下位階層に移動された一つ以上のページ）についても同様の情報を表示する（すなわち階層格下げ容量2305およびそのページ群が割り当てられたホスト計算機の内訳2303）。格上げページ或いは格下げページの別は再配置ページ管理テーブル1800（図16参照）から判断することができる。また、ボリュームプール管理テーブル1200（図10）から、各ページのLBAの範囲が分かるので、これに基づいて各ページの容量を算出することができる。また、どのページがどのホスト計算機に割り当てられているかは、仮想ボリューム管理テーブル1100（図9）を参照して、ページIDに対応する仮想ボリュームを特定し、その仮想ボリュームIDが割り当てられたホスト計算機を、ホストボリューム管理テーブル（図17）を参照して取得することにより、決定することができる。

【0125】

ユーザ（管理者）は、プール内再配置発生量グラフ2301により、どのホスト計算機100が階層格上げおよび階層格下げの要因になっているかを判断できる。また、プール内再配置発生量グラフ2301は、再配置容量管理テーブル2400（図22参照）に基づいて表示される。

【0126】

(ii) ホスト内タスクI/O比率グラフ2306

ホスト内タスクI/O比率グラフ2306は、例えば、プール内再配置発生量グラフ2301上の、確認したい時間帯におけるホスト計算機（例えば、24:00の「格上げのホスト3」）を選択することにより表示されるグラフである。

【0127】

ホスト内タスクI/O比率グラフ2306は、ページ階層再配置処理の実行されたある時間帯における「各ホスト計算機100における各タスク500のホストI/O発行量の比率」を示すものである。

【0128】

ユーザは、ホスト内タスクI/O比率グラフ2306により、ある時間帯のページ階層再配置処理において、各ホスト計算機100に対応するページ群の階層格上げの要因となったタスク500を判断できる。図21に示されるように、24:00のホスト3の再配置処理では、アンチウイルスAが要因で格上げとなったことが分かる。

10

20

30

40

50

【 0 1 2 9 】

ホスト I / O 発行量の比率については、当該ページ階層再配置処理のステップ 1 7 0 1 において、管理ソフトウェア 6 0 0 が、ページの配置先階層の決定に用いたストレージ I / O 数を記録した時間帯における、各タスク 5 0 0 のホスト I / O 発行量に基づいて計算する。例えば、ステップ 1 7 0 1 において、最新のストレージ I / O 数だけでなく、履歴として保持された過去のストレージ I / O 数も利用した場合、過去のストレージ I / O 数も計算に含まれる。ただし、多くの場合、過去のストレージ I / O 数は最新のストレージ I / O 数よりページ再配置処理における考慮の対象として重要性が低い。このため、仮想ボリュームマネージャ 7 0 4 はステップ 1 7 0 1 実行時に、適当な定数（重み係数）を過去のストレージ I / O 数に乘算して、最新のストレージ I / O 数より寄与率を低くするようによい。このような場合、管理ソフトウェア 6 0 0 が「各ホスト計算機 1 0 0 における各タスク 5 0 0 のホスト I / O 発行量の比率」を計算する際に、ステップ 1 7 0 1 で仮想ボリュームマネージャ 7 0 4 が過去のストレージ I / O 数に対して乗算したのと同じ係数を、当該時間帯の各タスクのホスト I / O 発行量に乘算する。なお、「各ホスト計算機 1 0 0 における各タスク 5 0 0 のホスト I / O 発行量の比率」の計算は、タスク稼働情報管理テーブル 2 5 0 0（図 2 3 参照）に基づいて実行される。

10

【 0 1 3 0 】

ここで、タスク稼働情報管理テーブル 2 5 0 0 を用いて「24 時の h o s t 3 のホスト内タスク I / O 比率グラフ 2 3 0 6 の表示に用いる、ホスト I / O 発行量の比率」の計算の例を簡単に説明する。例えば、24 時のページ階層再配置処理において、ステップ 1 7 0 1 でページの配置先階層の決定に用いられたストレージ I / O 数が 1 8 時から 2 4 時に記録されているとする（なおストレージ I / O 数が記録された時間帯は仮想ボリュームマネージャ 7 0 4 から取得できる）。管理ソフトウェア 6 0 0 は、タスク稼働情報管理テーブル 2 5 0 0 の h o s t 3 の各タスク、すなわち O L T P 3 とアンチウィルス A のそれぞれについて、1 8 時から 2 4 時のホスト I / O 発行量を合計する。この合計値の比率が「24 時の h o s t 3 のホスト内タスク I / O 比率グラフ 2 3 0 6 の表示に用いる、ホスト I / O 発行量の比率」となる。

20

【 0 1 3 1 】

(iii) 全ホスト階層格上げページ I / O 寄与率グラフ 2 3 0 7

全ホスト階層格上げページ I / O 寄与率グラフ 2 3 0 7 は、例えば、プール内再配置発生量グラフ 2 3 0 1 上の、確認したい時間帯のホスト計算機（例えば、6 : 00 の「格上げのホスト 3」及び「格上げのホスト 2」）を選択することにより表示されるグラフである。

30

【 0 1 3 2 】

全ホスト階層格上げページ I / O 寄与率グラフ 2 3 0 7 は、ページ階層再配置処理の実行された時間帯ごとに、「全ホスト計算機 1 0 0 における各タスク 5 0 0 の階層格上げ寄与率」を示すものである。

【 0 1 3 3 】

ユーザは、全ホスト階層格上げページ I / O 寄与率グラフ 2 3 0 7 を参照することにより、ある時間帯のページ階層再配置処理において、どのホスト計算機 1 0 0 のどのタスク 5 0 0 が階層格上げの要因になったかを判断できる。図 2 1 に示されるように、6 : 00 の再配置処理では、ホスト 2 の O L T P 2 が支配的であり、これが格上げの要因となったことが分かる。従って、O L T P 2 の処理が重要であれば I / O モニタ要否を「要」のままにし、重要でなければ「否」に設定すれば良いことになる。なお、割合が少ない O L T P 3 やアンチウィルス A を「否」にしても、下位階層に配置されてしまうことを防止することによる性能劣化防止の効果は少ないと言える。

40

【 0 1 3 4 】

なお、全ホスト階層格上げページ I / O 寄与率グラフ 2 3 0 7 で表示される「全ホスト計算機 1 0 0 における各タスク 5 0 0 の階層格上げ寄与率」は、ホスト内タスク I / O 比率グラフ 2 3 0 6 で表示する「各ホスト計算機 1 0 0 における各タスク 5 0 0 のホスト I

50

「I/O発行量の比率」と「階層格上げ容量2304のホスト計算機の内訳2302」とから、管理ソフトウェア600が計算する。すなわち、「階層格上げ容量2304のホスト計算機の内訳2302」の比率を「各ホスト計算機100における各タスク500のホストI/O発行量の比率」に対して、乗算したものが「全ホスト計算機100における各タスク500の階層格上げ寄与率」である。

【0135】

ここで、「全ホスト計算機100における各タスク500の階層格上げ寄与率」の計算の例を示す。階層格上げ容量2304が100GBであり、その内訳がhost2:70GB、host3:30GBであった場合、「階層格上げ容量2304のホスト計算機の内訳2302」の比

10

率は、host2:70%、host3:30%となる。「各ホスト計算機100における各タスク500のホストI/O発行量の比率」が、それぞれ、host2においてOLTP2:100%、host3においてアンチウイルスA:40%、OLTP3:60%である場合、「全ホスト計算機100における各タスク500の階層格上げ寄与率」は、OLTP2:70%(=70%×100%)、アンチウイルスA:12%(=30%×40%)、OLTP3:18%(=30%×60%)となる。

【0136】

<再配置容量管理テーブル>

図22は、管理ソフトウェア600が作成し、管理サーバ200内に保持される、再配置容量管理テーブル2400の構成例を示す図である。

20

【0137】

再配置容量管理テーブル2400は、プールID列2401と、ホストID列2402と、再配置時間帯列2403と、階層格上げ容量列2404と、階層格下げ容量列2405とによって構成される。

【0138】

プールID列2401には、ボリュームプール700を特定・識別するための識別子(識別情報)が登録される。

【0139】

ホストID列2402には、プールID列2401の値によって特定されるボリュームプールから作成された仮想ボリュームが割り当てられているホスト計算機を特定・識別するための識別子(識別情報)が登録される。

