

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2011 年 6 月 30 日(30.06.2011)

(10) 国際公開番号
WO 2011/077489 A1

- (51) 国際特許分類:
G06F 3/06 (2006.01)
- (21) 国際出願番号: PCT/JP2009/007207
- (22) 国際出願日: 2009 年 12 月 24 日(24.12.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人 (米国を除く全ての指定国について): 株式会社日立製作所(HITACHI, LTD.) [JP/JP]; 〒1008280 東京都千代田区丸の内一丁目 6 番 6 号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 折笠大典(ORIKASA, Daisuke) [JP/JP]; 〒2500872 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内 Kanagawa (JP). 松田健(MATSUDA, Ken) [JP/JP]; 〒2500872 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内 Kanagawa (JP). 高田豊(TAKATA, Yutaka) [JP/JP]; 〒2500872 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内 Kanagawa (JP). 江口賢哲(EGUCHI, Yoshiaki) [JP/JP]; 〒2150013 神奈川県川崎市麻生区王禅寺

1 0 9 9 番地 株式会社日立製作所 システム開発研究所内 Kanagawa (JP). 里山愛(SATOYAMA, Ai) [JP/JP]; 〒2150013 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内 Kanagawa (JP). 水野陽一(MIZUNO, Yoichi) [JP/JP]; 〒2150013 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内 Kanagawa (JP).

(74) 代理人: 特許業務法人ウィルフォート国際特許事務所 (WILLFORT INTERNATIONAL); 〒1010044 東京都千代田区鍛冶町二丁目 3 番 2 号 神田センタービルディング 5 階 Tokyo (JP).

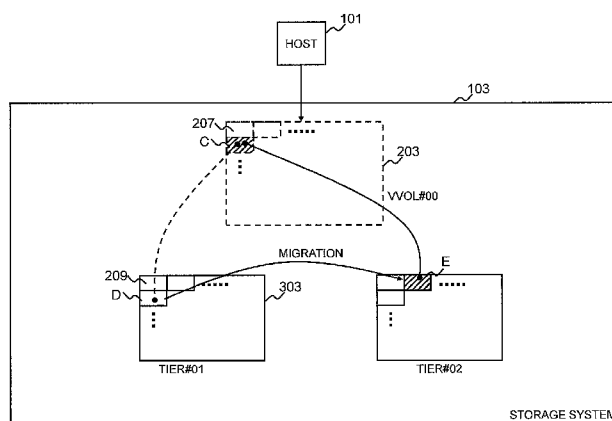
(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: STORAGE SYSTEM FOR PROVIDING VIRTUAL VOLUME

(54) 発明の名称: 仮想ボリュームを提供するストレージシステム

FIG. 1



(57) Abstract: A plurality of storage area groups which are obtained by classifying a plurality of storage areas provided from a plurality of storage devices in accordance with attributes of the storage areas are managed. In response to a write request to at least one address included in a plurality of addresses in a logical volume, a plurality of logical volumes to which at least one storage area included in the plurality of storage areas is assigned are provided. In accordance with an access situation of the at least one address in the logical volume, data which is written to the at least one address in response to the write request is shifted from the at least one storage area included in one of the plurality of storage area groups to at least one storage area in another storage area group included in the plurality of storage area groups.

(57) 要約:

[続葉有]



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ,

CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

複数の記憶デバイスから提供される複数の記憶領域を記憶領域の属性に従って分類された複数の記憶領域グループを管理する。論理ボリューム内の複数のアドレスに含まれる少なくとも1つのアドレスへのライト要求に応じて前記複数の記憶領域に含まれる少なくとも1つの記憶領域が割り当てられる前記論理ボリュームの複数の提供。前記論理ボリューム内の前記少なくとも1つのアドレスのアクセス状況に応じて、前記ライト要求によって前記少なくとも1つのアドレスにライトされているデータを、前記複数の記憶領域グループの1つに含まれる前記少なくとも1つの記憶領域から前記複数の記憶領域グループに含まれる他の記憶領域グループ内の少なくともひとつの記憶領域へ移行する。

明 細 書

発明の名称： 仮想ボリュームを提供するストレージシステム

技術分野

[0001] 本発明は、Thin Provisioning（Dynamic Provisioningとも呼ばれる）が適用される記憶制御に関する。

背景技術

[0002] データ移行技術が知られている。データ移行は、一般に、ボリューム単位で行われる。例えば、特許文献1によれば、第1の論理ボリュームに格納されている全てのデータが、第1の論理ボリュームから第2の論理ボリュームに移行される。

[0003] また、ストレージ階層化技術も知られている。例えば、特許文献2によれば、複数のストレージシステムが有する複数の論理ボリュームが、それぞれ、複数のティアのうちのいずれかのティアに分類され、或るティアに属する論理ボリュームが、別のティアに移行される。複数のティアには、例えば、高信頼性のティア、低コストのティアなどが含まれる。

先行技術文献

特許文献

[0004] 特許文献1：特開2006-302077号公報

特許文献2：特開2008-047156号公報

発明の開示

発明が解決しようとする課題

[0005] ところで、Thin Provisioningと呼ばれる（Dynamic Provisioningとも呼ばれる）技術が知られている。Thin Provisioningによれば、複数の仮想的な記憶領域（仮想ページ）で構成された仮想的な論理ボリューム（仮想ボリューム）と、複数の実体的な記憶領域（実ページ）とが用意される。仮想ページに対するライトが行われた場合に、その仮想ページに実ページが割り当てられ、割り当てられた実ページに、ライト対象のデータが書き込まれる。

- [0006] 上述したデータ移行技術及びストレージ階層化技術がThin Provisioningに適用されたとする。また、第1のティアに第1の仮想ボリュームが属し第2のティアに第2の仮想ボリュームが属しているとする。この場合、第1のティアに属する第1の仮想ボリュームから第2のティアに属する第2の仮想ボリュームに全てのデータを移行することができる。
- [0007] この技術によれば、第1の仮想ボリューム内の全てのデータが、第1のティアから第2のティアに移行されるが、全てのデータが必ずしも第2のティアに存在することが好ましいとは限らない。
- [0008] そこで、本発明の目的は、Thin Provisioningに従う仮想ボリューム内のデータの配置を適切にすることにある。

課題を解決するための手段

- [0009] 複数の記憶デバイスから提供される複数の記憶領域を記憶領域の属性に従って分類された複数の記憶領域グループを管理する。論理ボリューム内の複数のアドレスに含まれる少なくとも1つのアドレスへのライト要求に応じて前記複数の記憶領域に含まれる少なくとも1つの記憶領域が割り当てられる前記論理ボリュームの複数の提供する。前記論理ボリューム内の前記少なくとも1つのアドレスのアクセス状況に応じて、前記ライト要求によって前記少なくとも1つのアドレスにライトされているデータを、前記複数の記憶領域グループの1つに含まれる前記少なくとも1つの記憶領域から前記複数の記憶領域グループに含まれる他の記憶領域グループ内の少なくともひとつの記憶領域へ移行する。

図面の簡単な説明

- [0010] [図1]本発明の実施例1の概要を示す。
- [図2]ストレージシステム103の構成を示す。
- [図3]ストレージシステム103における各種記憶領域の関係を示す。
- [図4]CMPK119が記憶する情報及びコンピュータプログラムを示す。
- [図5]ティア定義テーブル401の構成を示す。
- [図6]ティア管理テーブル402の構成を示す。

- [図7] プール管理テーブル403の構成を示す。
- [図8] VVOL管理テーブル404の構成を示す。
- [図9] PVOL管理テーブル405の構成を示す。
- [図10] アロケーションテーブル407の構成を示す。
- [図11] 実ページ管理テーブル408の構成を示す。
- [図12] モニタテーブル409の構成を示す。
- [図13] 度数分布テーブル410の構成を示す。
- [図14] 移行ページテーブル411の構成を示す。
- [図15] 移行定義テーブル412の構成を示す。
- [図16] プール作成処理のフローチャートを示す。
- [図17] プール定義GUI1700を示す。
- [図18] VVOL作成処理のフローを示す。
- [図19] VVOL作成GUI1900を示す。
- [図20] 本発明の実施例5に係るVVOL管理テーブル4404を示す。
- [図21] データ要素の集約処理の概要を示す。
- [図22] 情報表示処理のフローを示す。
- [図23] 度数分布GUI2300を示す。
- [図24] ライト処理（キャッシュメモリへのライト）のフローを示す。
- [図25] デスページ処理のフローを示す。
- [図26] リード処理のフローを示す。
- [図27] I/O頻度のモニタと再配置との関係の概要を示す。
- [図28] 再配置処理のフローを示す。
- [図29] ティアレンジ決定処理の概要を示す。
- [図30] ティアレンジ決定処理のフローを示す。
- [図31] 図30のS4402のティア境界値算出処理のフローを示す。
- [図32] 本発明の実施例5に係るプール管理テーブル4403を示す。
- [図33] 本発明の実施例5に係る比較方法／比較対象の設定処理のフローを示す。

[図34] 図 28 の S 2808 での移行ページテーブル作成処理のフローを示す。

[図35] ページ単位の移行処理の概要を示す。

[図36] ページ単位の移行処理のフローを示す。

[図37] ティア閾値の変更処理のフローを示す。

[図38] 閾値オーバー対処処理のフローを示す。

発明を実施するための形態

[0011] 以下、本発明の後述する複数の実施例に基づく観点を述べる。

[0012] <観点 1>

複数の記憶デバイスから提供される複数の記憶領域を記憶領域の属性に従って分類された複数の記憶領域グループを管理し、

論理ボリューム内の複数のアドレスに含まれる少なくとも 1 つのアドレスへのライト要求に応じて前記複数の記憶領域に含まれる少なくとも 1 つの記憶領域が割り当てられる前記論理ボリュームの複数の提供し、

前記論理ボリューム内の前記少なくとも 1 つのアドレスのアクセス状況に応じて、前記ライト要求によって前記少なくとも 1 つのアドレスにライトされているデータを、前記複数の記憶領域グループの 1 つに含まれる前記少なくとも 1 つの記憶領域から前記複数の記憶領域グループに含まれる他の記憶領域グループ内の少なくともひとつの記憶領域へ移行する、ことを特徴とするストレージシステムの制御方法。

[0013] <観点 2>

観点 1 に従う方法であって、

前記複数の記憶領域グループに含まれる第一の記憶領域グループと第二の記憶領域グループとは、それぞれのグループに相応しい前記アクセス状況の程度をグループ毎に属性として有しており、前記第一と前記第二の記憶領域グループとの前記アクセス状況の程度はグループ毎に異なっているものの一部重複しており、

前記移行処理においては、前記複数の記憶領域グループの 1 つに含まれる前

記少なくとも 1 つの記憶領域が、前記第一又は前記第二の記憶領域グループに含まれており且つ前記少なくとも 1 つのアドレスの前記アクセス状況が前記アクセス状況の程度のうち前記一部重複している範囲内である場合、前記データの移行は実施されない、ことを特徴とするストレージシステムの制御方法。

[0014] <観点 3>

一以上のホストに接続されるストレージシステムが、複数種類の物理記憶デバイス群と、記憶制御装置とを備える。前記記憶制御装置が、記憶資源と、前記複数種類の物理記憶デバイス群、前記記憶資源及び前記ホストに接続されるコントローラとを有する。物理記憶デバイス群は、一以上の物理記憶デバイスで構成されている。前記複数種類の物理記憶デバイス群が、複数種類の実ページ群の基になっている。実ページ群の種類は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいている。実ページ群が、複数の実ページを有する。前記コントローラが、一以上の仮想ボリュームを前記一以上のホストに提供する。前記仮想ボリュームは、複数の仮想ページで構成される仮想的な論理ボリュームである。前記仮想ページは、仮想的な記憶領域である。前記実ページは、実体的な記憶領域である。前記記憶資源が、どの仮想ページにどの実ページが割り当てられているかを表す情報である割り当て情報と、仮想ページ毎の I / O (Input/Output) 頻度を表す情報であるモニタ情報と、実ページ群の種類毎に設定された I / O 頻度レンジを表す情報を含んだ情報である実ページ群管理情報とを記憶する (テーブル形式又はキュー形式など、情報の形式は何でもよい)。前記コントローラが、前記ホストからライトコマンドを受信し、前記ライトコマンドからライト先の仮想ページを特定し、特定した仮想ページにフリーの実ページを割り当てるよう前記割り当て情報を更新し、割り当てた実ページにライト対象のデータを格納する。前記コントローラが、仮想ページ又はその仮想ページに割り当てられる実ページに対する I / O に応じて、前記モニタ情報における、その仮想ページの I / O 頻度を更新する。前記実ページは、実体的な記憶領域であ

り、実ページに格納されるデータはその実ページの基になっている物理記憶デバイスに記憶される。前記コントローラが、以下の（Ａ）乃至（Ｃ）の処理を含んだ処理である移行処理：

（Ａ）前記モニタ情報から特定されたＩ／Ｏ頻度が所定の条件に適合していない仮想ボリューム又は仮想ページを特定する；

（Ｂ）特定された仮想ボリューム又は仮想ページに割り当てられている、第１種の実ページ群内の第１の実ページ、に格納されているデータを、第２種の実ページ群内の第２の実ページに移行する；

（Ｃ）前記第１の実ページの割当て先の仮想ページである第１の仮想ページに前記第１の実ページに代えて前記第２の実ページを割り当てるよう前記割当て情報を更新する、

を行う。前記所定の条件とは、前記第１種の実ページ群のＩ／Ｏ頻度レンジに収まることである。前記第２種の実ページ群は、前記（Ａ）で特定された仮想ボリューム又は仮想ページのＩ／Ｏ頻度にＩ／Ｏ頻度レンジが収まっている実ページ群である。前記コントローラは、Ｉ／Ｏ頻度レンジの決定処理を行い、その決定処理において、前記モニタ情報を基に、各種の実ページ群のＩ／Ｏ頻度レンジを決定する。コントローラは、例えば、少なくとも一つのプロセッサを有してよい。記憶資源は、例えば、メモリであって良い。ホストは、計算機であっても良いし、別のストレージシステム又は別の記憶制御装置であっても良い。

[0015] これにより、仮想ボリューム内のデータを、仮想ページのＩ／Ｏ頻度に応じて、ページ単位で再配置することができる。その際、データを、そのデータを格納する実ページが割り当てられている仮想ページのＩ／Ｏ頻度が収まるＩ／Ｏ頻度レンジに対応した種類の実ページ群にすることができる。

[0016] <観点４>

観点３に従うストレージシステムにおいて、前記コントローラが、前記決定処理において、以下の（ａ）及び（ｂ）の処理を行う、

（ａ）前記モニタ情報を用いて、Ｉ／Ｏ頻度と仮想ページ数との関係を表す

情報である関係情報を作成する、

(b) 前記関係情報を基に、各種実ページ群の I / O 頻度レンジを決定する

。

[0017] <観点 5>

観点 4 に従うストレージシステムにおいて、前記コントローラが、前記 (b) の処理において、以下の (b 1) 乃至 (b 3) の処理：

(b 1) 前記関係情報を用いて、前記複数種類の実ページ群から実ページが割り当てられている一以上の仮想ボリューム内の全ての仮想ページの I / O 頻度の合計である I / O 頻度合計を算出する；

(b 2) 各種実ページ群の性能値の比と、前記 I / O 頻度合計とを基に、実ページ群の境界の I / O 頻度である境界値を実ページ群の境界毎に算出する；

(b 3) 各種実ページ群の I / O 頻度と実ページ群の境界毎の境界値とを基に、各種実ページ群の I / O 頻度レンジを決定する、
を行う。実ページ群の前記性能値は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいて決定された、性能に関する値である。前記実ページ群の境界とは、実ページ群とその上位及び／又は下位の実ページ群との境である。各種実ページ群の I / O 頻度レンジの最大値及び／又は最小値は、その実ページ群とその上位及び／又は下位の実ページ群との境の境界値に基づいて決定された値である。