30

【0140】

再配置時間帯列2403には、ストレージ装置400においてページ階層再配置処理が実行された時間帯を示す情報が登録される。

【0141】

階層格上げ容量列2404および階層格下げ容量列2405には、ページ階層再配置処理において、下位階層から上位階層、および上位階層から下位階層に移動されたページ群の合計容量がそれぞれ登録される。

【0142】

階層格上げ容量列2404および階層格下げ容量列2405の値については、管理ソフトウェア600が、ホストボリューム管理テーブル1900と、ストレージ装置400から取得する仮想ボリューム管理テーブル1100と、ボリュームプール管理テーブル1200と、再配置ページ管理テーブル1800とにより計算する。例えば「pool1のhost1の階層格上げ容量」を計算する場合、再配置ページ管理テーブル1800に登録されたpool1のページ群のうち、「移動先階層列1804の値が移動元階層列1803より上位」であり、かつ「対応する仮想ボリュームの割り当て先のホスト計算機がhost1であるもの」が特定される。この特定は仮想ボリューム管理テーブル1100およびホストボリューム管理テーブル1900を組み合わせることで実行できる。特定した各ページの容量をボリュームプール管理テーブル1200から取得し、合計することで「pool1のhost1の階層格上げ容量」を計算できる。

40

50

【 0 1 4 3 】

< タスク稼働情報管理テーブル >

図 2 3 は、管理サーバ 2 0 0 が保持する、タスク稼働情報管理テーブル 2 5 0 0 の構成例を示す図である。

【 0 1 4 4 】

タスク稼働情報管理テーブル 2 5 0 0 は、各ホスト計算機 1 0 0 のタスク稼働情報 2 2 0 0 の履歴情報を保持しており、管理ソフトウェア 6 0 0 が各ホスト計算機 1 0 0 から定期的にタスク稼働情報 2 2 0 0 を取得することにより、生成・更新されるテーブルである。

【 0 1 4 5 】

タスク稼働情報管理テーブル 2 5 0 0 は、ホスト ID 列 2 5 0 1 と、タスク名列 2 5 0 2 と、稼働状態列 2 5 0 3 と、ホスト I / O 発行量列 2 5 0 4 と、採取時刻列 2 5 0 5 と、によって構成される。

【 0 1 4 6 】

ホスト ID 列 2 5 0 1 には、ホスト計算機 1 0 0 を特定・識別するための識別子（識別情報）が登録される。

【 0 1 4 7 】

タスク名列 2 5 0 2、稼働状態列 2 5 0 3、及びホスト I / O 発行量列 2 5 0 4 に登録される情報は、タスク稼働情報 2 2 0 0 における同名の項目と同じであり、ここでは説明を省略する。

【 0 1 4 8 】

採取時刻列 2 5 0 5 には、管理ソフトウェア 6 0 0 がタスク稼働情報 2 2 0 0 を取得した時刻が登録される。

【 0 1 4 9 】

< I / O モニタ OFF 推奨表示処理 (I) >

図 2 4 は、I / O モニタ OFF 推奨表示処理 (例) を説明するためのフローチャートである。管理ソフトウェア 6 0 0 は、本処理により、ユーザに対して I / O モニタを OFF とすべきタスク 5 0 0 を自動的に提示する。図 2 1 では、表示された各グラフによりユーザ (管理者) が I / O モニタを OFF とするタスクを判断しているので、自動的に対象のタスクを提示する点で図 2 1 の場合と異なり、ユーザが判断しなくても良いという利点がある。

【 0 1 5 0 】

本処理によって提示される、I / O モニタを OFF とすべきタスク 5 0 0 は、階層格上げを発生させており、かつ一般的に I / O モニタを OFF とすべき種別のタスクである。また、本処理は、定期的あるいはユーザの指示を契機として、再配置容量管理テーブル 2 4 0 0 のプール毎に実行される (ステップ 2 6 0 0)。ただしステップ 2 6 0 0 の前に、少なくとも一度はページ階層再配置処理が実行されており、再配置容量管理テーブル 2 4 0 0 やタスク稼働情報管理テーブル 2 5 0 0 にデータが格納されている必要がある。ここで、「定期的」の意味については上述した通りである。以下、ステップごとに説明を記述する。

【 0 1 5 1 】

ステップ 2 6 0 1 : 管理ソフトウェア 6 0 0 は、再配置容量管理テーブル 2 4 0 0 を参照し、階層格上げ容量列 2 4 0 4 の値が所定の閾値を超過しているホスト計算機 1 0 0 およびその再配置時間帯を検索する。例えば、閾値を 2 0 G B (以上) とすると、再配置容量管理テーブル 2 4 0 0 からは、20:00-24:00 のホスト 1、12:00-18:00 のホスト 2、6:00-12:00 のホスト 2、24:00-6:00 のホスト 3 が検出される。

【 0 1 5 2 】

ステップ 2 6 0 2 : 管理ソフトウェア 6 0 0 は、ステップ 2 6 0 1 の検索で見つかった各「ホスト計算機 1 0 0 および再配置時間帯」について、「各ホスト計算機 1 0 0 における各タスク 5 0 0 のホスト I / O 発行量の比率」を計算する。なお、「各ホスト計算機 1

10

20

30

40

50

00における各タスク500のホストI/O発行量の比率」はホスト内タスクI/O比率グラフ2306の表示時に用いるものと同じである。

【0153】

ステップ2603：管理ソフトウェア600は、ステップ2602の計算結果（各「ホスト計算機100および再配置時間帯」に対する「各ホスト計算機100における各タスク500のホストI/O発行量の比率」）を用いて、「タスク種別毎I/OモニタOFF候補管理テーブル2700（図25）にI/OモニタOFF候補として登録されている種別のタスクである」かつ「各ホスト計算機100における各タスク500のホストI/O発行量の比率」が所定の閾値を超過している」ようなタスク500を検索する。

【0154】

ステップ2604：管理ソフトウェア600は、ステップ2603の検索で得られたタスク500があれば、I/OモニタをOFFとすべきタスクとしてユーザに提示する。

【0155】

なお、図32は、ステップ2604において提示されるI/Oモニタ制御設定推奨表示画面3400の構成例を示す図である。I/Oモニタ制御設定推奨表示画面3400は、ホスト名3401と、タスク名3402と、I/OモニタOFFチェックボックス3403と、実行ボタン3404と、キャンセルボタン3405とによって構成される。

【0156】

ホスト名3401及びタスク名3402には、ステップ2603の検索で得た結果が表示される。

【0157】

ユーザは、I/OモニタOFFチェックボックス3403にチェックを入れ、実行ボタン3404を押下することにより、当該タスクのI/OモニタをOFFにすることができるようになっている。すなわち、I/Oモニタ要否管理テーブル2000における当該タスクの行のI/Oモニタ要否列2004の値を「否」に変更できる。また、これ以外の情報を表示してもよい。さらに、本ステップにおいてユーザにI/OモニタをOFFとすべきタスクを提示するだけでなく、自動的にI/OモニタをOFFとしてもよい。

【0158】

<タスク種別毎I/OモニタOFF候補管理テーブル>

図25は、管理サーバ200で保持され、管理ソフトウェア600が参照するタスク種別毎I/OモニタOFF候補管理テーブル2700の構成例を示す図である。

【0159】

タスク種別毎I/OモニタOFF候補管理テーブル2700は、タスク種別列2701と、I/OモニタOFF候補列2702とにより構成される。

【0160】

タスク種別列2701には、ホスト計算機100上で実行されるタスクの一般的な種別が登録される。

【0161】

I/OモニタOFF候補列2702には、各タスク種別が、一般的にI/OモニタOFFとすべき(yes)か否(no)かが登録される。例えば、一般的にインデックス構築のタスクは、インデックスを構築するために多量のI/Oを発行するが、本来高レスポンスが不要なデータに対してもI/Oを発行するため、このI/Oは記録すべきではない（I/OモニタOFFとすべき）。なお、タスク種別毎I/OモニタOFF候補管理テーブル2700は管理ソフトウェア600が初めから固定値（Default値）として保持していてもよいし、ユーザが適宜入力してもよい。

【0162】

<I/OモニタOFF推奨表示処理(II)>

図26は、図24とは別のI/OモニタOFF推奨表示処理を説明するためのフローチャートである。管理ソフトウェア600は本処理により、ユーザに対してI/OモニタをOFFとすべきタスク500を提示する。本処理によって提示される、I/OモニタをO

10

20

30

40

50

FFとすべきタスク500(タスクA)は、あるページ階層再配置時間帯において所定の閾値以上の階層格上げを発生させており、かつ、その次のページ階層再配置時間帯においては別の1つ以上のタスク500(タスクB群)が大量の階層格上げを発生させているものである。つまり、2つ(必ずしも2つでなくても良い)の連続する時間帯のうちで所定閾値以上の再配置処理が実行された時間帯が検出される。これは単発的に階層格上げの発生量が増加した時間帯ではなく、ある程度定常的に階層格上げが発生している時間帯を処理対象とする趣旨である。このようなタスクAは、一時的にホストI/O発行量が増大し、ページ階層再配置処理においてタスクAで使用されるページ群を階層格上げさせ、結果としてタスクB群のページを階層格下げさせている可能性がある。本処理は、定期的あるいはユーザの指示を契機として、再配置容量管理テーブル2400のプール毎に実行される(ステップ2800)。「定期的」の意味については上述した通りである。なお、ステップ2800の処理の前に、少なくとも2度はページ階層再配置処理が実行されており、再配置容量管理テーブル2400やタスク稼働情報管理テーブル2500にデータが格納されている必要がある。以下、ステップごとに説明を記述する。

10

【0163】

ステップ2801:管理ソフトウェア600は、再配置容量管理テーブル2400を参照し、着目する再配置時間帯(自再配置時間帯とも言う)およびその直後の再配置時間帯ともに、階層格上げ容量の全ホスト合計値が所定の閾値を超過しているような再配置時間帯を検索する。なお、「自再配置時間帯および直後の再配置時間帯」ではなく、「自再配置時間帯および直後のk個以内の再配置時間帯のいずれか」などとしてもよい(kは任意の自然数である)。例えば、24:00-6:00のホスト2とホスト3の階層格上げ容量の合計値(50GB)と、6:00-12:00のホスト2とホスト3の階層格上げ容量の合計値(40GB)の両方が所定の閾値(例えば、35GB)を超えているので、これらの時間帯がその後の処理を続行する対象として検出される。

20

【0164】

ステップ2802:管理ソフトウェア600は、ステップ2801で得られた各再配置時間帯、およびその各々の直後の再配置時間帯について、「全ホスト計算機100における各タスク500の階層格上げ寄与率」を計算する。「全ホスト計算機100における各タスク500の階層格上げ寄与率」は全ホスト階層格上げページI/O寄与率グラフ2307の表示に用いる場合の情報と同じである。

30

【0165】

ステップ2803:管理ソフトウェア600は、ステップ2802で計算した、任意の再配置時間帯における「全ホスト計算機100における各タスク500の階層格上げ寄与率」において寄与率が閾値を超過しているタスク500で、かつその直後の再配置時間帯における「全ホスト計算機100における各タスク500の階層格上げ寄与率」において寄与率が閾値を下回っているタスク500を取得する。例えば、アンチウィルスAについて、18:00-24:00の時間帯での寄与率が閾値を上回っているが、24:00-6:00の時間帯での寄与率がその閾値を下回っている場合には、そのアンチウィルスAが該当するタスクとして取得されることになる。

40

【0166】

ステップ2804:管理ソフトウェア600は、ステップ2803の検索で得られたタスク500があれば、I/OモニタをOFFとすべきタスクとしてユーザに提示する。なお、ステップ2804において提示される情報は、上述の図32で示されるI/Oモニタ制御設定推奨表示画面3400と同様である。なお、ユーザにI/OモニタをOFFとすべきタスクを提示するだけでなく、自動的にI/OモニタをOFFとしてもよい。

【0167】

図24および図26に示したI/OモニタOFF推奨表示処理は例であり、他にもI/OモニタOFF推奨表示処理がある。例えば、下記のようなタスク500をI/OモニタをOFFとすべきタスクとしてユーザに提示してもよい。つまり、あるページ階層再配置時間帯に大量の階層格上げを発生させているタスク500であり、かつその時に階層格上

50

げされたページ群の多くが階層格上げ以降、あまりアクセスされていないようなタスク 500 である。このようなタスク 500 は、一時的にホスト I/O 発行量が増大して大量の階層格上げを発生させたが、それ以降はアクセスが減少し、結果として無意味な階層格上げを発生させた可能性が高い。なお、このようなタスク 500 は、ボリューム管理テーブル 1100、再配置ページ管理テーブル 1800、仮想ホストボリューム管理テーブル 1900、再配置容量管理テーブル 2400、タスク稼働情報管理テーブル 2500 を用いることで特定できる。

【0168】

(2) 第 2 の実施形態

次に、本発明の第 2 の実施形態による計算機システムについて、第 1 の実施形態で用いた図 1 から図 12、図 15 から図 18、及び図 20 から図 26、さらに新たに図 27 から 29 を参照して、説明する。なお、第 1 及び第 2 の実施形態の違いは、I/O モニタの ON/OFF 制御の粒度である。すなわち、第 1 の実施形態では仮想ボリューム単位で I/O モニタの ON/OFF を制御するが、第 2 の実施形態ではページ単位で I/O モニタの ON/OFF を制御する。

【0169】

第 2 の実施形態におけるシステム構成は、第 1 の実施形態において図 1 から図 5 を用いて説明したものと同様である。また図 6 から図 12、図 15 から図 18、図 20 から図 26 に示す概念図、管理テーブル、フローチャートは、本実施形態にも適用される。ただし、第 2 の実施形態では、仮想ボリューム 701 をそのままホストボリュームとして利用するのではなく、1 つ以上の仮想ボリューム 701 上の、一部または複数の領域を、仮想的なホストボリュームとしてホスト計算機 100 上で利用するためのプログラム（ホストボリュームマネージャ）が、ホスト計算機 100 上や別の計算機上に存在してもよい。なお、ホストボリュームマネージャは、当該仮想的なホストボリュームと、当該一部または複数の領域の間のマッピング情報を保持する。また、ホスト計算機 100 として、物理マシンだけでなく、仮想マシンが存在してもよい。

【0170】

< ページ I/O モニタ ON/OFF 管理テーブル >

図 27 は、ページ I/O モニタ ON/OFF 管理テーブル 2900 の構成例を示す図である。このテーブルは、ページ単位で I/O モニタの ON 及び OFF を管理するためのものであり、第 1 の実施形態における図 13（仮想ボリューム単位で I/O モニタの ON 及び OFF を管理するテーブル）に対応する。

【0171】

ページ I/O モニタ ON/OFF 管理テーブル 2900 は、プール ID 列 2901 と、ページ ID 列 2902 と、I/O モニタ ON/OFF 列 2903 と、によって構成される。

【0172】

プール ID 列 2901 には、ボリュームプール 700 を特定・識別するための識別子（識別情報）が登録される。

【0173】

ページ ID 列 2902 には、各ボリュームプール 700 の各ページを特定・識別するための識別子（識別情報）が登録される。

【0174】

I/O モニタ ON/OFF 列 2903 には I/O モニタ処理時に当該ページに対する I/O を記録する（ON）か否（OFF）かが登録される。なお、この情報は、図 13 における I/O モニタ ON/OFF 列 1502 と同様に、初期値としてデフォルト値が設定されているが、システムを順次運用していくうちに、図 29 の処理によって値（ON/OFF）は変化するようになっている。

【0175】

< I/O モニタ処理 >

図28は、第2の実施形態において、仮想ボリュームマネージャ704によって実行されるI/Oモニタ処理(例)を説明するためのフローチャートである。本処理はホスト計算機100からストレージ装置400内の仮想ボリューム701に対してI/Oが発行されたことを契機として開始される(ステップ3000)。

【0176】

まず、仮想ボリュームマネージャ704は、ページI/OモニタON/OFF管理テーブル2900を参照し、I/O発行先のページのI/OモニタON/OFF2903がONになっているかを確認する(ステップ3001)。このときI/OモニタON/OFF2903がOFFであれば処理は終了する(ステップ3003)。I/OモニタON/OFF2903がONであれば処理はステップ3002に移行する。

10

【0177】

そして、仮想ボリュームマネージャ704は、ボリュームプール管理テーブル1200の、I/O発行先のページのページIDおよびボリュームプールのプールIDで決定される行のストレージI/O数列1206の数字を1増加させ(ステップ3002)、処理を終了する(ステップ3003)。

【0178】

< I/OモニタのON/OFF制御処理 >

図29は、管理ソフトウェア600が実行するI/OモニタのON/OFF制御処理(例)を説明するための示すフローチャートである。本処理は定期的に、ホスト計算機100毎に実行開始される(ステップ3100)。なお、「定期的」の意義については、第1の実施形態において上述した通りである。以下ステップごとに説明を記述する。