[0018] <観点 6>

観点 3 乃至 5 のうちのいずれかの観点に従うストレージシステムにおいて、前記 I / O 頻度レンジの最大値は、その I / O 頻度レンジに対応した大きい方の境界値より大きい、及び／又は、その I / O 頻度レンジの最小値は、その I / O 頻度レンジに対応した小さい方の境界値より小さい。

[0019] <観点 7>

観点 3 又は 4 に従うストレージシステムにおいて、実ページ群の前記性能値は、その実ページ群における、前記移行処理での移行先となり得る実ペー

ジの数と、その実ページ群内の一つの実ページあたりの性能値との掛け算により算出された値である。実ページ群内の一つの実ページあたりの性能値は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいて決定された、性能に関する値である。

[0020] <観点 8>

観点 3 乃至 7 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラが、前記決定処理において、以下の処理：

(a) 前記モニタ情報を用いて、前記複数種類の実ページ群から実ページが割り当てられている一以上の仮想ボリューム内の全ての仮想ページの I/O 頻度の合計である I/O 頻度合計を算出する；

(b) 各種実ページ群の性能値の比と、前記 I/O 頻度合計とを基に、実ページ群の境界の I/O 頻度である境界値を実ページ群の境界毎に算出し；

(c) 各種実ページ群の I/O 頻度と実ページ群の境界毎の境界値とを基に、各種実ページ群の I/O 頻度レンジを決定し、
を行う。実ページ群の前記性能値は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいて決定された、性能に関する値である。前記実ページ群の境界とは、実ページ群とその上位及び／又は下位の実ページ群との境である。各種実ページ群の I/O 頻度レンジの最大値及び／又は最小値は、その実ページ群とその上位及び／又は下位の実ページ群との境の境界値に基づいて決定された値である。

[0021] <観点 9>

観点 3 乃至 8 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラが、前記決定処理において、以下の (d) 及び (e) の処理：

(d) 実ページ群の I/O 頻度レンジに収まる I/O 頻度となっている全ての仮想ページに割り当てられている実ページ内のデータ要素がその実ページ群に移行されたと仮定して、その実ページ群の使用割合がその実ページ群の割合閾値を超えるか否かを判断する；

(e) 前記 (d) での判断の結果が否定的の場合に、決定された I/O 頻度レンジを前記実ページ群管理情報に含める、
を行う。実ページ群の前記使用割合とは、その実ページ群を構成する実ページの数に対する、仮想ページに割り当てられている実ページの数割合である。前記実ページ群の割合閾値とは、前記使用割合の閾値である。

[0022] これにより、I/O 頻度レンジの適否を判断することができる。

[0023] <観点 10>

観点 9 に従うストレージシステムにおいて、前記コントローラが、前記決定処理において、更に、以下の (f) の処理：

(f) 前記 (d) での判断の結果が肯定的の場合に、実ページ群の使用割合がその実ページ群の割合閾値を超えないような大きさに I/O 頻度レンジを修正する、
を行う。

[0024] <観点 11>

観点 10 に従うストレージシステムにおいて、前記コントローラが、前記決定処理において、更に、以下の (g) の処理：

(g) 修正後の I/O 頻度レンジを前記実ページ群管理情報に含める、
を行う。

[0025] <観点 12>

観点 3 乃至 11 のうちのいずれかの観点に従うストレージシステムにおいて、前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I/O 頻度とは、その仮想ボリューム又は仮想ページの、前記モニタ情報に含まれている I/O 頻度である。

[0026] <観点 13>

観点 3 乃至 12 のうちのいずれかの観点に従うストレージシステムにおいて、前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I/O 頻度とは、前記モニタ情報における、その仮想ボリュームとは別の仮想ボリュームの I/O 頻度、又はその仮想ページとは別の仮想ページの I/O 頻

度、との相対的な値である。

[0027] <観点 1 4>

観点 3 乃至 1 2 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記 (A) での I/O 頻度を絶対値とするか相対値とするかの指定を受ける。絶対値とすることの指定を受けた場合、前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I/O 頻度とは、その仮想ボリューム又は仮想ページの、前記モニタ情報に含まれている I/O 頻度である。相対値とすることの指定を受けた場合、前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I/O 頻度とは、前記モニタ情報における、その仮想ボリュームとは別の仮想ボリュームの I/O 頻度、又はその仮想ページとは別の仮想ページの I/O 頻度、との相対的な値である。

[0028] <観点 1 5>

観点 3 乃至 1 4 のうちのいずれかの観点に従うストレージシステムにおいて、前記 (A) での I/O 頻度は、或る時間における I/O 数の平均である。

[0029] <観点 1 6>

観点 3 乃至 1 4 のうちのいずれかの観点に従うストレージシステムにおいて、前記 (A) での I/O 頻度は、或る時間帯の複数の時間にそれぞれ対応した複数の I/O 数のうちの最大の I/O 数である。

[0030] <観点 1 7>

観点 3 乃至 1 4 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記 (A) での I/O 頻度を平均 I/O 数とするか最大 I/O 数とするかの指定を受ける。平均 I/O 数とすることの指定を受けた場合、前記 (A) での I/O 頻度は、或る時間における I/O 数の平均である。最大 I/O 数とすることの指定を受けた場合、前記 (A) での I/O 頻度は、或る時間帯の複数の時間にそれぞれ対応した複数の I/O 数のうちの最大の I/O 数である。

[0031] <観点 1 8>

観点 3 乃至 1 7 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、I/O コマンドを前記ホストから受信し、その I/O コマンドから I/O 先の仮想ページを特定する。前記コントローラは、前記 I/O 先の仮想ページに割り当てられる実ページに対して I/O を行う場合、前記モニタ情報における、前記 I/O 先の仮想ページに対応した I/O 頻度を更新し、一方、前記 I/O 先の仮想ページに割り当てられる実ページに対して I/O を行わない場合、前記モニタ情報における、前記 I/O 先の仮想ページに対応した I/O 頻度を更新しない。

[0032] これにより、実ページに対して I/O が行われた場合にだけ、その実ページが割り当てられている仮想ページの I/O 頻度を更新することができる。

[0033] <観点 1 9>

観点 1 8 に従うストレージシステムにおいて、前記記憶資源は、前記モニタ情報として、状態がアクティブのモニタ情報と、状態がスタンバイのモニタ情報とを記憶する。前記コントローラは、前記アクティブのモニタ情報における I/O 頻度を更新する。前記コントローラは、前記ページ単位移行処理において、前記アクティブのモニタ情報を基に、前記第 1 の仮想ページを特定し、且つ、前記スタンバイのモニタ情報の状態をスタンバイからアクティブに変更し、その後、仮想ページが特定される I/O コマンドに従って実ページに対して I/O を行う場合、状態がアクティブに変更されたモニタ情報における、その特定された仮想ページの I/O 頻度を更新する。

[0034] これにより、前記移行処理が行われるときでも、I/O 頻度の監視（更新）を継続することができる。

[0035] <観点 2 0>

観点 4 に従うストレージシステムにおいて、前記コントローラは、前記作成した関係情報が表示されるよう前記関係情報を出力する。前記コントローラは、移行開始の指示を受けたときに、前記移行処理を開始する。

[0036] <観点 2 1>

観点 3 乃至 20 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の実データ群の使用割合が所定の割合閾値を超えた場合に、前記移行処理を開始する。

[0037] このため、前記第 1 の実データ群の使用割合が所定の割合閾値を超えた場合には、前記移行処理により、第 1 の実データ群にフリーの実ページが増えることが期待できる。

[0038] <観点 22>

観点 3 乃至 21 のうちのいずれかの観点に従うストレージシステムにおいて、前記第 1 種の実データ群が、最上位の実データ群である。前記コントローラは、前記ライトコマンドから特定されたライト先の仮想ページには、必ず、前記第 1 種の実データ群から実ページを割り当てる。

[0039] <観点 23>

観点 3 乃至 22 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記仮想ボリュームに割り当てられている全ての実ページのうちの所定割合以上の実ページが或る種の実ページ群に存在する場合、前記仮想ボリュームに割り当てられている全ての実ページのうち、前記或る種の実ページ群以外の種類の実ページ群内の実ページ内のデータを、前記或る種の実ページ群内の実ページに移行する。

[0040] これにより、仮想ボリュームのデータの配置状況に応じて、その仮想ボリュームにとって適切と考えられる種類の実ページ群に、その仮想ボリューム内のデータを集約することができる。

[0041] <観点 24>

観点 23 に従うストレージシステムにおいて、前記或る種の実ページ群内の実ページに移行した後に前記仮想ページに割り当てられる実ページは、常に、前記或る種の実ページ群内の実ページである。

[0042] これにより、移行処理によってデータが集約された後は、前記第 2 の仮想ボリュームには、前記或る種の実ページ群から実ページが割り当てられる。

[0043] <観点 25>

観点 3 乃至 2 4 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記第 1 の実ページから前記第 2 の実ページへのデータの移行が開始されていない場合、その移行をキャンセルし、前記ライトコマンドに従うデータを前記第 2 種の実ページ群内の実ページに書き込み、且つ、前記第 1 の仮想ページに前記第 1 の実ページに代えてその書き込み先の実ページを割り当てるよう前記割当て情報を更新する。

[0044] 移行前に移行先の実ページ群にデータが書き込まれるので、そのデータのライト先の仮想ページについては、移行を行わなくて済む。

[0045] <観点 2 6>

観点 3 乃至 2 5 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記第 1 の実ページから前記第 2 の実ページへのデータの移行が開始されていない場合、前記ライトコマンドに従うデータを前記第 1 の実ページに書き込み、その後、前記 (A) 及び (B) を行う。

[0046] <観点 2 7>

観点 3 乃至 2 6 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記第 1 の実ページから前記第 2 の実ページへデータを移行中の場合、その移行をキャンセルし、前記ライトコマンドに従うデータを前記第 2 種の実ページ群内の実ページに書き込み、且つ、前記第 1 の仮想ページに前記第 1 の実ページに代えてその書き込み先の実ページを割り当てるよう前記割当て情報を更新する。

[0047] <観点 2 8>

観点 3 乃至 2 6 のうちのいずれかの観点に従うストレージシステムにおい

て、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記第 1 の実ページから前記第 2 の実ページへデータを移行中の場合、前記ライトコマンドに従うデータを前記第 1 の実ページに書き込み、その後、前記第 1 の実ページから前記第 2 種の実ページ群内の実ページへのデータの移行を再開する。

[0048] <観点 29>

観点 3 乃至 26 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記第 1 の実ページから前記第 2 の実ページへデータを移行中の場合、その移行が済んだ後に、前記割当て情報に従って、前記ライトコマンドに従うデータを前記第 2 の実ページに書き込む。

[0049] <観点 30>

観点 3 乃至 26 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記第 1 の実ページから前記第 2 の実ページへデータを移行中の場合、前記第 1 の実ページと前記第 2 の実ページの両方に、前記ライトコマンドに従うデータを書き込む。

[0050] <観点 31>

観点 3 乃至 30 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記第 1 の仮想ページが特定されるライトコマンドを前記ページ単位移行処理が開始された後に前記ホストから受信し、且つ、前記 (A) 及び (B) が済んでいる場合、前記割当て情報に従って、前記第 2 の実ページに、前記ライトコマンドに従うデータを書き込む。

[0051] <観点 32>

観点 3 乃至 31 のうちのいずれかの観点に従うストレージシステムにおい

て、前記第 1 の仮想ページは、I / O が行われた最新の時刻である最終 I / O 時刻が所定の条件を満たす仮想ページである。

[0052] <観点 3 3>

観点 3 2 に従うストレージシステムにおいて、前記所定の条件とは、最終 I / O 時刻が現在時刻から所定時間以上に過去であることである。前記第 2 種の実ページ群は、前記第 1 種の実ページ群よりも下位の実ページ群である。

[0053] <観点 3 4>

観点 3 乃至 3 3 のうちのいずれかの観点に従うストレージシステムにおいて、前記第 1 の仮想ページは、シーケンシャルアクセスに従う I / O が発生した仮想ページである。前記第 2 種の実ページ群は、前記第 1 種の実ページ群よりも下位の実ページ群である。

[0054] <観点 3 5>

観点 3 乃至 3 4 のうちのいずれかの観点に従うストレージシステムにおいて、前記コントローラは、前記モニタ情報の更新を所定の時間帯においてのみ行う。前記コントローラは、所定の時間帯を過ぎる都度に、前記モニタ情報を保存する。前記第 1 の仮想ページは、指定された時間帯に対応したモニタ情報から特定された I / O 頻度が所定の条件に適合していない仮想ページである。

[0055] <観点 3 6>

ホストに接続されるストレージシステムの記憶制御装置が、記憶資源と、前記記憶資源及び前記ホストに接続されるコントローラとを備える。前記コントローラが、一以上の仮想ボリュームを前記一以上のホストに提供する。前記仮想ボリュームは、複数の仮想ページで構成される仮想的な論理ボリュームである。前記仮想ページは、仮想的な記憶領域である。前記実ページは、実体的な記憶領域である。前記記憶資源が、どの仮想ページにどの実ページが割り当てられているかを表す情報である割当て情報と、仮想ページ毎の I / O (Input/Output) 頻度を表す情報であるモニタ情報と、実ページ群の

種類毎に設定された I / O 頻度レンジを表す情報を含んだ情報である実ページ群管理情報とを記憶する。前記コントローラが、前記ホストからライトコマンドを受信し、前記ライトコマンドからライト先の仮想ページを特定し、特定した仮想ページに複数種類の実ページ群のいずれかからフリーの実ページを割り当てるよう前記割当て情報を更新し、割り当てた実ページにライト対象のデータを格納する。前記コントローラが、仮想ページ又はその仮想ページに割り当てられる実ページに対する I / O に応じて、前記モニタ情報における、その仮想ページの I / O 頻度を更新する。前記実ページは、実体的な記憶領域である。実ページ群が、複数の実ページを有する。前記コントローラが、以下の (A) 乃至 (C) の処理を含んだ処理である移行処理：

(A) 前記モニタ情報から特定された I / O 頻度が所定の条件に適合していない仮想ボリューム又は仮想ページを特定する；

(B) 特定された仮想ボリューム又は仮想ページに割り当てられている、第 1 種の実ページ群内の第 1 の実ページ、に格納されているデータを、第 2 種の実ページ群内の第 2 の実ページに移行する；

(C) 前記第 1 の実ページの割当て先の仮想ページである第 1 の仮想ページに前記第 1 の実ページに代えて前記第 2 の実ページを割り当てるよう前記割当て情報を更新する、

を行う。前記所定の条件とは、前記第 1 種の実ページ群の I / O 頻度レンジに収まることである。前記第 2 種の実ページ群は、前記 (A) で特定された仮想ボリューム又は仮想ページの I / O 頻度に I / O 頻度レンジが収まっている実ページ群である。前記コントローラは、I / O 頻度レンジの決定処理を行い、その決定処理において、前記モニタ情報を基に、各種の実ページ群の I / O 頻度レンジを決定する。

[0056] 前述した観点 1 及び 2 は、ストレージシステムの制御方法を例に採った観点であり、観点 3 乃至 3 5 は、ストレージシステムを例に採った観点であり、観点 3 6 は、記憶制御装置を例に採った観点であるが、それらのうちのいずれの観点も、ストレージシステム、記憶制御装置、記憶制御方法、ストレ

ージシステム又は記憶制御装置の制御方法、ストレージシステム又は記憶制御装置、コンピュータプログラム（例えば、（例えば、記憶制御のためのコンピュータプログラム、或いは、ストレージシステム又は記憶制御装置で実行されるコンピュータプログラム）のうちの少なくとも一つの観点になることができる。

[0057] 以下、図面を参照して、本発明の幾つかの実施例に係るストレージシステムを説明する。なお、以下の説明において、対象が、参照符号に代えて名称とIDの組み合わせで表されることがある。例えば、論理ボリュームが参照符号500で表される場合には、「論理ボリューム500」と表記され、IDが00の論理ボリュームが表される場合には、「論理ボリューム#00」と表記される。また、以下、一つの実ページに格納されるデータを、「データ要素」と呼ぶ。また、仮想ページに割り当てられた実ページに格納されているデータ要素を、「仮想ページ内のデータ要素」又はそれに類似した表現で表すことがある。