20

【0179】

ステップ3101：管理ソフトウェア600は、ホスト計算機100上のOS501から、タスク稼働情報2200(図20参照)を取得する。

【0180】

ステップ3102：管理ソフトウェア600は、ホスト計算機100に割り当てられたページ群を特定する。この処理では、まず管理ソフトウェア600が、ホスト計算機100上のOS501からホスト計算機100の使用する仮想ボリューム701上のLBA範囲を取得する。この場合、ホスト計算機100のOS501は、そのホスト計算機の使用する仮想ボリューム701上のLBA範囲を管理している。また、ホスト計算機100がホストボリュームマネージャにより仮想ボリューム701上の一部領域を仮想的なホストボリュームとして利用している場合は、管理ソフトウェア600は、当該領域に対応するLBA範囲をホストボリュームマネージャから取得する。次に、管理ソフトウェア600は、取得したLBA範囲を、仮想ボリューム管理テーブル1100(図9参照)を用いて照合することで、ホスト計算機100に割り当てられたページ群を特定することができる。

30

【0181】

ステップ3103：管理ソフトウェア600は、ステップ3101で取得したタスク稼働情報2200において、「稼働状態2202がON」かつ「I/Oモニタ要否管理テーブル2000においてI/Oモニタ要否列2004の値が「否」である」ようなタスクが1つでも存在するかを判定する。ステップ3103の判定結果がyesの場合、処理はステップ3104に移行し、当該判定結果がnoの場合、処理はステップ3105に移行する。なお、条件として「稼働状態2202がON」の代わりに「ホストI/O発行量2203が閾値以上であるもの」としてもよい。また「1つでも存在するかを判定する」としたが「閾値以上存在するかを判定する」などとしてもよい。当該変形例については、第1の実施形態の場合と同様である。

40

【0182】

ステップ3104：管理ソフトウェア600は、ステップ3102で特定したページ群のI/OモニタをOFFにする。すなわち、ストレージ装置400に対してボリュームプールの識別子およびページ群のページIDを含むI/OモニタのOFF命令を送信する。

50

そして、それを受信したストレージ装置400が、ページI/OモニタON/OFF管理テーブル2900(図27)の該当するプールIDかつページIDのI/OモニタON/OFF列2903の値をOFFにする。

【0183】

ステップ3105:管理ソフトウェア600は、ステップ3102で特定したページ群のI/OモニタをONにする。すなわち、ストレージ装置400に対してボリュームプールのプールIDおよびページ群のページIDを含むI/OモニタのON命令を送信する。そして、それを受信したストレージ装置400が、ページI/OモニタON/OFF管理テーブル2900の該当するプールIDかつページIDのI/OモニタON/OFF列2903の値をONにする。

10

【0184】

(3)第3の実施形態

続いて、本発明の第3の実施形態による計算機システムについて、第1の実施形態で用いた図1から図18、及び図20から図26、さらに新たに図30を参照して、説明する。なお、第3の実施形態と第1の実施形態との違いは、I/OモニタのON/OFF制御処理が管理サーバ200の管理ソフトウェア600によって実行されるのではなく、各ホスト計算機100上のOS501によって、或いは、代表のホスト計算機のOS501によって実行されることである。従って、管理ソフトウェア600の役割は、ユーザに対してI/OモニタON/OFF判断用情報表示画面2300(図21)を提示することとなる。

20

【0185】

本実施形態におけるシステム構成は、第1の実施形態において図1から図5を用いて説明したものと同様である。また図6から図18、図20から図26に示す概念図、管理テーブル、フローチャートは、本実施形態にも適用される。ただし、I/OモニタON/OFF判断用情報表示画面2300を提示する必要がない場合、管理ソフトウェア600は必ずしも必要ない。また、I/Oモニタ要否管理テーブル2000は、管理ソフトウェア600が保持するとともに、各々のホスト計算機100が各自ホストIDの部分情報(タスク名、タスク種別、I/Oモニタ要否)を保持する。管理ソフトウェア600の保持するI/Oモニタ要否管理テーブル2000と、各ホスト計算機100の保持するI/Oモニタ要否管理テーブルの部分情報は同期される。このため、I/Oモニタ要否管理テーブル2000をユーザが入力する場合、管理ソフトウェア600上で入力してもよいし、各ホスト計算機100上で入力してもよい。

30

【0186】

<I/OモニタのON/OFF制御処理>

図30は、I/OモニタのON/OFF制御処理の一例を示すフローチャートである。本処理は定期的に、各ホスト計算機100上のOS501により実行開始される(ステップ3200)。「定期的」の意義については上述の通りである。以下ステップごとに説明を記述する。

【0187】

ステップ3201:各ホスト計算機100上のOS501は、各ホスト計算機(自ホスト)のタスク稼働情報2200(図20)を参照する。

40

【0188】

ステップ3202:各ホスト計算機100上のOS501は、タスク稼働情報2200において、「稼働状態2202がON」かつ「I/Oモニタ要否管理テーブル2000においてI/Oモニタ要否列2004の値が「否」である」ようなタスクが1つでも存在するかを判定する。ステップ3202の判定結果がyesの場合、処理はステップ3203に移行し、当該判定結果がnoの場合、ステップ3204に移行する。なお、タスクの前提条件として「稼働状態がON」の代わりに「ホストI/O発行量2203が閾値以上である」としてもよい。また「1つでも存在するかを判定する」としたが「閾値以上存在するかを判定する」などとしてもよい。当該変形例については、第1の実施形態の場合と同

50

様である。

【 0 1 8 9 】

ステップ 3 2 0 3 : 各ホスト計算機 1 0 0 上の OS 5 0 1 は、該当する各ホスト計算機 (自ホスト計算機) に割り当てられたボリュームの I / O モニタを OFF にする。この処理は、例えば SCSI Inquiry により、仮想ボリューム ID を含む I / O モニタの OFF 命令が、ストレージ装置 4 0 0 に発行される。そして、この OFF 命令を受信したストレージ装置 4 0 0 は仮想ボリューム I / O モニタ ON / OFF 管理テーブル 1 5 0 0 (図 1 3) の該当する仮想ボリューム ID の I / O モニタ ON / OFF 列 1 5 0 2 の値を OFF にする。

【 0 1 9 0 】

ステップ 3 2 0 4 : 各ホスト計算機 1 0 0 上の OS 5 0 1 は、該当する各ホスト計算機 (自ホスト計算機) に割り当てられたボリュームの I / O モニタを ON にする。この処理は、例えば SCSI Inquiry により、仮想ボリューム ID を含む I / O モニタの ON 命令が、ストレージ装置 4 0 0 に発行される。そして、この ON 命令を受信したストレージ装置 4 0 0 は仮想ボリューム I / O モニタ ON / OFF 管理テーブル 1 5 0 0 の該当する仮想ボリューム ID の I / O モニタ ON / OFF 列 1 5 0 2 の値を ON にする。

【 0 1 9 1 】

(4) 第 4 の実施形態

さらに続いて、本発明の第 4 の実施形態による計算機システムについて、第 1 の実施形態で用いた図 1 から図 1 2、図 1 5 から 1 8、及び図 2 0 から図 2 6、第 2 の実施形態で用いた図 2 7 及び 2 8、さらに新たに図 3 1 を参照して、説明する。なお、第 1 の実施形態と第 4 の実施形態との違いは、「 I / O モニタの ON / OFF 制御を管理ソフトウェア 6 0 0 が実行するのではなく、各ホスト計算機 1 0 0 上の OS 5 0 1 が実行すること」および「 I / O モニタの ON / OFF 制御はページ単位で実行されること」である。つまり、第 2 の実施形態と第 3 の実施形態とを組み合わせたものとなっている。従って、管理ソフトウェア 6 0 0 の役割は、ユーザに対して I / O モニタ ON / OFF 判断用情報表示画面 2 3 0 0 を提示することとなる。

【 0 1 9 2 】

本実施形態におけるシステム構成は、第 1 の実施形態において図 1 から図 5 を用いて説明したものと同様である。また、図 6 から図 1 8、図 2 0 から図 2 6 に示す概念図、管理テーブル、フローチャートは、本実施形態にも適用される。ただし、 I / O モニタ ON / OFF 判断用情報表示画面 2 3 0 0 を提示する必要がない場合、管理ソフトウェア 6 0 0 は必ずしも必要ない。また、 I / O モニタ要否管理テーブル 2 0 0 0 は、管理ソフトウェア 6 0 0 が保持するとともに、各々のホスト計算機 1 0 0 が各自ホスト ID の部分情報 (タスク名、タスク種別、 I / O モニタ要否) を保持することとなる。また、管理サーバ 2 0 0 が保持し、管理ソフトウェア 6 0 0 によって使用される I / O モニタ要否管理テーブル 2 0 0 0 と、各ホスト計算機 1 0 0 が保持する I / O モニタ要否管理テーブルの部分情報は同期される。このため、 I / O モニタ要否管理テーブル 2 0 0 0 をユーザが入力する場合、管理サーバ 2 0 0 の管理ソフトウェア 6 0 0 上で入力してもよいし、各ホスト計算機 1 0 0 上で入力してもよい。さらに、仮想ボリューム 7 0 1 をそのままホストボリュームとして利用するのではなく、1 つ以上の仮想ボリューム 7 0 1 上の、一部または複数の領域を、仮想的なホストボリュームとしてホスト計算機 1 0 0 上で利用するためのプログラム (ホストボリュームマネージャ) が、ホスト計算機 1 0 0 上や別の計算機上に存在してもよい。なお、ホストボリュームマネージャは、当該仮想的なホストボリュームと、当該一部または複数の領域の間のマッピング情報を保持する。

【 0 1 9 3 】

< I / O モニタの ON / OFF 制御処理 >

図 3 1 は、第 4 の実施形態による I / O モニタの ON / OFF 制御処理 (例) を説明するためのフローチャートである。本処理は定期的に、各ホスト計算機 1 0 0 上の OS 5 0 1 により実行開始される (ステップ 3 3 0 0)。「定期的」の意義については、上述の通

10

20

30

40

50

りである。以下ステップごとに説明を記述する。

【0194】

ステップ3301：各ホスト計算機100上のOS501は、該当する各ホスト計算機（自ホスト計算機）のタスク稼働情報2200（図20）を参照する。

【0195】

ステップ3302：各ホスト計算機100上のOS501は、タスク稼働情報2200において、「稼働状態2202がON」かつ「I/Oモニタ要否管理テーブル2000においてI/Oモニタ要否列2004の値が「否」である」ようなタスクが1つでも存在するかを判定する。ステップ3302の判定結果がyesの場合、処理はステップ3303に移行し、当該判定結果がnoの場合、処理はステップ3304に進む。なお、タスクの前提条件として「稼働状態がON」の代わりに「ホストI/O発行量2203が閾値以上である」としてもよい。また「1つでも存在するかを判定する」としたが「閾値以上存在するかを判定する」などとしてもよい。当該変形例については、第1の実施形態の場合と同様である。

【0196】

ステップ3303：各ホスト計算機100上のOS501は、該当する各ホスト計算機（自ホスト計算機）に割り当てられた、仮想ボリューム701上の1つ以上の領域のI/OモニタをOFFにする。この処理では、例えばSCSI Inquiryにより、仮想ボリュームIDおよびLBA範囲を含んだI/OモニタOFF命令をストレージ装置400に送信する。このOFF命令を受信したストレージ装置400は、仮想ボリューム管理テーブル1100（図9）を参照して、指示された仮想ボリュームIDおよびLBA範囲に対応する、プールIDおよびページIDを特定し、この特定したプールIDおよびページIDをもとに、ページI/OモニタON/OFF管理テーブル2900（図27）のI/OモニタON/OFF列2903の値をOFFにする。この場合、OS501は、該当する各ホスト計算機（自ホスト計算機）が使用する仮想ボリューム701上のLBA範囲を管理している。なお、ホスト計算機100がホストボリュームマネージャにより仮想ボリューム701上の一部領域を仮想的なホストボリュームとして利用している場合は、ホスト計算機100上のOS501は、当該仮想的なホストボリュームに関するI/OモニタOFF命令を、ホストボリュームマネージャに対して発行する。ホストボリュームマネージャはこの命令を受信し、例えばSCSI Inquiryにより、仮想ボリュームID、および当該仮想的なホストボリュームに対応する仮想ボリューム701上のLBA範囲、を含んだI/OモニタOFF命令をストレージ装置400に送信する。

【0197】

ステップ3304：各ホスト計算機100上のOS501は、該当する各ホスト計算機（自ホスト計算機）に割り当てられた、仮想ボリューム701上の1つ以上の領域のI/OモニタをONにする。この処理では、仮想ボリュームIDおよびLBA範囲を含んだI/OモニタON命令を、例えばSCSI Inquiryにより、ストレージ装置400に送信する。このON命令を受信したストレージ装置400は仮想ボリューム管理テーブル1100を参照して、指示された仮想ボリュームIDおよびLBA範囲に対応する、プールIDおよびページIDを特定し、この特定したプールIDおよびページIDをもとに、ページI/OモニタON/OFF管理テーブル2900のI/OモニタON/OFF列1502の値をONにする。この場合、OS501は、該当する各ホスト計算機（自ホスト計算機）が使用する仮想ボリューム701上のLBA範囲を管理している。なお、ホスト計算機100がホストボリュームマネージャにより仮想ボリューム701上の一部領域を仮想的なホストボリュームとして利用している場合は、ホスト計算機100上のOS501は、当該仮想的なホストボリュームに関するI/OモニタON命令を、ホストボリュームマネージャに対して発行する。ホストボリュームマネージャはこの命令を受信し、例えばSCSI Inquiryにより、仮想ボリュームID、および当該仮想的なホストボリュームに対応する仮想ボリューム701上のLBA範囲、を含んだI/OモニタON命令をストレージ装置400に送信する。

10

20

30

40

50

【 0 1 9 8 】

(5) まとめ

(i) 本発明の各実施形態において、ストレージ装置では、複数の計算機（ホスト計算機）上で実行されるタスクに関するアクセス特性（例えば、I / O数）に基づいて再配置処理が実行される。この際、管理計算機或いはホスト計算機は、仮想論理ボリューム単位のI / O数、仮想ボリュームを構成するページ単位（仮想論理ボリュームの実記憶領域に対するアクセス単位）のI / O数、或いは、ホスト計算機で発行される（タスクを実行するホスト計算機によるストレージサブシステムに対するアクセス単位で）I / O数をモニタすることにより、再配置処理実行の有無を判断している。ここで、本実施形態では、I / Oモニタを実行するか否かを示すI / OモニタON / OFF管理情報（図13及び27参照）が設定される。そして、ストレージ装置は、このI / OモニタON / OFF管理情報を参照し、モニタされたホスト計算機からのI / O数に基づいて再配置処理を実行する。実施形態では、I / Oモニタをするか否かの情報（I / OモニタON / OFF管理情報）を設定しているが、I / O数（アクセス特性）をモニタしたとしてもそのI / O数を再配置処理の要否判断に用いなければ良いので、I / OモニタON / OFF管理情報ではなく、単に「再配置処理における移動先決定の判断基準として考慮するか否かを示す再配置基準情報を設定する」と表現することができる。このようにすることにより、所定の仮想ボリューム、ページやホストのボリュームについて再配置処理の要否判断に用いる必要がなくなるので、不必要な再配置処理を防止することができるようになる。

10

【 0 1 9 9 】

また、本発明の各実施形態では、管理計算機或いはホスト計算機は、複数のホスト計算機（各ホスト計算機が管理計算機の機能を有する場合（第3の実施形態）には自身のホスト計算機）で実行されるタスクが使用する記憶領域毎のデータの再配置発生量（少なくとも階層格上げの容量）を示す情報（再配置発生情報：ホスト内タスクI / O比率や全ホスト階層格上げページI / O寄与率を含む）を生成し、これを出力する（図21参照）ようにしている。このようにすることにより、ユーザ（管理者）は、出力された情報を参照して、各タスクについてI / Oモニタ要否（再配置処理の要否）を決定することができる。従って、各タスクの実行の傾向に合致して、ユーザはI / Oモニタ要否を決定できるため、ユーザにとって管理しやすい計算機システム（ストレージシステム）を提供することが可能となる。

20

30

【 0 2 0 0 】

一方、I / Oモニタ要否（再配置処理の要否）を自動的に設定、或いは、モニタOFFにすることが推奨されるタスクをユーザに提示するようにしても良い。このようにすることにより、I / Oモニタ要否判断に精通していないユーザも適切にストレージシステムを管理することができるようになる。なお、具体的には、モニタOFF推奨タスク表示は、図24及び26に示す処理を実行することにより、実現することができる。