実施例 1

[0058] 図1は、本発明の実施例1の概要を示す。

[0059] ストレージシステム103にホスト装置（以下、ホスト）101が接続されている。ホスト101は、一つであっても複数であってもよい。ストレージシステム103は、仮想ボリューム（以下、VVOL（Virtual VOLUME））203をホスト101に提供し、複数種類のティア303を有する。

[0060] VVOL203は、Thin Provisioningに従う仮想的な論理ボリューム、すなわち、物理的な記憶デバイス（以下、PDEV）を基としない論理ボリュームである。VVOL203は、複数の仮想ページ207で構成されている。仮想ページ207は、仮想的な記憶領域である。VVOL203として、一つのVVOL#00があるとする。以下、VVOL#a内の仮想ページ#bを、「仮想ページ#（a-b）」と表記することにする。Thin ProvisioningのVVOL203は、仮想的な容量を有するものとしてホスト101へ提供されるものであり、ホスト101から仮想ページのあるアドレスへのライ

ト要求に応じて実ページが割り当てられるものである。したがって、仮想的な容量を満たすように実ページが割り当てられてしまった状態を除いては、あるVVOL203へ割り当てられた全ての実ページのトータル容量は、仮想的な容量よりも小さいものとなる。1つのVVOL203は、1つ以上のホスト101へ提供されるものであり、複数のホスト101へ提供される場合には、これらの複数のホスト101によって共有されることになる。

[0061] ティア303は、複数の実ページ209で構成されている。実ページ209は、実体的な記憶領域である。ティア303として、例えば、二つのティア#01及び#02があるとする。以下、ティア#c内の実ページ#dを、「実ページ#(c-d)」と表記することにする。なお、ティア303は、例えば、一以上の実ボリューム（以下、RVOL (Real VOLUME)）で構成されてよい。RVOLは、実体的な論理ボリューム、すなわち、PDEVを基とする論理ボリュームである。1つのプール内の複数のティア303の各々は、データ移行前の状態には、1つ又は複数のVVOL203に利用されるものとして設定される。

[0062] ホスト101は、一般的には計算機であるが、計算機に代えて、別のストレージシステムであってもよい。ホスト101は、例えば、I/O (Input/output) コマンドをストレージシステム103に送信する。I/Oコマンドは、例えば、ライトコマンド又はリードコマンドであり、I/O先情報を有する。I/O先情報は、I/O先を表す情報であり、例えば、VVOL203のID（例えばLUN (Logical Unit Number)）とI/O先のアドレス（例えばLBA (Logical Block Address)）とを含む。I/O先情報から、I/O先のVVOL203及び仮想ページ207が特定される。

[0063] ストレージシステム103が、ホスト101からライトコマンドを受信し、そのライトコマンドが有するI/O先情報を基に、ライト先として、仮想ページ#(00-C)を特定したとする。ストレージシステム103は、特定された仮想ページ#(00-C)にいずれの実ページ209が割り当てら

れていなければ、その仮想ページ207に、いずれかのフリー（未割当て状態）の実ページ#（01-D）を割り当て、割り当てた実ページ（01-D）に、ライトコマンドに従うライト対象のデータ要素を書き込む。

[0064] 本実施例では、データ移行が、ページ単位で行われる。具体的には、例えば、図1に示すように、ストレージシステム103が、以下の処理：

- *仮想ページ#（00-C）に割り当てられている実ページ#（01-D）内のデータ要素を、フリー（未割当て状態）の実ページ#（02-E）に移行する；

- *仮想ページ#（00-C）の割当元を、実ページ#（01-D）から実ページ#（02-E）に変更する；

- *実ページ#（01-D）の状態をフリー（未割当て状態）に更新する、を行う。

[0065] データ移行を移行するか否かや、移行先のティアをどこにするかが、後述のティアレンジに基づいて決定される。ティアレンジは、自動で決定される。

[0066] 以下、実施例1を詳細に説明する。

[0067] 図2は、ストレージシステム103の構成を示す。

[0068] ストレージシステム103は、複数のPDEV105と、複数のPDEV105に接続された記憶制御装置と、電源ユニット100とを有する。

[0069] 複数のPDEV105には、複数種類のPDEV、例えば、SSD（Solid State Drive）及びHDD（Hard Disk Drive）が含まれる。HDDとしては、HDD-FC（Fibre Channel）、HDD-SATA（Serial ATA）、及びHDD-SAS（Serial Attached SCSI）などがある。

[0070] 記憶制御装置は、コントローラと、コントローラに接続された複数のCMPK（キャッシュメモリパッケージ）119を有する。コントローラは、複数のCHA（チャネルアダプタ）111と、複数のDKA（ディスクアダプタ）113と、複数のMPPK（マイクロプロセッサパッケージ）121と、SW（スイッチ）117とを有する。CHA111、DKA113、MP

P K 1 2 1、S W 1 1 7 及び電源ユニット 1 0 0 は、冗長化の観点から、それぞれ複数存在するが、それらのうちの少なくとも一つは、図示の 2 に限らず、2 より多くても少なくても良い。C M P K 1 1 9 の数は、2 より多くても良い。

[0071] 電源ユニット 1 0 0 は、商用電源からの電力に基づく電力を各パッケージ 1 1 9、1 1 1、1 1 3、1 2 1 及び 1 1 7 に供給する。

[0072] C H A 1 1 1 は、ホスト 1 0 1 に接続されるインタフェース装置である。C H A 1 1 1 は、ホスト 1 0 1 から、I / O コマンド（ライトコマンド又はリードコマンド）を受信し、受信した I / O コマンドを、複数の M P P K 1 2 1 のうちのいずれかに転送する。

[0073] C H A 1 1 1 は、プロトコルチップ 1 1 2 と、L R（Local Router）1 1 4 とを有する。プロトコルチップ 1 1 2 は、ホスト 1 0 1 との通信のためのプロトコル変換を行う。L R 1 1 4 は、受信した I / O コマンドを、その I / O コマンドが有する I / O 先情報から特定される I / O 先に対応した M P P K 1 2 1 に、転送する。

[0074] D K A 1 1 3 は、H D D 1 0 5 に接続されるインタフェース装置である。D K A 1 1 3 は、H D D 1 0 5 からデータを読み出して C M P K 1 1 9 の D R A M（Dynamic Random Access Memory）に転送したり、C M P K 1 1 9 からのデータを H D D 1 0 5 に書込んだりする。

[0075] M P P K 1 2 1 は、一又は複数の M P（マイクロプロセッサ）を有する装置である。M P は、C H A 1 1 1 からの I / O コマンドを処理する。

[0076] S W 1 1 7 に、複数のパッケージ、すなわち、C H A 1 1 1、D K A 1 1 3、C M P K 1 1 9 及び M P P K 1 2 1 が接続されている。S W 1 1 7 は、P K（パッケージ）間の接続を制御する。

[0077] C M P K 1 1 9 は、揮発性メモリ及び／又は不揮発性メモリを有する。C M P K 1 1 9 には、例えば、仮想ページ 2 0 7（実ページ 2 0 9）に対する I / O の対象のデータ要素が一時的に格納される記憶領域（以下、キャッシュメモリ或いはそれを略して C M）がある。また、C M P K 1 1 9 には、種

々の情報及びコンピュータプログラムが記憶される記憶領域がある。情報及びコンピュータプログラムについては、後に説明する。

[0078] ストレージシステム 103 に、管理装置 107 が接続されている。管理装置 107 は、例えば、表示装置及び入力装置を有する計算機である。管理者は、管理装置 107 からストレージシステム 103 に対して種々の設定を行うことができる。

[0079] 図 3 は、ストレージシステム 103 における各種記憶領域の関係を示す。

[0080] 図 3 に示すように、下位から上位にかけて、一以上の RG 301、一以上のプール 201、一以上の VVOL 203 が管理される。

[0081] RG 301 は、RAID グループの略である。一つの RG 301 は、同一種類の PDEV で構成されている。PDEV 種類は、例えば、性能及び単位コストのうちの少なくとも一方で定義される。性能とは、例えば、データの I/O の速度、又は、レスポンスタイム（ホストからコマンドを受けてから応答を返すまでの時間長）である。単位コストは、単位サイズのデータを保存するのに要する価格（例えばいわゆるビットコスト）である。例えば、RG#00 は、複数の SSD で構成されており、RG#01 は、複数の HDD-SAS で構成されており、RG#02 は、複数の HDD-SATA で構成されている。なお、一つの RG を構成する複数の PDEV の容量は、例えば同じである。

[0082] プール 201 は、実ページ 207 の集合である。具体的には、例えば、プール 201 は、一以上の RVOL 205 で構成されており、各 RVOL 205 が、一以上の実ページ 207 で構成されている。以下、プール 201 を構成する RVOL を、プール VOL を略して「PVOL」と表す。

[0083] 一つのプール 201 を構成する実ページ 207 が階層化されている。すなわち、一つのプール 201 に、複数のティア 303 が存在する。ティア 303 は、例えば、RVOL の種類毎に存在する。例えば、図 3 によれば、RVOL の種類として 3 種類あるため、ティア 303 の数も 3 つである。このため、一つのティア 303 には、そのティアに対応した種類の RVOL 205

が属することになる。具体的には、例えば、ティア#00は、SSDという種類であり、SSDで構成されたRG301に基づくRVOLが属することになり、ティア#01は、HDD-SASという種類であり、HDD-SASで構成されたRG301に基づくRVOLが属することになり、ティア#03は、HDD-SATAという種類であり、HDD-SATAで構成されたRG301に基づくRVOLが属することになる。なお、ティア303は、必ずしも、厳密にRVOL種類毎に設けられる必要も無く、同一のティア303に、類似の種類のRGに基づく複数のRVOL205が属しても良い。

[0084] VVOL203には、一以上のプール201から、割当元となるプール201が関連付けられる。また、VVOL203には、選択されたプール201から、割当元として使用可能なティア（以下、割当ティア）303が関連付けられる。また、VVOL203には、一以上の割当ティア303のうちの一つがデフォルトのティア303として関連付けられる。図3によれば、VVOL#01の割当ティアは、ティア#00及び#02である。

[0085] また、上述の説明によれば、複数のRVOLは、それらが同一のティアに属していれば、一つのRGが基になっている可能性があるが、それらが異なるティアにそれぞれ属していれば、それぞれ別々のRGが基になっていることになる。同一のティア303に関連する1つ又は複数のVVOL203の複数の仮想ページに対して複数の実ページを順次割り当てる場合には、このティア303に関連するRGをまたいで実ページが順次選択されるようすることが望ましい。具体的には、このティア303に含まれる複数のRVOL205から複数の実ページが選択される場合に、RGとの関係を考慮して、順次異なるRGから実ページが順次選択される。これにより、アクセス性能を向上させることが可能になる。

[0086] 図4は、CMPK119が記憶する情報及びコンピュータプログラムを示す。なお、本実施例では、情報の一例としてテーブルが採用されるが、情報は、テーブル以外の形式（例えばキュー形式）であっても良い。

[0087] CMPK 119は、情報として、ティア定義テーブル401、ティア管理テーブル402、プール管理テーブル403、VVOL管理テーブル404、PVOL管理テーブル405、アロケーションテーブル407、実ページ管理テーブル408、モニタテーブル409、度数分布テーブル410、移行ページテーブル411及び移行定義テーブル412を記憶する。また、CMPK 119は、コンピュータプログラムとして、UI (User Interface) を制御するプログラムであるUI制御プログラム413、I/Oを制御するプログラムであるI/O制御プログラム414、及び、再配置を行うプログラムである再配置プログラム415を記憶する。なお、これらのテーブル及びコンピュータプログラムのうちの少なくとも一つがCMPK 119以外の記憶資源に記憶されていても良い。例えば、少なくとも一つのコンピュータプログラム、及び／又は、移行ページテーブル411は、MPPK 121内のメモリに記憶されても良い。

[0088] 以下、各テーブルを説明する。なお、以下の説明では、対象は、IDを用いて特定されるが、IDに代えて又は加えて、名称及び番号のうちの少なくとも一つを用いて特定されても良い。

[0089] 図5は、ティア定義テーブル401の構成を示す。

[0090] ティア定義テーブル401は、ティアの定義を表すテーブルである。具体的には、例えば、テーブル401は、RVOL種類とティアIDとの関係を表す。テーブル401によれば、RVOL種類「SSD」に対応したティアとして、「00」というIDが付与されたティアが用意されることになる。

[0091] 図6は、ティア管理テーブル402の構成を示す。

[0092] ティア管理テーブル402は、ティアに関する情報を管理するためのテーブルである。具体的には、例えば、テーブル402は、ティア毎に、プールID、ティアID、ティアレンジ、PVOL ID、ティア閾値、使用割合、実ページIDを有する。一つのティア（以下、図6の説明において「対象ティア」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

- * 「プールID」は、対象ティアを有するプールのIDである。
- * 「ティアID」は、対象ティアのIDである。ティアIDから、対象ティアに属するPVOLの基になっているPDEVの種類が一義的に特定される。
- * 「ティアレンジ」は、対象ティアに存在すべき実ページの負荷（例えばI/O頻度）の範囲である。
- * 「PVOL ID」は、対象ティアに属するPVOLのIDである。
- * 「ティア閾値」は、対象ティアの使用割合の閾値である。
- * 「使用割合」は、対象ティアの使用割合、すなわち、対象ティアを構成する実ページの数に対する、割当て済みの実ページの数割合である。
- * 「実ページID」は、対象ティアに属する全ての実ページのIDである。

[0093] 図6のテーブル402の例によれば、ティア閾値は、ティア毎に設定されているが、それに代えて、一つのプールについて一つのティア閾値が設定され、その一つのティアが、その一つのプールが有する全てのティアについて共通であっても良い。

[0094] 図7は、プール管理テーブル403の構成を示す。

[0095] プール管理テーブル403は、プールに関する情報を管理するためのテーブルである。具体的には、例えば、テーブル403は、プール毎に、プールID、階層化、再配置、自動実行、再配置実行周期、再配置実行時刻、モニタ時間帯及びプールステータスを有する。一つのプール（以下、図7の説明において「対象プール」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

- * 「プールID」は、対象プールのIDである。
- * 「階層化」は、対象プール内の実ページを階層化するか否かである。“ON”は、階層化すること（つまり、対象プールに複数のティアが設けられること）を意味し、“OFF”は、階層化しないこと（つまり、対象プールにティアが設けられないこと）を意味する。
- * 「再配置」は、対象プール内のデータ要素を再配置するか否かである。“

“ON”は、再配置を行うことを意味し、“OFF”は、再配置を行わないことを意味する。“ON”の場合、対象プールが割り当てられているVVOL又はその仮想ページのI/O頻度がモニタされ、“OFF”の場合、対象プールが割り当てられているVVOL又はその仮想ページのI/O頻度がモニタされない。I/O頻度のモニタに関して大事なことは、I/O先のVVOL或いは仮想ページに割り当てられた実ページに実際にI/Oが行われなければ、I/O頻度は更新されず、その割り当てられた実ページに対してI/Oが行われたときに、I/O頻度が更新されることである。この点は、後述のライト処理及びリード処理の際にも説明する。

＊「自動実行」は、対象プール内のデータ要素の再配置を自動で開始するか手動で開始するかである。“ON”は、再配置が自動で開始されることを意味し、“OFF”は、再配置が手動で開始されることを意味する。

＊「再配置実行周期」は、対象プール内のデータ要素の再配置を行う周期である。例えば、“1日”は、1日（24時間）毎に再配置が開始されることを意味する。

＊「再配置実行時刻」は、対象プール内のデータ要素の再配置を開始する時刻である。

＊「モニタ時間帯」は、対象プールが割り当てられているVVOLに割り当てられた実ページのI/O頻度をモニタする時間帯である。

＊「プールステータス」は、対象プールのステータスである。「プールステータス」の値として、例えば、“モニタリング中”、“再配置中”、“非モニタリング中”がある。“モニタリング中”は、対象プールが割り当てられているVVOL又はそのVVOL内の仮想ページのI/O頻度がモニタリングされている最中であり、且つ、データ要素の再配置中ではないことを意味する。“再配置中”は、データ要素の再配置（対象プール内での再配置であっても良いし、対象プールから別のプールへのデータ要素の再配置であっても良い）が行われている最中であることを意味する。“非モニタリング中”は、I/O頻度のモニタリング中でもないしデータ要素の再配置中でもない

ことを意味する。

[0096] 図8は、VVOL管理テーブル404の構成を示す。

[0097] VVOL管理テーブル404は、VVOLに関する情報を管理するためのテーブルである。具体的には、例えば、テーブル404は、VVOL毎に、VVOL ID、PVOL ID、割当ティアID、デフォルトティアID、容量、再配置、自動実行、再配置実行周期、再配置実行時刻及びモニタ時間帯を有する。これらの情報要素のうち、再配置、自動実行、再配置実行周期、再配置実行時刻及びモニタ時間帯は、プール管理テーブル403における情報要素と同じであるが、このテーブル404における情報要素（例えば「再配置」）の値がプール管理テーブル403における同じ情報要素（例えば「再配置」）の値と違っていれば、このテーブル404における値が優先されて良い。つまり、プール管理テーブル403では、一つのプールについて情報要素の値が設定され、その結果、そのプールが割り当てられている全てのVVOLについての設定にその値を反映させることができるが、VVOL毎に、設定をすることができる。VVOLについて特に値が設定されていなければ、そのVVOLに割り当てられているプールについて設定されている値が、そのVVOLについての値とされる。一つのVVOL（以下、図8の説明において「対象VVOL」と言う）を例に採っていうと、このテーブル404における上記の情報要素は、具体的には、下記の通りである。

*「VVOL ID」は、対象VVOLのIDである。

*「PVOL ID」は、対象VVOLに割り当てられているプール（以下、割当プール）のIDである。

*「割当ティアID」は、割当プールが有する複数のティアのうちの対象VVOLに割り当てられるティアのIDである。対象VVOL内の仮想ページには、割当ティアから実ページが割り当てられる。

*「デフォルトティアID」は、デフォルトティアのIDである。デフォルトティアとは、ホスト101からのライト先の仮想ページに初めに割り当てられる実ページの提供元となるティアである。

*「容量」は、対象VVOLの容量である。

*「再配置」は、対象VVOL内のデータ要素を再配置するか否かである。

“ON”は、再配置を行うことを意味し、“OFF”は、再配置を行わないことを意味する。“ON”の場合、対象VVOL又はその仮想ページのI/O頻度がモニタされ、“OFF”の場合、対象VVOL又はその仮想ページのI/O頻度がモニタされない。

*「自動実行」は、対象VVOL内のデータ要素の再配置を自動で開始するか手動で開始するかである。“ON”は、再配置が自動で開始されることを意味し、“OFF”は、再配置が手動で開始されることを意味する。

*「再配置実行周期」は、対象VVOL内のデータ要素の再配置を行う周期である。

*「再配置実行時刻」は、対象VVOL内のデータ要素の再配置を開始する時刻である。

*「モニタ時間帯」は、対象VVOL又は仮想ページのI/O頻度をモニタする時間帯である。

[0098] 図9は、PVOL管理テーブル405の構成を示す。

[0099] PVOL管理テーブル405は、PVOLに関する情報を管理するためのテーブルである。具体的には、例えば、テーブル405は、PVOL毎に、RG ID、PDEV ID、種類、PVOL ID、及び容量を有する。一つのPVOL（以下、図9の説明において「対象PVOL」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

*「RG ID」は、対象PVOLの基になっているRGのIDである。

*「PDEV ID」は、対象PVOLの基になっているRGを構成する全てのPDEVのIDである。

*「種類」は、対象PVOLの種類である。これは、対象PVOLの基になっているPDEVの種類と同じである。例えば、RVOL種類が“SSD”であれば、対象PVOLの種類も“SSD”となる。

*「PVOL ID」は、対象PVOLのIDである。

* 「容量」は、対象PVOLの容量である。

[0100] 図10は、アロケーションテーブル407の構成を示す。

[0101] アロケーションテーブル407は、どの仮想ページにどの実ページが割り当てられているかを表すテーブルである。具体的には、例えば、テーブル407は、仮想ページ毎に、VVOL ID、仮想ページID、プールID、実ページID及びティアIDを有する。一つの仮想ページ（以下、図10の説明において「対象仮想ページ」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

* 「VVOL ID」は、対象仮想ページを有するVVOLのIDである。

* 「仮想ページID」は、対象仮想ページのIDである。

* 「プールID」は、対象仮想ページを有するVVOLに割り当てられているプールのIDである。

* 「実ページID」は、対象仮想ページに割り当てられている実ページのIDである。

* 「ティアID」は、対象仮想ページに割り当てられている実ページを有するティアのIDである。

[0102] 図11は、実ページ管理テーブル408の構成を示す。

[0103] 実ページ管理テーブル408は、各実ページの状態を管理するためのテーブルである。具体的には、例えば、テーブル408は、実ページ毎に、プールID、実ページID及びステータスを有する。一つの実ページ（以下、図11の説明において「対象実ページ」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

* 「プールID」は、対象実ページを有するプールのIDである。

* 「実ページID」は、対象実ページのIDである。

* 「ステータス」は、対象実ページのステータスである。ステータスの値として、例えば、“使用中”と“フリー”がある。“使用中”は、対象実ページがいずれかの仮想ページに割当て済みであることを意味する。“フリー”は、対象実ページがいずれの仮想ページにも割り当てられておらずそれ故に

ずれの仮想ページにも割当て可能な状態になっていることを意味する。

[0104] 図 12 は、モニタテーブル 409 の構成を示す。

[0105] モニタテーブル 409 は、VVOL 毎に存在する。テーブル 409 は、仮想ページの I/O 頻度を表すテーブルである。一つの VVOL（以下、図 12 の説明において「対象 VVOL」と言う）を例に採って、テーブル 409 を説明する。

[0106] テーブル 409 は、具体的には、例えば、仮想ページ毎に、VVOL ID、仮想ページ ID、合計 I/O 数、平均 I/O 数、最大 I/O 数及び最終 I/O 時刻を有する。一つの仮想ページ（以下、図 12 の説明において「対象仮想ページ」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

* 「VVOL ID」は、対象仮想ページを有する VVOL（つまり対象 VVOL）の ID である。

* 「仮想ページ ID」は、対象仮想ページの ID である。

* 「合計 I/O 数」は、モニタ時間帯に対象仮想ページに行われた I/O の数である。

* 「平均 I/O 数」は、「合計 I/O 数」の値を所定時間で割ることにより算出された値である。

* 「最大 I/O 数」は、モニタ時間帯を構成する複数の時間帯における複数の I/O 数（対象仮想ページについての I/O 数）のうちの最も大きい I/O 数である。

* 「最終 I/O 時刻」は、対象仮想ページに I/O が行われた最新の時刻である。

[0107] テーブル 409 も、例えば二つ（二つより多くても良い）存在する。二つのテーブル 409 が、交互に使用される。具体的には、例えば、再配置が実行される場合、その実行時点の直近のモニタ時間帯についてのテーブル 409 を用いて、後述する移行ページテーブル 411 が作成され、且つ、他方のテーブル 409 を用いて、I/O 頻度のモニタが行われる。以下の説明では

、使用されている方のテーブル４０９を、アクティブのテーブル４０９と言い、使用されていない方のテーブル４０９を、スタンバイのテーブル４０９と言うことがある。

[0108] 図１３は、度数分布テーブル４１０の構成を示す。

[0109] 度数分布テーブル４１０は、Ｉ／Ｏ数と仮想ページ数との相関を表すテーブルである。このテーブル４１０は、モニタテーブル４０９を基に作成される。Ｉ／Ｏ数としては、前述したテーブル４０９における合計Ｉ／Ｏ数、平均Ｉ／Ｏ数及び最大Ｉ／Ｏ数のうちの少なくとも一つが採用されて良い。図１３は、平均Ｉ／Ｏ数が採用された例である。すなわち、図１３に示すテーブル４１０は、平均Ｉ／Ｏ数別に仮想ページの数を表す。図１３によれば、平均Ｉ／Ｏ数が“０”の仮想ページ数は“５６１”であることがわかる。このテーブル４１０（又はこのテーブル４１０を基に作成されるグラフ）は、管理装置１０７に表示される。管理者は、テーブル４０１（又はそれに基づくグラフ）を基に、手動で、データ要素の再配置の開始をストレージシステム１０３に指示することができる。

[0110] 度数分布テーブル４１０は、モニタテーブル４０９が更新されることに伴って更新されても良いし、度数分布の表示を管理者から命じられたときに、テーブル４０９を基に作成されても良い。

[0111] 図１４は、移行ページテーブル４１１の構成を示す。

[0112] 移行ページテーブル４１１は、ページ単位のデータ移行の際に、モニタテーブル４０９を基に作成されるテーブルである。テーブル４１１は、移行元と移行先との関係を表す。具体的には、例えば、テーブル４１１は、移行対象の仮想ページ毎に、仮想ページＩＤ、移行元ティアＩＤ、移行先ティアＩＤ及び移行ステータスを有する。一つの仮想ページ（以下、図１４の説明において「対象仮想ページ」と言う）を例に採っていうと、これらの情報要素は、具体的には、下記の通りである。

* 「仮想ページＩＤ」は、対象仮想ページのＩＤである。

* 「移行元ティアＩＤ」は、対象仮想ページに割り当てられている実ページ

（移行元の実ページ）を有するティアのIDである。

＊「移行先ティアID」は、データ要素の移行先となるティアのIDである。移行先の実ページは、そのティアにおけるいずれかフリーの実ページである。

＊「移行ステータス」は、対象仮想ページについての移行のステータスである。「移行ステータス」の値として、例えば、“移行済”、“移行中”及び“未移行”がある。“移行済”は、移行元の実ページから移行先のティアにおけるいずれかの実ページにデータ要素を移行済みであることを意味する。

“移行中”は、移行元の実ページから移行先のティアにおけるいずれかの実ページにデータ要素を移行中であることを意味する。“未移行”は、移行元の実ページから移行先のティアにおけるいずれかの実ページにデータ要素を移行することが未だ開始されていないことを意味する。

[0113] なお、移行ページテーブル411は、仮想ページIDに代えて又は加えて、移行元に関する他種の情報要素を有してもよい。例えば、仮想ページに割り当てられている実ページのIDがテーブル411に含まれてもよい。

[0114] また、移行ページテーブル411は、移行先ティアIDに代えて又は加えて、移行先に関する他種の情報要素を有してもよい。例えば、移行先ティアにおけるフリーの実ページがテーブル402（図6）及び408（図11）を基に特定され、その特定された実ページのIDがテーブル411に含まれてもよい。その際、特定された実ページのステータスが、“フリー”から“使用中”（或いは他のステータス（例えば“リザーブ”）に更新されてもよい。

[0115] 図15は、移行定義テーブル412の構成を示す。

[0116] 移行定義テーブル412は、各VVOLについて、移行に関する定義を表す。具体的には、例えば、テーブル412は、VVOL毎に、VVOL ID、ティア維持、割当ティアID及び格納閾値を有する。一つのVVOL（以下、図15の説明において「対象VVOL」と言う）を例に採つていうと、これらの情報要素は、具体的には、下記の通りである。

*「VVOL ID」は、対象VVOLのIDである。

*「割当ティアID」は、対象VVOLの割当ティアのIDである。

*「格納閾値」は、割当ティアの格納割合の閾値である。この閾値は、割当ティア毎に設定される。「格納割合」とは、一つの割当ティア（以下、この段落の説明において「対象割当ティア」）について言えば、次のとおりである。すなわち、格納割合は、対象VVOL内の全てのデータ要素のうちどの程度の割合のデータ要素が対象割当ティアに格納されているかである。

[0117] 図15のテーブル412の例によれば、格納閾値は、ティア毎に設定されているが、それに代えて、一つのプールについて一つの格納閾値が設定され、その一つのティアが、その一つのプールが有する全てのティアについて共通であっても良い。

[0118] 以下、本実施例で行われる処理を説明する。なお、以下の説明において、コンピュータプログラムが行う処理は、実際には、コンピュータプログラムを実行するMP（マイクロプロセッサ）が行う。MPは、MPPK121（図2参照）にある。

[0119] <プールの作成>。

[0120] 図16は、プール作成処理のフローを示す。なお、この<プールの作成>の説明において、そのプール作成処理で新たに作成されるプールを「対象プール」と言う。

[0121] (S1601)

UI制御プログラム413が、プールの作成指示を、管理者から管理装置107を通じて受ける。そして、プログラム413は、例えば下記の準備処理を行う。

*プログラム413は、対象プールの新規のレコードを、テーブル402（図6）、403（図7）、408（図11）に追加する。

*プログラム413は、PVOLの指定を管理者から管理装置107を通じて受ける。具体的には、例えば、プログラム413は、RVOLに関する情報（例えば、ID、種類など）を表示し、PVOLとするRVOLの指定を

受ける。

[0122] 指定されたRVOL毎に、S 1 6 0 2～S 1 6 0 5が行われる。一つのRVOL（以下、図 1 6 の説明において「対象RVOL」と言う）を例に採り、S 1 6 0 2～S 1 6 0 5を説明する。

[0123] (S 1 6 0 2)

プログラム 4 1 3 は、指定されたRVOLの種類を特定する。具体的には、例えば、CMPK 1 1 9 が、RVOL毎にID、容量、種類（例えば、SSD、HDD-SAS）などを有するRVOL管理テーブルを記憶していて、プログラム 4 1 3 は、そのRVOL管理テーブルから、指定されたRVOLの種類を特定する。

[0124] (S 1 6 0 3)

プログラム 4 1 3 は、特定されたRVOL種類に対応したティアが対象プールに既に存在しているか否かを、ティア管理テーブル 4 0 3 から判断する。この判断の結果が否定的の場合、S 1 6 0 4 が行われ、この判断の結果が肯定的の場合、S 1 6 0 5 が行われる。

[0125] (S 1 6 0 4)

プログラム 4 1 3 は、特定されたRVOL種類に対応したティアのIDをティア定義テーブル 4 0 1 から特定する。そして、プログラム 4 1 3 は、対象プールについて、特定されたティアIDを、テーブル 4 0 2（図 6）に追加する。

[0126] (S 1 6 0 5)

プログラム 4 1 3 は、特定されたRVOL種類に対応したティア（以下、この段落において「対象ティア」と言う）に、対象RVOLを追加する。具体的には、例えば、プログラム 4 1 3 は、テーブル 4 0 2（図 6）に、対象ティアのティアIDに対応するPVOL IDとして、対象RVOLのIDを登録する。

[0127] (S 1 6 0 6)

プログラム 4 1 3 は、全ての指定されたRVOLについてS 1 6 0 2以降

が行われたか否かを判断する。指定されたRVOLに、S 1 6 0 2以降の処理が行われていないRVOLがあれば、そのRVOLについて、S 1 6 0 2以降が行われる。

[0128] (S 1 6 0 7)

プログラム4 1 3は、プール定義GUI (Graphical User Interface) を表示する。そして、プログラム4 1 3は、そのGUIを通じて種々の設定を受け付ける。

[0129] 図1 7は、プール定義GUI 1 7 0 0を示す。

[0130] GUI 1 7 0 0には、対象プールのIDが表示される。なお、プログラム4 1 3は、所定のツール（例えばプルダウンメニュー）が管理者に操作されたことに応答して、対象プール以外のプール（すなわち既存のプール）のIDを表示することもできる。

[0131] また、GUI 1 7 0 0は、表示されているIDが表すプールについて、テーブル4 0 3（図7）に設定すべき情報要素「階層化」、「再配置」、「自動実行」、「実行周期」、「実行時刻」、「モニタ時間帯」、「ティア閾値」の値を入力するために管理者が使用する入力ツールを有する。本実施例で言う入力ツールとしては、ラジオボタン、チェックボックス、テキスト入力欄などを任意のツールを採用することができる。なお、GUI 1 7 0 0は、「ティア閾値」について、一つの閾値を、対象プールが有する全てのティアに共通とさせることも、或いは、対象プールが有するティア毎に閾値を設定することもできるようになっている。