【 0 2 0 1 】

(ii) 本発明は、実施形態の機能を実現するソフトウェアのプログラムコードによっても実現できる。この場合、プログラムコードを記録した記憶媒体をシステム或は装置に提供し、そのシステム或は装置のコンピュータ（又はCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出す。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコード自体、及びそれを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク、CD-ROM、DVD-ROM、ハードディスク、光ディスク、光磁気ディスク、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどが用いられる。

40

【 0 2 0 2 】

また、プログラムコードの指示に基づき、コンピュータ上で稼動しているOS（オペレーティングシステム）などが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。さらに、記憶媒体から読み出された

50

プログラムコードが、コンピュータ上のメモリに書きこまれた後、そのプログラムコードの指示に基づき、コンピュータのCPUなどが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。

【0203】

さらに、実施の形態の機能を実現するソフトウェアのプログラムコードを、ネットワークを介して配信することにより、それをシステム又は装置のハードディスクやメモリ等の記憶手段又はCD-RW、CD-R等の記憶媒体に格納し、使用時にそのシステム又は装置のコンピュータ（又はCPUやMPU）が当該記憶手段や当該記憶媒体に格納されたプログラムコードを読み出して実行するようにしてもよい。

【0204】

最後に、ここで述べたプロセス及び技術は本質的に如何なる特定の装置に関連することはなく、コンポーネントの如何なる相応しい組み合わせによってでも実装できることを理解する必要がある。更に、汎用目的の多様なタイプのデバイスがここで記述した教授に従って使用可能である。ここで述べた方法のステップを実行するのに、専用の装置を構築するのが有益であることが判るかもしれない。また、実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態にわたる構成要素を適宜組み合わせてもよい。本発明は、具体例に関連して記述したが、これらは、すべての観点に於いて限定の為ではなく説明の為である。本分野にスキルのある者には、本発明を実施するのに相応しいハードウェア、ソフトウェア、及びファームウェアの多数の組み合わせがあることが解るであろう。例えば、記述したソフトウェアは、アセンブラ、C/C++、perl、Shell、PHP、Java（登録商標）等の広範囲のプログラム又はスクリプト言語で実装できる。

【0205】

さらに、上述の実施形態において、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示しているとは限らない。全ての構成が相互に接続されていてもよい。

【0206】

加えて、本技術分野の通常の知識を有する者には、本発明のその他の実装がここに開示された本発明の明細書及び実施形態の考察から明らかになる。記述された実施形態の多様な態様及び/又はコンポーネントは、データを管理する機能を有するコンピュータ化ストレージシステムに於いて、単独又は如何なる組み合わせでも使用することが出来る。明細書と具体例は典型的なものに過ぎず、本発明の範囲と精神は後続する請求範囲で示される。

【符号の説明】

【0207】

- 100・・・ホスト計算機
- 200・・・管理サーバ（管理計算機）
- 300・・・SAN
- 301・・・LAN
- 400・・・ストレージ装置
- 700・・・ボリュームプール
- 701・・・仮想ボリューム（仮想論理ボリューム）
- 800乃至803・・・物理ディスク
- 1100・・・仮想ボリューム管理テーブル（仮想ボリューム管理情報）
- 1200・・・ボリュームプール管理テーブル（ボリュームプール管理情報）
- 1300・・・プール階層管理テーブル（プール階層管理情報）
- 1400・・・未使用領域管理テーブル（未使用領域管理情報）
- 1500・・・仮想ボリュームI/OモニタON/OFF管理テーブル（仮想ボリュームI/OモニタON/OFF管理情報）

10

20

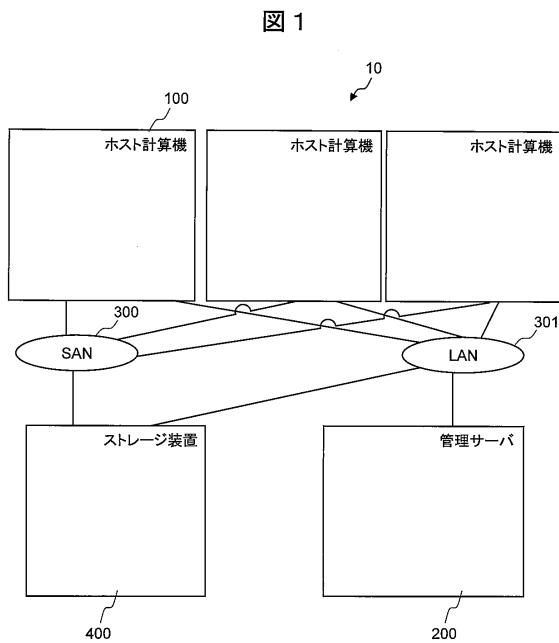
30

40

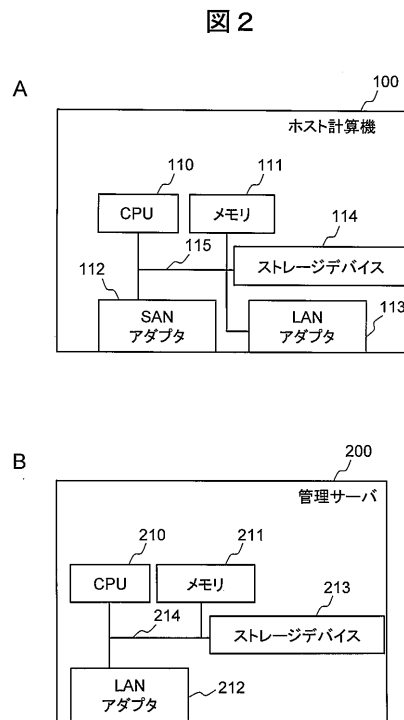
50

- 1800・・・再配置ページ管理テーブル（再配置ページ管理情報）
- 1900・・・ホストボリューム管理テーブル（ホストボリューム管理情報）
- 2000・・・I/Oモニタ要否管理テーブル（I/Oモニタ要否管理情報）
- 2200・・・タスク稼働情報
- 2400・・・再配置容量管理テーブル（再配置容量管理情報）
- 2500・・・タスク稼働情報管理テーブル（タスク稼働管理情報）
- 2700・・・タスク種別毎I/OモニタOFF候補管理テーブル（タスク種別毎I/OモニタOFF候補管理情報）
- 2900・・・ページI/OモニタON/OFF管理テーブル（ページI/OモニタON/OFF管理情報）