[0132] なお、前述したように、GUI 1 7 0 0には、既存のプールのIDも表示可能であるため、GUI 1 7 0 0を用いて、管理者所望の既存のプールについての設定を変更することもできる。その際、プログラム4 1 3は、例えば、管理者所望のティア閾値が、そのティア閾値が設定されるティアの使用割合よりも高ければ、不適切なティア閾値として、警告を管理装置1 0 5に表示し、そのティア閾値をテーブル4 0 2（図6）に登録しないようにしても良い。

[0133] 再び図 1 6 を参照する。

[0134] (S 1 6 0 8)

プログラム 4 1 3 は、テーブルを更新する。具体的には、例えば、プログラム 4 1 3 は、下記の処理を行う。

* プログラム 4 1 3 は、図 1 7 の G U I 1 7 0 0 に入力された、情報要素「階層化」、「再配置」、「自動実行」、「実行周期」、「実行時刻」、「モニタ時間帯」、「ティア閾値」の値を、テーブル 4 0 3 (図 7) に登録する。

* プログラム 4 1 3 は、対象プールを複数の実ページに区切り、それら複数の実ページの I D を、テーブル 4 0 2 (図 6)、及び、4 0 8 (図 1 1) に追加する。

[0135] 実ページの容量は、全てのプールにおいて共通であってよい。そのため、対象プールを構成する実ページの数、対象プールの容量(対象プールを構成する P V O L の総容量)に依存してよい。

[0136] なお、P V O L 毎に、実ページの容量が異なっても良い。具体的には、例えば、R A I D 5 (4 D + 1 P) の R G に基づく P V O L を構成する実ページの容量と、R A I D 6 (4 D + 2 P) の R G に基づく P V O L を構成する実ページの容量は、異なっても良い。なぜなら、一つの実ページに格納されるパリティの数が違うためである。

[0137] <V V O L の作成>。

[0138] 図 1 8 は、V V O L 作成処理のフローを示す。なお、この<V V O L の作成>の説明において、そのV V O L 作成処理で新たに作成されるV V O L を「対象V V O L」と言う。

[0139] (S 1 8 0 1)

U I 制御プログラム 4 1 3 は、V V O L 作成 G U I を表示する。そして、プログラム 4 1 3 は、その G U I を通じて種々の設定を受け付ける。

[0140] 図 1 9 は、V V O L 作成 G U I 1 9 0 0 を示す。

[0141] G U I 1 9 0 0 には、対象V V O L の I D が表示される。なお、プログラ

ム４１３は、所定のツール（例えばプルダウンメニュー）が管理者に操作されたことに応答して、対象ＶＶＯＬ以外のＶＶＯＬ（既存のＶＶＯＬ）のＩＤを表示することもできる。

[0142] ＧＵＩ１９００は、表示されているＩＤが表すＶＶＯＬについて、テーブル４０４（図８）に設定すべき情報要素「プールＩＤ」、「容量」、「再配置」、「自動実行」、「実行周期」、「実行時刻」、「モニタ時間帯」の値を入力するために管理者が使用する入力ツールを有する。

[0143] また、ＧＵＩ１９００は、入力されたプールＩＤに対応するティアＩＤをテーブル４０２（図６）から特定し、特定された全てのティアＩＤを表示する。ＧＵＩ１９００は、表示されているＩＤが表すＶＶＯＬについて、「割当ティア」、「格納閾値」及び「デフォルトティア」の値を入力するために管理者が使用する入力ツールを有する。管理者は、それらの入力ツールを用いて、入力されたプールＩＤが表すプール（ＶＶＯＬに割り当てたプール）が有する複数のティアのうちのどれを割当ティアにするか、割当ティアのうちのどれを「デフォルトティア」にするか、及び、各割当ティアの格納閾値を、入力することができる。

[0144] 再び図１８を参照する。

[0145] （１８０２）

プログラム４１３は、テーブルを更新する。具体的には、例えば、プログラム４１３は、下記の処理を行う。

* プログラム４１３は、図１９のＧＵＩ１９００に入力された、情報要素「プールＩＤ」、「容量」、「再配置」、「自動実行」、「実行周期」、「実行時刻」、「モニタ時間帯」、「割当ティアＩＤ」及び「デフォルトティアＩＤ」の値を、テーブル４０４（図８）に登録する。

* プログラム４１３は、対象ＶＶＯＬの容量に基づいて、対象ＶＶＯＬを構成する仮想ページの数を出算する。プログラム４１３は、対象ＶＶＯＬのＩＤと、算出された仮想ページ数分の仮想ページＩＤとを、テーブル４０７（図１０）に登録する。

*プログラム413は、対象VVOLについて、「再配置」が“ON”であれば、テーブル409（図12）を作成する。モニタ時間帯に、I/O頻度のモニタが行われるからである。なお、テーブル409には、上記算出された仮想ページ数分の仮想ページIDが登録される。

*プログラム413は、対象VVOLについて、「割当ティアID」及び「格納閾値」を、テーブル412（図15）に登録する。

[0146] <情報の表示>。

[0147] 図22は、情報表示処理のフローを示す。

[0148] UI制御プログラム413は、管理装置107から表示命令を受け（S2201）、表示命令で指定されている情報を管理装置107に表示する（S2202）。

[0149] 表示可能な情報の種類として、例えば、テーブル401～412のうちの少なくとも一つ（或いはそれを加工することにより得られる情報（例えば、グラフ、サマリーなど））がある。

[0150] 図23は、度数分布GUI2300を示す。

[0151] 度数分布GUI2300は、図22のS2202で表示される。具体的には、例えば、図22のS2201で、管理装置107が、管理者からの要求に従って、管理者所望のVVOL IDを指定した度数分布表示命令を送信する。図22のS2202で、プログラム413が、その命令に応答して、その命令で指定されているVVOL IDに対応した度数分布をテーブル410（図13）から抽出し、抽出した度数分布を、図23に示すように管理装置107に表示する。なお、度数分布は、表形式で表示されているが、それに代えて又は加えて、グラフ形式など他の形式で表示されても良い。

[0152] 管理者は、GUI2300に表示されている度数分布を見て、所望のVVOL内のデータ要素を移行するか否かを判断する。例えば、管理者は、平均I/O数が高い仮想ページ数がたくさん存在すると思った場合には、所望のVVOL内のデータ要素の移行をストレージシステム103に指示して良い（例えば、GUI2300にある「移行実行」ボタンを押して良い）。

[0153] このように、度数分布は、データ移行（VVOL内のデータ要素の移行）を行うか否かを管理者が判断するための判断材料として用いられる。別の言い方をすれば、データ移行が自動で実行されるのであれば、度数分布は計算されなくて良い。

[0154] <ライト処理（キャッシュメモリへのライト）>。

[0155] 図24は、ライト処理（キャッシュメモリへのライト）のフローを示す。

[0156] （S2401）

I/O制御プログラム414が、ライトコマンドをホスト101から受信する。

[0157] （S2402）

プログラム414が、受信したライトコマンドが有するI/O先情報から、ライト先のVVOL及び仮想ページを特定する。ここでは、複数の仮想ページが特定されることが多い。なぜなら、ライト対象のデータのサイズは、通常、一つの仮想ページの容量よりも大きいからである。特定された仮想ページ毎に、S2403以降が行われる。以下、一つの仮想ページを例に採り（図24の説明において、「対象仮想ページ」と言う）、S2403以降を説明する。また、図24の説明において、対象仮想ページに対するライト対象のデータ要素を「対象データ要素」と言う。

[0158] （S2403）

プログラム414は、対象仮想ページのIDが、移行ページテーブル411に存在するか否かを判断する。この判断の結果が否定的の場合、S2404～S2406が行われ、この判断の結果が肯定的の場合、S2407～S2408が行われる。

[0159] （S2404）

プログラム414は、アロケーションテーブル407（図10）を基に、対象仮想ページに、実ページが割り当てられているか否かを判断する。この判断の結果が否定的であれば、S2405が行われる。

[0160] （S2405）

プログラム 4 1 4 が、以下の処理を行う。

* プログラム 4 1 4 は、ライト先の VVOL に対応したプール及びそのプール内のデフォルトティアを、VVOL 管理テーブル 4 0 4（図 8）を基に特定する。

* プログラム 4 1 4 は、ティア管理テーブル 4 0 2（図 6）及び実ページ管理テーブル 4 0 8（図 1 1）を基に、特定されたデフォルトティア内のフリーの実ページを特定する。

* プログラム 4 1 4 は、対象仮想ページに、特定した実ページを割り当てる。すなわち、プログラム 4 1 4 は、アロケーションテーブル 4 0 7（図 1 0）における、対象仮想ページの ID に対応した欄に、割り当てた実ページの ID と、その実ページを有するティアの ID と、そのティアを有するプールの ID とを登録する。

* プログラム 4 1 4 は、デフォルトティアからフリーの実ページを特定できない場合には、ライト先の VVOL に対応した割当ティアからフリーの実ページを特定し、特定した実ページを対象仮想ページに割り当ててよい。

* プログラム 4 1 4 は、テーブル 4 0 2（図 6）における、割り当てた実ページを有するティア（この S 2 4 0 3 の説明において「割当元ティア」と言う）の「使用割合」の値を、更新する。

* プログラム 4 1 4 は、割当元ティアの「使用割合」の更新後の値が割当元ティアの「ティア閾値」の値を超えているか否かの判断（以下、「判断 Z」と言う）を行う。割当元ティアの「ティア閾値」の値は、テーブル 4 0 2（図 6）に登録されている値、又は、テーブル 4 1 2（図 1 5）に登録されている値であるが、テーブル 4 1 2 に登録されている値が優先的に採用されて良い。

* プログラム 4 1 4 は、判断 Z の結果が肯定的であれば、下記（a）及び（b）のうちの少なくとも一つの処理を行って良い。

（a）プログラム 4 1 4 は、エラーを、管理装置 1 0 7（及び／又は、ライトコマンドの送信元のホスト 1 0 1）に通知して良い。管理者は、割当元テ

ィアについてエラーを知った場合、割当元ティアにRVOLを追加することができる。

(b) プログラム 4 1 4 は、再配置プログラム 4 1 5 を呼び出す。再配置プログラム 4 1 5 が、図 3 8 に示す処理を行う。図 3 8 に示す処理は、後に説明する。

[0161] (S 2 4 0 6)

プログラム 4 1 4 は、対象仮想ページに格納されるデータ要素のための CM 領域をキャッシュメモリから確保する。

[0162] (S 2 4 0 7)

プログラム 4 1 4 は、移行ページテーブル 4 1 1 における、対象仮想ページに対応した「移行ステータス」の値に応じて、CM 領域を確保する。具体的には、例えば、下記の通りである。

[0163] <<ケース A : 「移行ステータス」が“未移行”>>

例えば、以下の（具体例 A 1）及び（具体例 A 2）のうちのいずれかが行われる。

（具体例 A 1）

* プログラム 4 1 4 は、CM 領域を確保し、対象仮想ページに対応した移行先ティアにおけるフリーの実ページを、ライト先として決定する。より具体的には、例えば、プログラム 4 1 4 は、移行先ティアにおけるいずれかのフリーの実ページ内のデータ要素がキャッシュメモリにあれば、そのデータ要素が存在する CM 領域を確保し、且つ、そのフリーの実ページを、ライト先として決定する。移行先ティアにおけるフリーの実ページは、前述したように、テーブル 4 0 2（図 6）、及び 4 0 8（図 1 1）を基に特定することができる。

* プログラム 4 1 4 は、対象仮想ページに対応した「移行ステータス」を“移行済”に更新する。この結果、対象仮想ページについて、データ要素の移行は行われない。すなわち、言い換えれば、対象仮想ページについてデータ要素の移行がキャンセルされたことになる。

*プログラム414は、対象仮想ページに、元の実ページに代えて、決定したライト先の実ページを割り当てる。すなわち、プログラム414は、アロケーションテーブル407における、対象仮想ページのIDに対応した欄に、上記決定したライト先の実ページのID、及び、その実ページを有するティアのID（移行先ティアのID）を上書きする。

*プログラム414は、実ページ管理テーブル408（図11）における、対象仮想ページに割り当てられていた元の実ページの「ステータス」を、“フリー”に更新する。また、プログラム414は、そのテーブル408における、対象仮想ページに新たに割り当てられた実ページ（決定したライト先の実ページ）の「ステータス」を“使用中”に更新する。

（具体例A2）

*プログラム414は、CM領域を確保し、対象仮想ページに割り当てられている実ページ（移行元ティア内の実ページ）を、ライト先として決定する。より具体的には、例えば、プログラム414は、移行先ティアにおけるいずれのフリーの実ページ内のデータ要素もキャッシュメモリになれば、CM領域（例えば、対象仮想ページに割り当てられている実ページ内のデータ要素が存在する領域）を確保し、且つ、対象仮想ページに割り当てられている実ページ（移行元ティア内の実ページ）を、ライト先として決定する。

*プログラム414は、対象仮想ページに対応した「移行ステータス」を“中断”に更新して良い。これにより、例えば、対象仮想ページについては、再配置プログラム415によるデータ移行は行われない。なお、この「移行ステータス」の値は、例えば、プログラム414によって、後述のデステージが済んだ場合に、“中断”から“未移行”に戻されて良い。

[0164] <<ケースB： 「移行ステータス」が“移行中”>>

例えば、以下の（具体例B1）～（具体例B3）のうちのいずれかが行われる。

（具体例B1）

*プログラム414は、対象仮想ページに対応した「移行ステータス」を“

中断”に更新する。そして、プログラム 4 1 4 は、前述の（具体例 A 1）又は（具体例 A 2）を行う。（具体例 A 1）が行われた場合、対象仮想ページに対応した「移行ステータス」は、“移行済”に更新されることになる。（具体例 A 2）が行われた場合、対象仮想ページに対応した「移行ステータス」は、“未移行”に更新されることになる。

（具体例 B 2）

* プログラム 4 1 4 は、CM 領域を確保する。（具体例 B 2）の場合、デステージは、対象仮想ページに対応した「移行ステータス」が“移行済”に更新された後に行われる。それ故、ライト先（デステージ先）は、移行先ティア内の実ページである。なぜなら、その時点で、対象仮想ページには、移行先ティア内の実ページが割り当てられているからである。

（具体例 B 3）

* プログラム 4 1 4 は、CM 領域を確保する。（具体例 B 3）の場合、デステージは、対象仮想ページに対応した「移行ステータス」が“移行中”の最中に行われ、データ要素は、対象仮想ページに割り当てられている実ページと、移行先ティア内の実ページとの両方にデステージされる（いわゆる二重ライトが行われる）。

[0165] << ケース C : 「移行ステータス」が“移行済” >>

例えば、以下の処理が行われる。

* プログラム 4 1 4 は、CM 領域を確保する。このケースでは、ライト先（デステージ先）は、移行先ティア内の実ページである。なぜなら、対象仮想ページに対応した「移行ステータス」が“移行済”に更新されているので、対象仮想ページに割り当てられている実ページが、後に説明するように、移行先となった実ページに変更されるからである。

[0166] （S 2 4 0 8）

S 2 4 0 2 で特定された全ての仮想ページについて CM 領域が確保された場合、プログラム 4 1 4 は、確保された一以上の CM 領域にライト対象のデータを書き込み、完了をホスト 1 0 1 に応答する。

[0167] <ライト処理（デステージ）>。

[0168] 図25は、デステージ処理のフローを示す。

[0169] I/O制御プログラム414は、ライト対象のデータを構成する各データ要素について、デステージ処理を行う。

[0170] (S2501)

プログラム414は、デステージを行う。すなわち、プログラム414は、キャッシュメモリに存在する複数のデータ要素のうち実ページに未だ書き込まれていないデータ要素（ダーティのデータ要素）を実ページに書き込む。デステージ先の実ページは、その実ページのライト先となった仮想ページに割り当てられている実ページ（アロケーションテーブル407から特定される実ページ）である。なお、このとき、その仮想ページに対応した「移行ステータス」が“中断”になっていれば、プログラム414は、その「移行ステータス」を“未移行”に更新して良い。

[0171] (S2502)

プログラム414は、テーブル403（図7）又は404（図8）を基に、モニタテーブル409を更新する必要があるか否かを判断する。テーブル404（図8）が優先されて良い。具体的には、例えば、プログラム414は、以下の二つの判断P及び判断Qを行う。

（判断P）プログラム414は、S2501でのデステージ先の実ページを有するプール（以下、図25の説明において「対象プール」と言う）に対応した「再配置」が“ON”か否か、又は、そのデステージ先の実ページが割り当てられている仮想ページを有するVVOL（以下、図25の説明において「対象VVOL」と言う）に対応した「再配置」が“ON”か否かを判断する。この判断Pの結果が否定的の場合、S2503は行われな（すなわち、モニタテーブル409を更新する必要が無いということになる）。

（判断Q）判断Pの結果が肯定的の場合、プログラム414は、現在の時刻が、対象プールに対応した「モニタ時間帯」の値が表す時間帯内か否かを判断する。この判断Qの結果が否定的の場合、S2503は行われな。一方

、判断Pだけでなくこの判断Qの結果も肯定的の場合、S 2 5 0 3が行われる（すなわち、モニタテーブル4 0 9を更新する必要があるということになる）。

[0172] (S 2 5 0 3)

プログラム4 1 4は、モニタテーブル4 0 9を更新する。具体的には、例えば、下記の通りである。

*プログラム4 1 4は、テーブル4 1 2（図1 5）において、対象VVOLに対応した「ページ単位移行」が“ON”になっていれば、対象VVOLに対応したアクティブのモニタテーブル4 0 9を更新する。具体的には、例えば、プログラム4 1 4は、アクティブのモニタテーブル4 0 9における、デステージされたデータ要素のライト先仮想ページに対応する「合計I/O数」の値を更新し（例えば1インクリメントし）、且つ、「最終I/O時刻」の値を更新する。また、プログラム4 1 4は、例えば、そのライト先仮想ページについて、デステージ時点の時刻を、前述のI/O履歴テーブルに追加する。この場合、I/O履歴テーブルには、ライト先の仮想ページのIDが記録されて良い。例えば、そのI/O履歴テーブルを基に、後に、プログラム4 1 4は、アクティブのモニタテーブル4 0 9における「平均I/O数」及び「最大I/O数」を更新して良い。

[0173] <リード処理>。

[0174] 図2 6は、リード処理のフローを示す。

[0175] (S 2 6 0 0)

I/O制御プログラム4 1 4が、リードコマンドをホスト1 0 1から受信する。そして、プログラム4 1 4が、受信したリードコマンドが有するI/O先情報から、リード元のVVOL及び仮想ページを特定する。

[0176] (S 2 6 0 1)

プログラム4 1 4は、キャッシュメモリ上に、リード対象のデータがあるか否かを判断する。この判断の結果が否定的の場合、S 2 6 0 3が行われ、この判断の結果が肯定的の場合、S 2 6 0 2が行われる。

[0177] (S 2 6 0 2)

プログラム 4 1 4 は、キャッシュメモリ上のリード対象のデータを、ホスト 1 0 1 に送信する。この場合、リード元の V V O L に対応するモニタテーブル 4 0 9 は更新されない。実ページに対する I / O が行われていないからである。

[0178] (S 2 6 0 3)

プログラム 4 1 4 は、アロケーションテーブル 4 0 7 を基に、S 2 6 0 0 で特定された仮想ページに実ページが割り当てられているか否かを判断する。この判断の結果が否定的の場合、S 2 6 0 4 が行われ、この判断の結果が肯定的の場合、S 2 6 0 5 が行われる。

[0179] (S 2 6 0 4)

プログラム 4 1 4 は、所定の応答（例えば、ゼロを表すデータ）をホスト 1 0 1 に送信する。

[0180] (S 2 6 0 5)

プログラム 4 1 4 は、S 2 6 0 0 で特定された一以上の仮想ページの I D のうちの少なくとも一つが移行ページテーブル 4 1 1 に登録されており、且つ、その登録されている仮想ページ I D に対応する「移行ステータス」が“移行中”か否かを判断する。この判断の結果が否定的の場合、S 2 6 0 7 が行われ、この判断の結果が肯定的の場合、S 2 6 0 6 が行われる。

[0181] (S 2 6 0 6)

プログラム 4 1 4 は、S 2 6 0 0 で特定された一以上の仮想ページにそれぞれ割り当てられている一以上の実ページ（アロケーションテーブル 4 0 7 から特定される一以上の実ページ）からデータ要素を読み出す。プログラム 4 1 4 は、それらのデータ要素で構成したリード対象のデータを、ホスト 1 0 1 に送信する。

[0182] (S 2 6 0 7)

プログラム 4 1 4 は、「移行ステータス」が“移行中”の仮想ページについて、「移行ステータス」が“移行済”に更新されるまで待つ。その後、プ

プログラム４１４は、Ｓ２６００で特定された一以上の仮想ページにそれぞれ割り当てられている一以上の実ページ（アロケーションテーブル４０７から特定される一以上の実ページ）からデータ要素を読み出す。プログラム４１４は、それらのデータ要素で構成したリード対象のデータを、ホスト１０１に送信する。

[0183] （Ｓ２６０８）

プログラム４１４は、テーブル４０３（図７）又は４０４（図８）を基に、モニタテーブル４０９を更新する必要があるか否かを判断する。テーブル４０４（図８）が優先されて良い。具体的には、例えば、プログラム４１４は、以下の二つの判断Ｖ及び判断Ｗを行う。

（判断Ｖ）プログラム４１４は、リード元のＶＶＯＬに割り当てられているプール（以下、図２６の説明において「リード元プール」と言う）に対応した「再配置」が“ＯＮ”か否か、又は、リード元のＶＶＯＬに対応した「再配置」が“ＯＮ”か否かを判断する。この判断Ｖの結果が否定的の場合、Ｓ２６０９は行われない（すなわち、モニタテーブル４０９を更新する必要が無いということになる）。

（判断Ｗ）判断Ｖの結果が肯定的の場合、プログラム４１４は、現在の時刻が、対象プールに対応した「モニタ時間帯」の値が表す時間帯内か否かを判断する。この判断Ｗの結果が否定的の場合、Ｓ２６０９は行われない。一方、判断Ｖだけでなくこの判断Ｗの結果も肯定的の場合、Ｓ２６０９が行われる（すなわち、モニタテーブル４０９を更新する必要があるということになる）。

[0184] （Ｓ２６０９）

プログラム４１４は、モニタテーブル４０９を更新する。具体的には、例えば、下記の通りである。

*プログラム４１４は、テーブル４１２（図１５）において、リード元のＶＶＯＬに対応した「ページ単位移行」が“ＯＮ”になっていれば、リード元のＶＶＯＬに対応したアクティブのモニタテーブル４０９を更新する。具体

的には、例えば、プログラム 414 は、アクティブのモニタテーブル 409 における、リード元のライト先仮想ページに対応する「合計 I/O 数」の値を更新し（例えば 1 インクリメントし）、且つ、「最終 I/O 時刻」の値を更新する。また、プログラム 414 は、例えば、そのリード元の仮想ページについて、リード時点の時刻を、前述の I/O 履歴テーブルに追加する。この場合、I/O 履歴テーブルには、リード元の仮想ページの ID が記録されて良い。例えば、その I/O 履歴テーブルを基に、後に、プログラム 414 は、アクティブのモニタテーブル 409 における「平均 I/O 数」及び「最大 I/O 数」を更新して良い。

[0185] < I/O 頻度のモニタと再配置処理との関係 >。

[0186] 図 27 は、I/O 頻度のモニタと再配置との関係の概要を示す。

[0187] 実ページに対する I/O がモニタされ、実ページに対する I/O が行われた場合に、その実ページが割り当てられている仮想ページ又はその仮想ページを有する VVOL の I/O 頻度が更新される。つまり、モニタテーブル 409 が更新される。別の言い方をすれば、前述したように、I/O 制御プログラム 414 がリードコマンドをホスト 101 から受けたものの、実ページからデータ要素をリードしなかった場合には、モニタテーブル 409 は更新されない。

[0188] モニタテーブル 409 の更新に応じて、度数分布テーブル 410 が更新される。なお、度数分布テーブル 410 は、予め用意されていなくて良く、例えば、管理者から度数分布の表示が命じられた場合に、モニタテーブル 409 を基に度数分布テーブル 410 が作成され、そのテーブル 410 を基に度数分布が表示されてよい。なお、度数分布に代えて、モニタテーブル 409 が表示されてよい。

[0189] 再配置処理は、ティア判定処理と移行処理とを有する。ティア判定処理では、移行ページテーブル 411 が作成される。そのテーブル 411 は、モニタテーブル 409 が表す I/O 頻度と、ティア管理テーブル 402（図 6）内のティアレンジとを基に作成されてよい。移行処理では、作成された移行

ページテーブル 4 1 1 を基に、ページ単位のデータ移行が行われる。

[0190] <再配置処理>。

[0191] 再配置処理は、或るティアの使用割合がそのティアのティア閾値を超えた場合に行われても良いし、定期的に（例えば「再配置実行時刻」の値が表す時刻の都度に）行われても良い。

[0192] 図 2 8 は、再配置処理のフローを示す。なお、図 2 8 は、一つの V V O L（図 2 8 の説明において「対象 V V O L」と言う）についてのフローである。

[0193] (S 2 8 0 1)

再配置プログラム 4 1 5 は、対象 V V O L についてデータ移行を実行するか否かを判断する。具体的には、例えば、以下の処理が行われる。

* プログラム 4 1 5 は、V V O L 管理テーブル 4 0 4（図 8）において、対象 V V O L に対応した「自動実行」が“ON”になっているか否かを判断する。対象 V V O L に対応した「自動実行」の値の設定が無い場合、プログラム 4 1 5 は、プール管理テーブル 4 0 3（図 7）において、対象 V V O L に割り当てられているプール（以下、図 2 8 の説明において「対象プール」）に対応した「自動実行」が“ON”になっているか否かを判断する。

* 対象 V V O L 又は対象プールに対応した「自動実行」が“ON”の場合、プログラム 4 1 5 は、対象 V V O L 又は対象プールに対応した「再配置実行時刻」の値が表す時刻に現在の時刻が達しているか否かを判断する。この判断の結果が肯定的の場合、S 2 8 0 1 の判断の結果が肯定的であり、この判断の結果が否定的の場合、S 2 8 0 1 の判断の結果が否定的である。

* 対象 V V O L 又は対象プールに対応した「自動実行」が“OFF”の場合、プログラム 4 1 5 は、管理者から管理装置 1 0 5 を通じて、対象 V V O L についてデータ移行の指示を受け付けたならば、S 2 8 0 1 の判断の結果は肯定的となる。

[0194] プログラム 4 1 5 は、対象 V V O L 又は対象プールに対応した「自動実行」が“OFF”であれば（S 2 8 0 2 : NO）、S 2 8 0 6 を行う。

[0195] プログラム 4 1 5 は、対象 V V O L 又は対象プールに対応した「自動実行」が“ON”であれば（S 2 8 0 2 : Y E S）、S 2 8 0 7 を行う。

[0196] （S 2 8 0 6）

プログラム 4 1 5 は、管理者からティアレンジが設定されているか否かを判断する。この判断の結果が否定的の場合、S 2 8 0 7 が行われた後に S 2 8 0 8 が行われ、この判断の結果が否定的の場合、S 2 8 0 7 が行われることなく S 2 8 0 8 が行われる。

[0197] （S 2 8 0 7）

プログラム 4 1 5 は、ティアレンジ決定処理を行う。この処理によって、ティアレンジが決定される。ティアレンジ決定の方法としては、幾つか考えられる。例えば、予め、複数種類のティアレンジが用意されていて良く、プログラム 4 1 5 が、ティア毎に、それら複数種類のティアレンジから任意のティアレンジを選択して良い。

[0198] （S 2 8 0 8）

プログラム 4 1 5 は、ティア判定処理を行う。ティア判定処理では、前述したように、移行ページテーブル 4 1 1 が作成される。

[0199] （S 2 8 0 9）

プログラム 4 1 5 は、作成した移行ページテーブル 4 1 1 に従って、ページ単位の移行処理を行う。ページ単位のデータ移行では、対象プール内において、或るティア内の実ページから別のティア内の実ページにデータ要素が移行される。ページ単位の移行処理については、後に詳細に説明する。

[0200] <<ページ単位の移行処理>>。

[0201] 図 3 4 は、図 2 8 の S 2 8 0 8 での移行ページテーブル作成処理のフローを示す。

[0202] （S 3 4 0 1）

プログラム 4 1 5 は、対象 V V O L の先頭の仮想ページを選択する。

[0203] （S 3 4 0 2）

プログラム 4 1 5 は、アロケーションテーブル 4 0 7 を基に、選択された

仮想ページに実ページが割り当てられているか否かを判断する。この判断の結果が否定的であれば、S 3 4 0 5が行われ、この判断の結果が肯定的であれば、S 3 4 0 3が行われる。

[0204] (S 3 4 0 3)

プログラム 4 1 5 は、モニタテーブル 4 0 9 を基に、選択された仮想ページに割り当てられている実ページ内のデータ要素が存在すべきティアを変更すべきか否かを判断する。具体的には、例えば、下記の通りである。

* プログラム 4 1 5 は、選択された仮想ページに割り当てられている実ページを有するティア (S 3 4 0 3 の説明において「対象ティア」) を、テーブル 4 0 2 (図 6) から特定する。

* プログラム 4 1 5 は、選択された仮想ページの I/O 頻度 (例えば、対象 VVOL に対応したアクティブのモニタテーブル 4 0 9 が表す合計 I/O 数、平均 I/O 数、及び最大 I/O 数のうちの少なくとも一つ) と、対象ティアに対応した「ティアレンジ」の値とを基に、選択された仮想ページに割り当てられている実ページ内のデータ要素を、対象ティアから別のティアに移行すべきかどうかを判断する。例えば、選択された仮想ページの対象ティアに対応したティアレンジに、選択された仮想ページの I/O 頻度が収まっていなければ、データ要素が移行されるべきと判断される。

[0205] この S 3 4 0 3 の判断の結果が否定的であれば、S 3 4 0 5が行われ、この判断の結果が肯定的であれば、S 3 4 0 4が行われる。

[0206] (S 3 4 0 4)

プログラム 4 1 5 は、選択された仮想ページの I/O 頻度 (例えば、対象 VVOL に対応したアクティブのモニタテーブル 4 0 9 が表す合計 I/O 数、平均 I/O 数、及び最大 I/O 数のうちの少なくとも一つ) と、対象ティアを有するプール内の他の割当ティアに対応した「ティアレンジ」の値とを基に、移行先ティアを決定する。ここでは、例えば、選択された仮想ページの I/O 頻度が収まっているティアレンジに対応する割当ティア (対象 VVOL の割当ティア) が、移行先ティアとして決定される。そして、プログラ

ム 4 1 5 は、移行先ページテーブル 4 1 1 に、選択された仮想ページの I D と、対象ティアの I D と、決定された移行先ティアの I D と、移行ステータス “未移行” とを登録する。

[0207] (S 3 4 0 5)

プログラム 4 1 5 は、選択された仮想ページが、対象 V V O L の末端の仮想ページであるか否かを判断する。この判断の結果が否定的であれば、S 3 4 0 6 が行われ、この判断の結果が肯定的であれば、この移行ページテーブル作成処理が終了となる。

[0208] (S 3 4 0 6)

プログラム 4 1 5 は、対象 V V O L から、直前回に選択された仮想ページの次の仮想ページを選択する。その後、その選択された仮想ページについて、S 3 4 0 2 以降が行われる。