【図1】

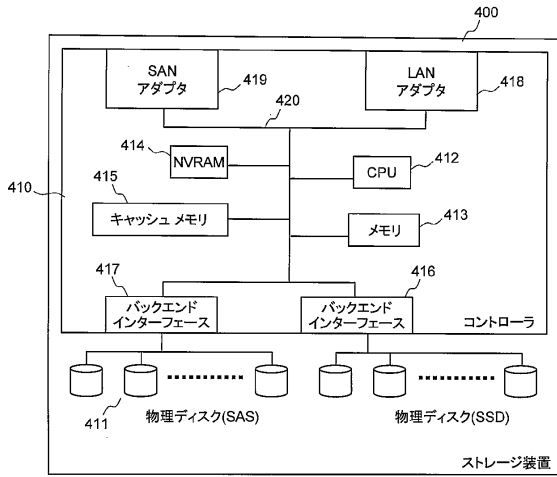


【図2】



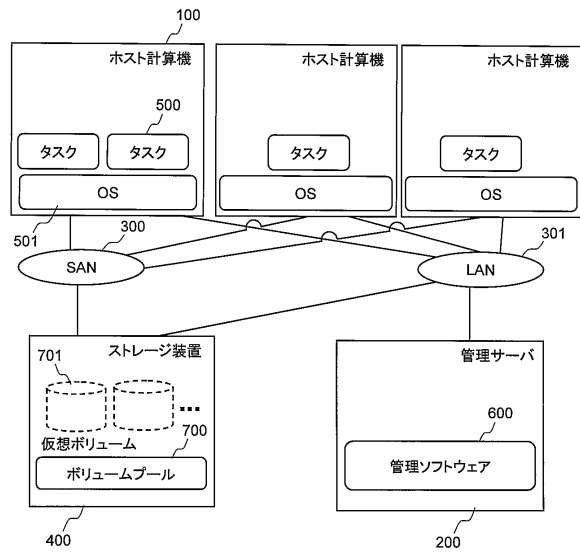
【図3】

図3



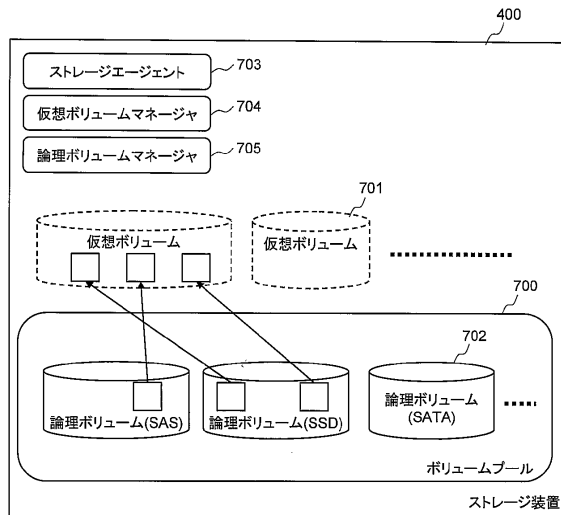
【図4】

図4



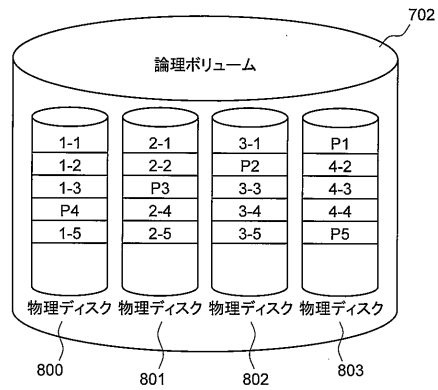
【図5】

図5



【図6】

図6

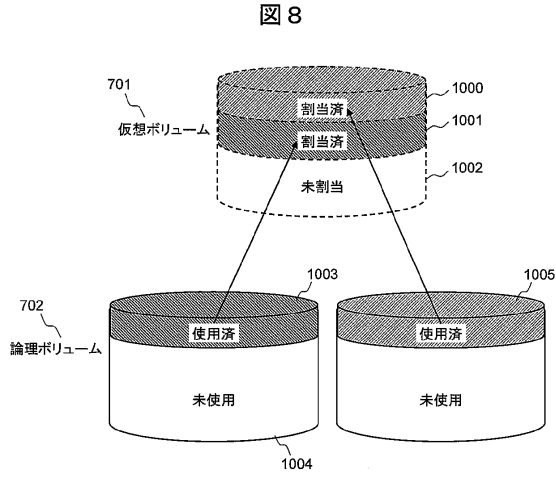


【図7】

図7

論理ボリュームID	ディスク	RAIDレベル	ストライプサイズ	ディスク種別
0	1, 2, 3, 4	5	32KB	SAS
1	5, 6, 7, 8	5	32KB	SSD
⋮	⋮	⋮	⋮	⋮
k	m, m+1, ...	5	32KB	SAS

【図8】



【図9】

図9

仮想ボリュームID	開始LBA	終了LBA	プールID	ページID
0	0x00000000	0x0001af0f	pool 1	0
	0x0001af10	0x0001ffff		1
	0x00020000	0x0020ffff		2
	0x00021000	0x0008ffff		4
	⋮	⋮		⋮
1	0x00000000	0x0001af0f	pool 2	0
	0x0001af10	0x0001ffff		1
	0x00020000	0x0020ffff		2
	0x00021000	0x0008ffff		3
	⋮	⋮		⋮
⋮	⋮	⋮	⋮	⋮

【図10】

図10

プールID	ページID	論理ボリュームID	開始LBA	終了LBA	ストレージI/O数
pool 1	0	0	0x00000000	0x0001af0f	4000
	1	0	0x00030000	0x000350ef	1000
	2	0	0x00050000	0x000500ff	5000
	3	1	0x00000000	0x0006efff	1000
	⋮	⋮	⋮	⋮	⋮
pool 2	0	5	0x00000000	0x0001a000	0
	1	5	0x00030000	0x000320ef	0
	2	5	0x00050000	0x000500fe	0
	⋮	⋮	⋮	⋮	⋮

【図12】

図12

論理ボリュームID	開始LBA	終了LBA
0	0x0001af10	0x0002ffff
	0x000350f0	0x0004ffff
⋮	⋮	⋮

【図11】

図11

プールID	階層	論理ボリュームID
pool1	Tier0	1
		2
	Tier1	0
		3
⋮	⋮	⋮

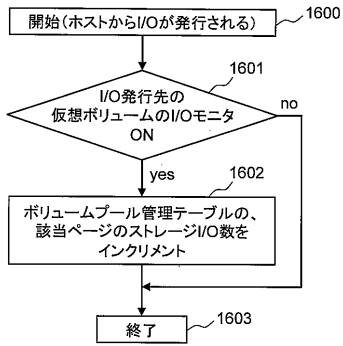
【図13】

図13

仮想ボリュームID	I/OモニタON/OFF
0	ON
1	OFF
⋮	⋮

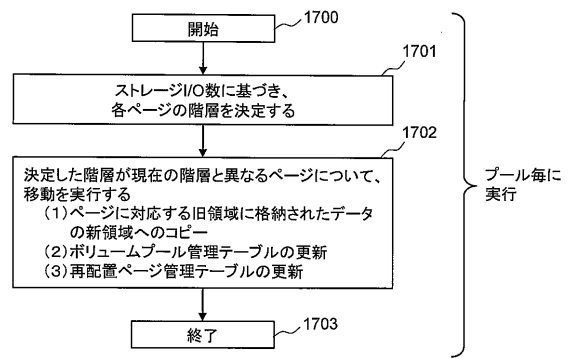
【図14】

図14



【図15】

図15



【図16】

図16

プールID	ページID	移動元階層	移動先階層
pool1	0	Tier1	Tier0
	1	Tier0	Tier1
	2	Tier1	Tier0
	⋮	⋮	⋮
pool2	0	Tier1	Tier0
	1	Tier1	Tier0
	2	Tier0	Tier1
	⋮	⋮	⋮

【図17】

図17

ホストID	ホスト WWN	ホスト ボリュームID	ストレージ ボリュームID
host1	WWN1-1	0	仮想ボリューム0
	WWN1-2	1	仮想ボリューム2
	⋮	⋮	⋮
host2	WWN2	0	仮想ボリューム1
	WWN2	1	論理ボリューム3
	⋮	⋮	⋮
host3	WWN3	0	仮想ボリューム4
⋮	⋮	⋮	⋮

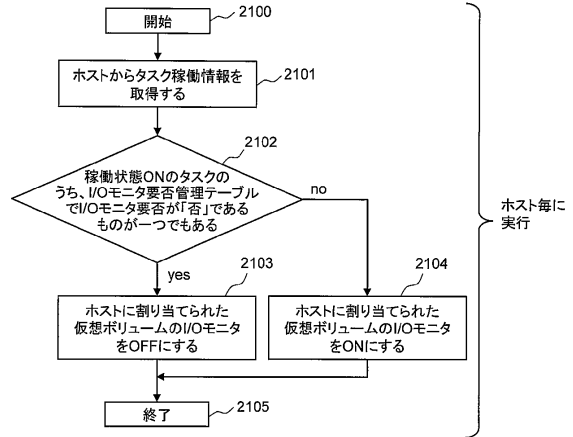
【図18】

図18

ホストID	タスク名	タスク種別	I/Oモニタ要否
host1	夜間集計バッチ1	バッチ	否
host1	商品インデックス構築	インデックス構築	否
host1	OLTP1	OLTP	要
host2	OLTP2	OLTP	要
host3	OLTP3	OLTP	要
host3	アンチウイルスA	アンチウイルス	否