[0209] この移行ページテーブル作成処理により、ページ単位の移行処理のための移動ページテーブル 4 1 1 が作成される。

[0210] 図 3 5 は、ページ単位の移行処理の概要を示す。

[0211] ページ単位の移行処理によれば、テーブル 4 1 1 に登録されている仮想ページ毎に、その仮想ページに割り当てられている実ページ内のデータ要素が、その仮想ページに対応した移行先ティア内のフリーの実ページに移行される。その仮想ページに、移行元実ページに代えて移行先実ページが割り当てられる。図 3 5 に示すように、この移行処理によって、より上位のティアに移行されるデータ要素もあれば、より下位のティアに移行されるデータ要素もある。

[0212] どのティアについてどのようなティアレンジを設定するかによって、仮想ページの I / O 頻度に適したティアにその仮想ページ内のデータ要素を配置することができる。

[0213] 例えば、上位のティアほど、高いティアレンジが設定され、下位のティアほど、低いティアレンジが設定されとする。具体的には、例えば、ティア # 1 1 が、高速且つ高信頼のティア（例えば、R V O L 種類が “S S D” の

ティア)であって、高いティアレンジが設定されたとする。また、例えば、ティア#13が、低速又は低信頼のティア(例えば、RVOL種類が“HDD-SATA”のティア)であって、低いティアレンジが設定されたとする。また、例えば、ティア#12が、VVOL#50のデフォルトティアであって、中程度のティアレンジが設定されたとする。この場合、VVOL#50を構成する仮想ページに対するライト対象のデータ要素は、一旦は、デフォルトティア#12内の実ページに書き込まれる。その後、例えば、VVOL#50内の或る仮想ページ#AのI/O頻度が、高いティアレンジに収まるぐらいに高いI/O頻度になった場合には、その仮想ページ#Aに割り当てられている実ページ#X内のデータ要素が、ティア#12からティア#11に移行される。一方、例えば、VVOL#50内の或る仮想ページ#BのI/O頻度が、低いティアレンジに収まるぐらいに低いI/O頻度になった場合には、その仮想ページ#Bに割り当てられている実ページ#Y内のデータ要素が、ティア#12からティア#13に移行される。

[0214] このように、どのティアにどのようなティアレンジを設定するかと、仮想ページ毎のI/O頻度のモニタ結果と、ページ単位のデータ移行とにより、VVOL内のデータ要素を適切なティア(物理記憶資源)に配置することができる。

[0215] 図36は、ページ単位の移行処理のフローを示す。

[0216] (S3601)

再配置プログラム415が、対象VVOLに割り当てられているプールの「プールステータス」を“再配置中”に更新する。

[0217] (S3602)

プログラム415が、作成された移行ページテーブル411の先頭のレコードを選択する。

[0218] (S3603)

プログラム415は、選択されたレコードにおける「移行ステータス」の値を“移行中”に更新する。

[0219] (S 3 6 0 4)

プログラム 4 1 5 は、選択されたレコードが表す移行先ティアからフリーの実ページを選択し、選択した実ページ（移行先実ページ）に、移行元ティア内の移行元実ページに記憶されているデータ要素を移行する。移行元実ページは、対象仮想ページに割り当てられている実ページである。

[0220] (S 3 6 0 5)

プログラム 4 1 5 は、テーブルを更新する。具体的には、例えば、下記の通りである。

* プログラム 4 1 5 は、移行元実ページの代わりに移行先実ページが、選択されたレコードが表す仮想ページ（以下、S 3 6 0 5 の説明において「選択仮想ページ」）に割り当てるよう、アロケーションテーブル 4 0 7 を更新する。つまり、プログラム 4 1 5 は、選択仮想ページに、移行先実ページの ID と、移行先ティアの ID とを対応付ける。

* プログラム 4 1 5 は、選択仮想ページの I/O 頻度（例えば、アクティブのテーブル 4 0 9）を更新する。具体的には、例えば、選択仮想ページ又は対象 VVOL の合計 I/O 数が 1 インクリメントされる。

* プログラム 4 1 5 は、テーブル 4 0 2（図 6）における、移行先ティアの「使用割合」の値を、更新する。

[0221] (S 3 6 0 6)

プログラム 4 1 5 は、選択されたレコードにおける「移行ステータス」の値を“移行済”に更新する。

[0222] (S 3 6 0 7)

プログラム 4 1 5 は、選択されたレコードが、仮想ページテーブル 4 1 1 における末端のレコードであるか否かを判断する。この判断の結果が否定的であれば、S 3 6 0 8 が行われ、この判断の結果が肯定的であれば、この移行処理が終了となる。

[0223] (S 3 6 0 8)

プログラム 4 1 5 は、仮想ページテーブル 4 1 1 から、直前回に選択され

たレコードの次のレコードを選択する。その後、その選択されたレコードについて、S 3 6 0 3以降が行われる。

[0224] 以上が、ページ単位の移行処理のフローである。なお、ページ単位の移行処理の変形例として、例えば、下記の複数の変形例のうちの少なくとも一つが考えられる。

* S 3 6 0 2のS 3 6 0 3の間で、プログラム 4 1 5が、データ要素を移行先ティアに移行したと仮定して、移行先ティアの使用割合が移行先ティアのティア閾値を超えるか否かを判断してよい。超えると判断された場合、プログラム 4 1 5は、S 3 6 0 3～S 3 6 0 6をスキップしてS 3 6 0 7を行ってよい（すなわち、プログラム 4 1 5は、選択されたレコードについてデータ要素の移行を行わなくて良い）。或いは、超えると判断された場合、プログラム 4 1 5は、移行先ティアを割当ティアとしているV V O Lについて再配置処理を行うことで、移行先ティアにフリーの実ページを増やし、その後で、S 3 6 0 3を開始してもよい。

* S 3 6 0 5で移行先ティアの使用割合を更新したときに、プログラム 4 1 5が、移行先ティアの使用割合が移行先ティアのティア閾値を超えるか否かを判断してよい。超えたと判断された場合、プログラム 4 1 5は、管理装置 1 0 7に警告を表示して良く、及び/又は、移行先ティアを割当ティアとしているV V O Lについて再配置処理を行うことで、移行先ティアにフリーの実ページを増やしてよい。

[0225] <ティア閾値の変更>。

[0226] U I制御プログラム 4 1 3が、管理者からティア閾値の変更を受け付けることができる。

[0227] 図 3 7は、ティア閾値の変更処理のフローを示す。

[0228] (S 3 7 0 1)

プログラム 4 1 3が、プール I D及びティア I Dと、そのティア I Dについての変更後のティア閾値とを受け付ける。

[0229] (S 3 7 0 2)

プログラム 4 1 3 は、S 3 7 0 1 で受けたプール ID 及びティア ID に対応した「使用割合」の値が、変更後のティア閾値を超えているか否かを判断する。

[0230] (S 3 7 0 3)

プログラム 4 1 3 は、S 3 7 0 2 の判断の結果が肯定的の場合、エラーを管理装置 1 0 7 に表示する。

[0231] <使用割合の監視>。

[0232] 前述したように、I/O 制御プログラム 4 1 4 又は再配置プログラム 4 1 5 は、実ページにデータ要素を書き込んだ場合、テーブル 4 0 2 (図 6) における、その実ページを有するティアの「使用割合」の値を、更新する。その後、プログラム 4 1 4 又は 4 1 5 は、「使用割合」の更新後の値が、そのティアに対応する「ティア閾値」の値を超えたか否かを判断してよい。この判断の結果が肯定的であれば、プログラム 4 1 4 又は 4 1 5 は、図 3 8 に示す閾値オーバー対処処理を行って良い。

[0233] 図 3 8 は、閾値オーバー対処処理のフローを示す。

[0234] プログラム 4 1 4 又は 4 1 5 は、使用割合がティア閾値を超えたティアを有するプールの「プールステータス」が“再配置中”か否かを判断する (S 3 8 0 1)。

[0235] S 3 8 0 1 の判断の結果が肯定的の場合、プログラム 4 1 4 又は 4 1 5 は、管理装置 1 0 7 又はホスト 1 0 1 に、警告を出す (S 3 8 0 2)。その警告は、例えば、使用割合がティア閾値を超えたティアの ID と、そのティアを有するプールの ID と、そのプールの「プールステータス」の値 (“再配置中”) とを含んでよい。

[0236] S 3 8 0 1 の判断の結果が否定的の場合、プログラム 4 1 4 又は 4 1 5 は、管理装置 1 0 7 又はホスト 1 0 1 に、警告を出し (S 3 8 0 3)、そのティアの使用割合がそのティアのティア閾値以下になるよう再配置処理 (図 2 8) を実行する (S 3 8 0 4)。なお、S 3 8 0 3 での警告は、例えば、使用割合がティア閾値を超えたティアの ID と、そのティアを有するプールの

ＩＤと、そのプールについて再配置処理を行うことを意味するメッセージとを含んでよい。

[0237] <データ要素の集約>。

[0238] 本実施例では、データ要素の集約処理が行われて良い。

[0239] 図２１は、データ要素の集約処理の概要を示す。

[0240] Ｉ／Ｏ制御プログラム４１４又は再配置プログラム４１５は、ＶＶＯＬ＃３９について、定期的に、或いは、データ要素を実ページに書き込んだときに（例えば、ページ単位の移行処理の終了時に）、ＶＶＯＬ＃３９内のデータ要素の配置をチェックする。具体的には、例えば、プログラム４１４又は４１５は、ＶＶＯＬ＃３９の各割当ティア＃００、＃０１、＃０３について、ＶＶＯＬ＃３９内の全てのデータ要素のうちどの程度の割合のデータ要素が存在するか（つまり格納割合）を算出する。プログラム４１４又は４１５は、ＶＶＯＬ＃３９の割当ティア＃００、＃０１、＃０３に、格納割合が格納閾値を超える割当ティアがあるか否かを判断する。

[0241] プログラム４１４又は４１５は、格納割合が格納閾値を超える格納割合の割当ティアがある場合、その割当ティア＃００以外の割当ティア＃０１及び＃０３内のデータ要素（ＶＶＯＬ＃３９内のデータ要素）を、割当ティア＃００（具体的には、例えば、割当ティア＃００のフリーの実ページ）に移行する。

[0242] なお、その後、例えば、プログラム４１４は、ＶＶＯＬ＃３９内の仮想ページ（特に、実ページが割り当てられていない仮想ページ）に対してライトが発生した場合には、その仮想ページに、必ず、割当ティア＃００から実ページを割り当てて良い。

[0243] なお、割当ティア＃０１及び＃０３内のデータ要素（ＶＶＯＬ＃３９内のデータ要素）を割当ティア＃００に移行すると割当ティア＃００の使用割合が割当ティア＃００のティア閾値を超えてしまうようであれば、プログラム４１４又は４１５は、下記の（ｘ）又は（ｙ）を行って良い。

（ｘ）プログラム４１４は、割当ティア＃０１及び＃０３内のデータ要素を

割当ティア#00に移行しない。

(y) プログラム414は、対象VVOL以外のVVOL内のデータ要素であって割当ティア#00内のデータ要素を他のティアに移行することで割当ティア#00にフリーの実ページを増やす。

[0244] また、本実施例では、図28のS2807において、ティアレンジを自動で決定することができる。ティアレンジは、例えば、図示しないティアレンジ決定プログラムがMPに実行されることで決定される。

[0245] 以下、一つのプール（ティアレンジの自動決定の説明において「対象プール」と言う）を例に採り、ティアレンジ決定処理を説明する。

[0246] 図29は、ティアレンジ決定処理の概要を示す。

[0247] 対象プールが、3つのティア#01、#02及び#03を有するとする。ティア#01が最上位であり、ティア#02が2番目に上位であり、ティア#03が最下位のティアであるとする。言い換えれば、ティア#01の性能ポテンシャルが最も高く、ティア#02の性能ポテンシャルが2番目に高く、ティア#03の性能ポテンシャルが最も低いとする。「性能ポテンシャル」とは、再配置処理での、データ要素の配置の重みである。性能ポテンシャルについては後に詳述する。

[0248] ティアレンジ決定プログラムが、対象プールが割り当てられている全てのVVOLの度数分布を集約することで、対象プールについての度数分布4300を取得する。度数分布の横軸は、I/O頻度（負荷）であり、縦軸は、仮想ページ数である。

[0249] ティアレンジ決定プログラムは、上位のティアから下位のティアにかけて順に（又は、下位のティアから上位のティアにかけて順に）、ティアレンジを決定する。以下、ティア#01、#02、#03の順でティアレンジが決定されるとする。以下、各ティアについてのティアレンジの決定のために行われる処理を述べる（なお、図29において、ティアレンジの上限はひし形のマークであり、ティアレンジの下限は丸型のマークである）。

[0250] <<ティア#01のティアレンジ1の決定>>

*ティアレンジ決定プログラムは、ティア#01のティアレンジ1の上限として、無限大を設定する。つまり、最上位のティアのティアレンジの上限として、無限大が設定される。

*ティアレンジ決定プログラムは、度数分布4300において、矢印4301で示すように、I/O頻度が最も高い方から低い方へと仮想ページ数を累計していく。その累計値が、ティア#01の性能ポテンシャルに達した場合、ティアレンジ決定プログラムは、ティア#01の性能ポテンシャルに達する直前に加算された仮想ページ数に対応するI/O頻度を、ティア(0-1)境界値(ティア#00とティア#01の境界値)として設定する。

*ティアレンジ決定プログラムは、ティア(0-1)境界値より或る程度のマージンだけ低いI/O頻度を、ティア#01のティアレンジ1の下限として設定する。具体的には、例えば、その下限は、ティア(0-1)境界値より所定割合低いI/O頻度(例えば10%低いI/O頻度)である。

*ティアレンジ決定プログラムは、ティアレンジ1に収まるI/O頻度となっている全ての仮想ページ内のデータ要素がティア#01に移行されたと仮定して、ティア#01の使用割合がティア#01のティア閾値を超えるか否かを判断する。この判断の結果が肯定的の場合、ティアレンジ決定プログラムは、ティア#01の使用割合がティア#01のティア閾値を超えないような大きさにティアレンジ1を修正する。

[0251] <<ティア#02のティアレンジ2の決定>>

*ティアレンジ決定プログラムは、ティア#02のティアレンジ1の上限として、ティア(0-1)境界値より或る程度のマージンだけ高いI/O頻度を設定する。具体的には、例えば、その上限は、ティア(0-1)境界値より所定割合高いI/O頻度(例えば10%高いI/O頻度)である。

*ティアレンジ決定プログラムは、度数分布4300において、ティア(0-1)境界値から低い方へと仮想ページ数を累計していく。その累計値が、ティア#02の性能ポテンシャルに達した場合、ティアレンジ決定プログラムは、ティア#02の性能ポテンシャルに達する直前に加算された仮想ペー

ジ数に対応する I / O 頻度を、ティア (1 - 2) 境界値 (ティア # 0 1 とティア # 0 2 の境界値) として設定する。

*ティアレンジ決定プログラムは、ティア (1 - 2) 境界値より或る程度のマージンだけ低い I / O 頻度を、ティア # 0 2 のティアレンジ 2 の下限として設定する。

*ティアレンジ決定プログラムは、ティアレンジ 2 に収まる I / O 頻度となっている全ての仮想ページ内のデータ要素がティア # 0 2 に移行されたと仮定して、ティア # 0 2 の使用割合がティア # 0 2 のティア閾値を超えるか否かを判断する。この判断の結果が肯定的の場合、ティアレンジ決定プログラムは、ティア # 0 2 の使用割合がティア # 0 2 のティア閾値を超えないような大きさにティアレンジ 2 を修正する。

[0252] <<ティア # 0 3 のティアレンジ 3 の決定>>

*ティアレンジ決定プログラムは、ティア # 0 3 のティアレンジ 3 の上限として、ティア (1 - 2) 境界値より或る程度のマージンだけ高い I / O 頻度を設定する。

*ティアレンジ決定プログラムは、ティア # 0 2 のティアレンジ 2 の下限として、ゼロを設定する。つまり、最下位のティアのティアレンジの下限として、ゼロが設定される。