【図19】

図19



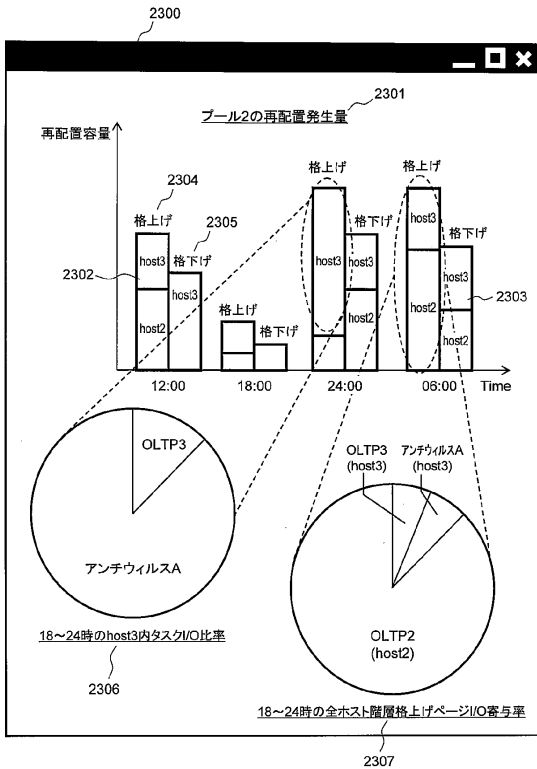
【図20】

図20

タスク名	稼働状態	ホストI/O発行量
夜間集計バッチ	ON	10000
商品インデックス構築	ON	5000
OLTP0	OFF	0

【図21】

図21



【図22】

図22

プールID	ホストID	再配置時間帯	階層格上げ容量	階層格下げ容量
pool1	host1	12:00-16:00	10GB	10GB
		16:00-20:00	10GB	5GB
		20:00-24:00	30GB	30GB
		⋮	⋮	⋮
pool2	host2	12:00-18:00	20GB	0GB
		18:00-24:00	5GB	0GB
		24:00-6:00	10GB	20GB
	host3	6:00-12:00	30GB	15GB
		12:00-18:00	10GB	25GB
		18:00-24:00	5GB	5GB
⋮	⋮	⋮	⋮	

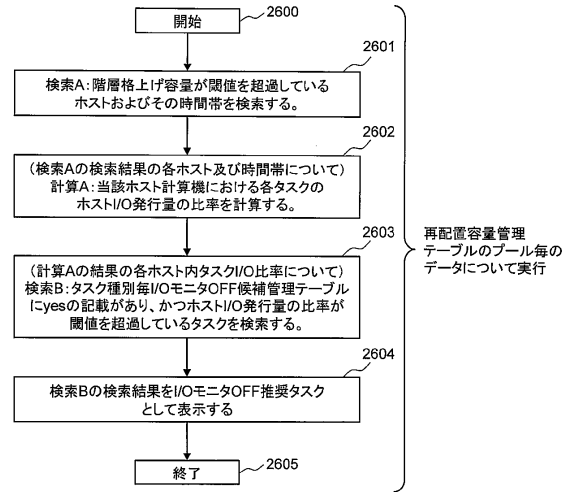
【図23】

図23

2501 ホストID	2502 タスク名	2503 稼働状態	2504 ホストI/O発行量	2505 採取時刻
host1	夜間集計バッチ	OFF	0	12:00
	インデックス構築	ON	500	12:00
	OLTP1	ON	3000	12:00
host2	OLTP2	ON	5000	12:00
		ON	10000	15:00
		ON	6000	18:00
host3	OLTP3	ON	3000	12:00
		ON	2000	15:00
		ON	1000	18:00
		OFF	0	21:00
	アンチウイルスA	ON	200	12:00
		ON	300	15:00
		ON	1000	18:00
			10000	21:00

【図24】

図24



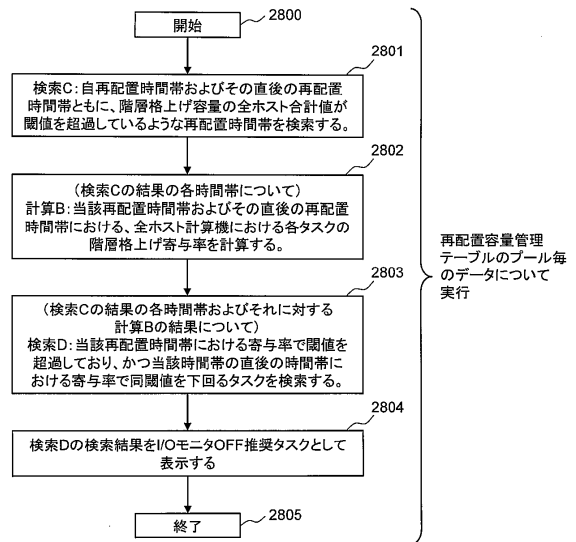
【図25】

図25

2701 タスク種別	2702 I/OモニタOFF候補
バッチ	yes
インデックス構築	yes
OLTP	no
アンチウイルス	yes
⋮	⋮

【図26】

図26



【図 27】

図 27

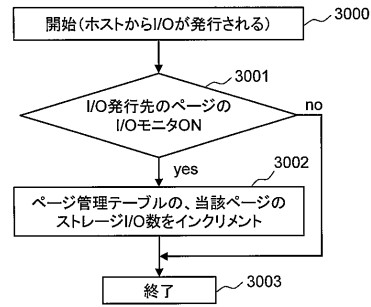
2900

プールID	ページID	I/OモニタON/OFF
pool1	0	ON
	1	ON
	2	ON
	⋮	⋮
	k	OFF
pool2	0	OFF
	1	OFF
	2	OFF
	⋮	⋮

2901 2902 2903

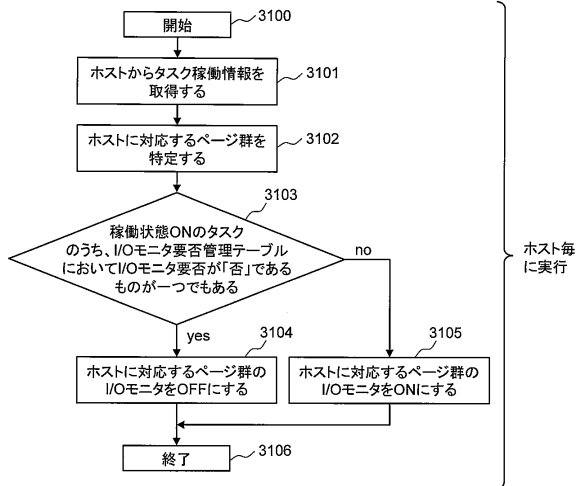
【図 28】

図 28



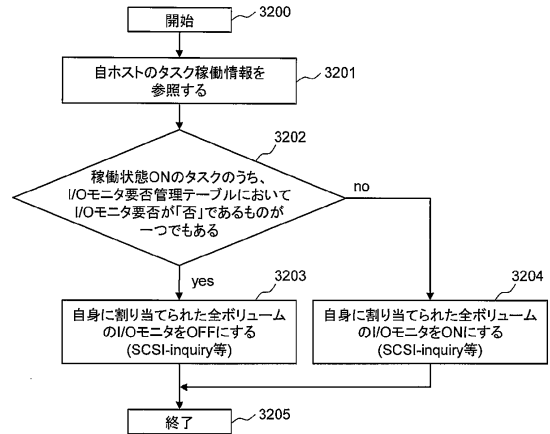
【図 29】

図 29

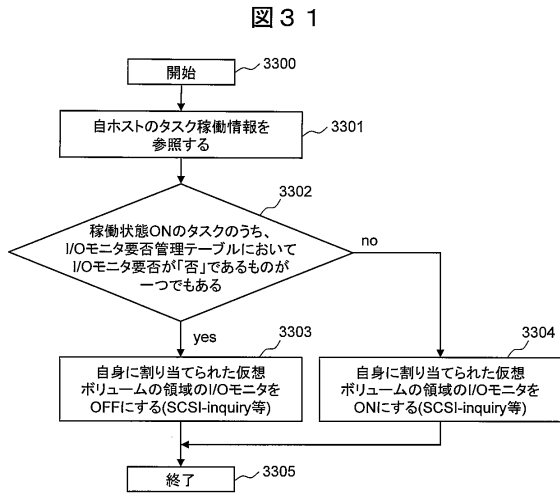


【図 30】

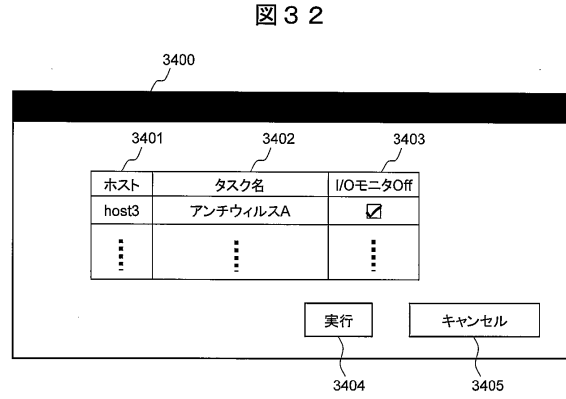
図 30



【図31】



【図32】



フロントページの続き

(72)発明者 原 純一

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

審査官 桜井 茂行

(56)参考文献 国際公開第2011/096017(WO, A1)

特開2007-066259(JP, A)

特開2003-015915(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 12/00

G06F 17/30