*ティアレンジ決定プログラムは、ティアレンジ 3 に収まる I / O 頻度となっている全ての仮想ページ内のデータ要素がティア # 0 3 に移行されたと仮定して、ティア # 0 3 の使用割合がティア # 0 3 のティア閾値を超えるか否かを判断する。この判断の結果が肯定的の場合、ティアレンジ決定プログラムは、ティア # 0 3 の使用割合がティア # 0 3 のティア閾値を超えないような大きさにティアレンジ 3 を修正する。

[0253] 以上のようにして、ティアレンジ 1 ~ 3 が決定される。なお、ティアレンジの上限及び／又は下限に、ティア境界値のマージンが付与されるのは、I / O 頻度の変動に応じて頻繁にデータ要素の移行が発生しないようにするためである。具体的には、例えば、マージンが無く、且つ、或る仮想ページの

I/O頻度がティア（0-1）境界値を頻繁に越えたり下回ったりすると、その仮想ページ内のデータ要素が頻繁にティア#01に移行されたりティア#02に移行されたりしてしまう。しかし、上記のようにマージンが設けられれば、そのような移行が行われる頻度を減らすことができる。

[0254] 以下、ティアレンジ決定処理をより詳細に説明する。

[0255] 図30は、ティアレンジ決定処理のフローを示す。

[0256] (S4401)

ティアレンジ決定プログラムは、対象プールの各ティアについて性能ポテンシャルを算出する。一つのティア（以下、S4401の説明で「対象ティア」と言う）を例に採ると、性能ポテンシャルは、例えば、下記（式1）、

性能ポテンシャル＝性能レベル×実効実ページ数（実効容量）・・・

（式1）

により算出される。この（式1）は、性能レベルが低くても実行実ページ数が多ければ、性能ポテンシャルが高いことを意味する。

[0257] なお、「実効実ページ数」とは、対象ティアを構成する全実ページのうちの、再配置処理においてデータ要素の移行先となり得る実ページの総数である。言い換えれば、プールには、再配置処理中にホストからのライト対象のデータ要素を書きこめるよう、再配置処理では移行先とされることのない実ページが設けられていて良い。

[0258] 「性能レベル」とは、対象ティアの単位容量（例えば一つの実ページ）当たりの性能ポテンシャルである。性能レベルは、例えば、下記（式2）、

値K＝RG性能/RG実効容量・・・（式2）

により算出される値Kに基づいている。（式2）における「RG性能」とは、例えば、リードの重みとライトの重みの比が1：1のときのRG性能であって、対象ティアの基になっているRGが単位時間当たりに処理可能な最大I/O数である。（式2）における「RG実効容量」とは、例えば、対象ティアの基になっているRGについての、データ要素を格納する実ページの総数である。なお、性能レベルは、上記値Kに代えて又は加えて、リードの重

みとライトの重みの比が参酌されても良い。その比は、例えば、ホストからの I / O の状況、キャッシュメモリの状況（例えば、ダーティのデータ要素とクリーンのデータ要素（実ページに格納済みのデータ要素）の配分）、R G の R A I D レベルのうちの少なくとも一つに基づいて変更されて良い。また、性能レベルは、管理者によって手動で設定されても良い。

[0259] (S 4 4 0 2)

ティアレンジ決定プログラムは、ティア境界値算出処理を行う。すなわち、ティアレンジ決定プログラムは、対象プールの各ティアについて算出された性能ポテンシャルを基に、ティア境界値を算出する。

[0260] (S 4 4 0 3)

ティアレンジ決定プログラムは、算出された各ティア境界値を基に、各ティアについてティアレンジを算出する。

[0261] (S 4 4 0 4)

ティアレンジ決定プログラムは、各ティアについてティアレンジが不適切か否かを判断する。一つのティア（以下、S 4 4 0 3 及び S 4 4 0 5 の説明において「対象ティア」と言う）を例に採って具体的に言うと、例えば、ティアレンジ決定プログラムは、対象ティアのティアレンジに収まる I / O 頻度となっている全ての仮想ページ内のデータ要素が対象ティアに移行されたと仮定して、対象ティアの使用割合が対象ティアのティア閾値を超えるか否かを判断する。

[0262] (S 4 4 0 5)

S 4 4 0 4 の判断の結果が肯定的の場合、ティアレンジ決定プログラムは、管理装置 1 0 7 に警告を表示するか、或いは、対象ティアの使用割合が対象ティアのティア閾値を超えないような大きさにティアレンジを修正する。

[0263] 図 3 1 は、図 3 0 の S 4 4 0 2 のティア境界値算出処理のフローを示す。

[0264] (S 4 5 0 1)

ティアレンジ決定プログラムが、対象プールが割り当てられている全ての V V O L の度数分布を集約することで、対象プールの度数分布を取得する。

度数分布は、対象プールが割り当てられている各VVOLに対応するモニターテーブル409を用いて取得することができる。

[0265] (S4502)

ティアレンジ決定プログラムが、対象プールのI/O頻度合計を算出する。具体的には、ティアレンジ決定プログラムが、対象プールが割り当てられている全てのVVOLの全ての仮想ページのI/O頻度を合計する。

[0266] (S4503)

ティアレンジ決定プログラムが、S4502で算出されたI/O頻度合計を基に、対象プールの各ティアについての配分I/O頻度を算出する。具体的には、ティアレンジ決定プログラムが、S4502で算出されたI/O頻度合計を、対象プールの複数のティアに配分する。具体的には、例えば、ティアレンジ決定プログラムが、図30のS4401で算出された、複数のティアの性能ポテンシャルの比に基づいて、それら複数のティアにI/O頻度合計を配分する。これにより、例えば、性能ポテンシャルの比が、ティア#01：ティア#02：ティア#03=4：2：1であり、且つ、I/O頻度合計が7000であった場合、ティア#01の配分I/O頻度は4000（ $7000 \times (4 / (4 + 2 + 1))$ ）となり、ティア#02の配分I/O頻度は2000（ $7000 \times (2 / (4 + 2 + 1))$ ）となり、ティア#03の配分I/O頻度は1000（ $7000 \times (1 / (4 + 2 + 1))$ ）となる。

[0267] (S4504)

ティアレンジ決定プログラムが、対象プールの各ティアの配分I/O頻度を基に、ティア境界値を決定する。例えば、S4503の例で言えば、ティア（0-1）境界値は、2000となり、ティア（1-2）境界値は、1000となる。

[0268] 以上のティアレンジ決定処理によれば、各ティアの性能ポテンシャルと対象プールの度数分布とを基に、ティア毎に、対象プールの全体のI/O状況に応じた適切なティアレンジを設定することができる。

[0269] 以上が、実施例1についての説明である。

[0270] なお、実施例 1 において、例えば、VVOL について、デフォルトティアは、必ず、VVOL の割当ティアのうち I/O 性能（例えば I/O 速度又はレスポンスタイム）が最大のティア（例えば割当ティアのうちの最上位のティア）であってもよい。VVOL の I/O 性能をなるべく大きくするためである。この場合、VVOL に書き込まれるデータ要素は、なるべく上位の割当ティアに配置されてよい。但し、割当ティアの使用容量がその割当ティアのティア閾値を超えないように制御される。

[0271] また、例えば、各プール、或いは、複数のプールに共通して、退避用のティアが設けられて良い。プログラム 4 1 4 及び 4 1 5 は、退避用のティアには、通常、ホストからのライトコマンドに従うライトでも、再配置処理における移行処理でも、データ要素を書き込まない。しかし、プログラム 4 1 4 又は 4 1 5 は、例えば、データ要素の書き込み先のティア（以下、この段落の説明において「対象ティア」）の使用割合がそのティアのティア閾値を超える場合、対象ティア内の任意の実ページ内のデータ要素（或いは、対象ティアのティアレンジに I/O 頻度が収まっていない仮想ページに割り当てられている、対象ティアにおける実ページ内のデータ要素）を、退避用のティアに移行し、それにより、対象ティアにフリーの実ページを増やしてもよい。この場合、プログラム 4 1 4 又は 4 1 5 は、対象ティア内の移行元実ページに割り当てられている仮想ページに、その移行元実ページに代えて、退避用のティア内の移行先実ページを割り当ててよい。その後、プログラム 4 1 4 又は 4 1 5 は、退避用ティア内の実ページに割り当てられている仮想ページの I/O 頻度が収まっているティアレンジが設定されているティアに、その退避用ティア内の実ページに格納されているデータ要素を移行してもよい。退避用ティアは、例えば、最上位のティアと同種の RVOL で構成されたティアであってもよい。

[0272] また、デフォルトティアは設定されなくて良い。VVOL に対するホスト 101 からのライト対象のデータ要素は、なるべく上位のティアに書き込まれて良い。上位のティアの使用割合がティア閾値を超えた場合には、その上

位のティアより一つ下位のティアにデータ要素が書き込まれて良い。

[0273] また、モニタテーブル 409B 及び 409A のうち 409A は必ずしも無くてもよい。この場合、モニタテーブル 409B が有する仮想ページ毎の I/O 頻度を合計することで VVOL の I/O 頻度が算出されても良い。

[0274] また、ページ単位の移行処理中断することもできる。例えば、再配置プログラム 415 は、管理者から中断の指示を受けた場合、移行ステータス“移行中”については移行ステータス“移行済”にするまで処理をすすめてから、ページ単位の移行処理を中断することができる。プログラム 415 は、例えば、中断の指示を受けてから一定時間後に、又は、再開の指示を受けたとき、移行ページテーブル 411 を基に移行処理を再開しても良い。或いは、プログラム 415 は、中断の指示を受けた場合、移行ページテーブル 411 を破棄し、次の移行処理の際には、アクティブのモニタテーブル 409 を基に移行ページテーブル 411 を作成しても良い。

[0275] また、本実施例では、Thin Provisioning の VVOL 203 は、ホスト 101 から仮想ページのあるアドレスへのライト要求に応じて実ページが割り当てられるものであるとしたが、本発明はこの場合に限られるものでなく、以下の例に挙げるものも好ましい。これらの例の場合でも、実際に使用される容量の削減を実現できるからである。

[0276] 1 つ目の例は、次のとおりである。すなわち、VVOL 203 の複数の仮想ページのうちの一部の仮想ページに対しては又は VVOL 203 の各仮想ページに含まれる一部の領域は、ホスト 101 から仮想ページのあるアドレスへのライト要求を受信する前に事前に実領域が割り当てられていて（プレアロケーション）、ホスト 101 から仮想ページのあるアドレスへのライト要求に応じて、既に割当済みの領域で十分でなく更なる割当が必要な場合に、追加の実領域が割り当てられる。

[0277] 2 つ目の例は、次のとおりである。すなわち、VVOL 203 の複数の仮想ページは、ホスト 101 から仮想ページのあるアドレスへのライト要求を受信する前に事前にゼロデータが格納されている実領域や制御用の領域（こ

これらの領域はSSDなどの記憶デバイス上の領域に対応するものである）に仮に割り当てられていて（仮アロケーション）、ホスト101から仮想ページのあるアドレスへのライト要求に応じて、仮の割当領域からデータが格納されるべき実ページへ割当先が変更される（本アロケーション）。この場合、本アロケーション実施前にホスト101から仮想ページのあるアドレスへリード要求を受信した場合、仮アロケーション先の実領域や制御用の領域に格納されているゼロデータがホスト101へ応答されることになる。

[0278] また、本実施例では、複数のプール201を一元的に管理するべく1つのテーブルで複数のプール201を管理することにしたが、本発明はそれに限られることなく、各プール201毎に実ページ207を一元的に管理するべく、プール201毎に管理用のテーブル（テーブル形式以外の形式の情報でも良い）を分けることも好ましい。

[0279] また、ティア303の種類は、本実施例で記載したSSD、HDD-SAS、HDD-SATAなどの属性毎に分類されることに限られるものでなく、FCインタフェースを有するHDDや、HDD形式でない記憶デバイス、例えばフラッシュメモリなどの半導体記憶媒体で構成されたデバイス、等の属性毎に分類されることも好ましい。この場合、各ティア303に関連付けられる仮想ページの属性や各ティア303に含まれる各実ページの属性も、本実施例で記載したSSD、HDD-SAS、HDD-SATAなどの属性毎に分類されることに限られるものでなく、FCインタフェースを有するHDDや、HDD形式でない記憶デバイス、例えばフラッシュメモリなどの半導体記憶媒体で構成されたデバイス、等の属性毎に分類されることになる。

[0280] さらに、ティア303の種類は、上記記憶デバイスのタイプという属性だけで分類されることに限られるものでもなく、上記記憶デバイスのタイプとRAID1からRAID5等のRAIDタイプとを組み合わせた属性や、上記記憶デバイスのタイプと上記記憶デバイスの性能（アクセス速度など）や記憶容量とRAID1からRAID5等のRAIDタイプとを組み合わせた属性毎に分類されることも好ましい。これらの場合、各ティア303に関連

付けられる仮想ページの属性や各ティア 303に含まれる各実ページの属性も、上記記憶デバイスのタイプという属性だけで分類されることに限られるものでもなく、上記記憶デバイスのタイプと RAID1 から RAID5 等の RAID タイプとを組み合わせた属性や、上記記憶デバイスのタイプと上記記憶デバイスの性能（アクセス速度など）や記憶容量と RAID1 から RAID5 等の RAID タイプとを組み合わせた属性毎に分類されることになる。

実施例 2

[0281] 以下、本発明の実施例 2 を説明する。その際、実施例 1 との相違点を主に説明し、実施例 1 との共通点については説明を省略或いは簡略する（これは、後の実施例 3 以降についても同様である）。

[0282] 本発明の実施例 2 では、I/O 頻度に代えて又は加えて、VVOL 毎又は仮想ページ毎の最終 I/O 時刻を基に、ページ単位の移行処理が行われる。最終 I/O 時刻は、実施例 1 で説明したように、実ページにデータ要素が書き込まれた場合に、モニタテーブル 409 に登録される。再配置プログラムは、仮想ページの I/O 頻度とその仮想ページ内のデータ要素を有するティアのティアレンジとの関係に代えて又は加えて、仮想ページの最終 I/O 時刻を基に、移行ページテーブル 411 に登録する仮想ページを決定する。例えば、最終 I/O 時刻が現在の時刻から所定時間以上に過去である仮想ページが、移行ページテーブル 411 に登録されて良い。その際、その仮想ページ内のデータ要素の移行先は、その仮想ページ内のデータ要素を有するティアよりも下位のティアとされて良い。なぜなら、最終 I/O 時刻が古い仮想ページ内のデータ要素は、今後も入出力される可能性が低く、より下位のティアに配置するのが好ましいと考えられるからである。

実施例 3

[0283] 本発明の実施例 3 では、シーケンシャルアクセスとランダムアクセスのどちらのアクセスパターンで I/O が行われるかを基に、ページ単位の移行処理が行われる。

[0284] I/O対象のデータが、シーケンシャルアクセスされるデータならば、I/O先のR/VOLの基になっているPDEVがHDDの場合に、HDDのシークの時間が短くて済む。この場合、比較的性能が低いHDDにI/O対象のデータが入っていても、シーケンシャルアクセスされるが故に、性能低下の度合いは小さいと考えられる。また、シーケンシャルアクセスされるデータを安価で性能の低いPDEVに移すことで、ランダムアクセスされる他のデータのために、性能の高いPDEVに基づく実ページのより多くをフリーにすることができる。このため、ストレージシステム全体のI/O性能の向上が期待できる。

[0285] 実施例3では、例えば、I/O制御プログラム414が、実ページにデータ要素を書き込んだ場合に、I/O先の仮想ページについてのアクセスパターン（シーケンシャルアクセスかランダムアクセスか）を、例えばアロケーションテーブル407に登録する。アロケーションテーブル407には、仮想ページ毎に、アクセスパターン毎の実行頻度が書き込まれても良い。

[0286] 再配置プログラム415は、例えば、移行ページテーブル411の作成において、仮想ページに対するアクセスパターンがシーケンシャルアクセスであれば（或いは、シーケンシャルアクセスの実行頻度が高ければ）、その仮想ページに対応した移行先として、より下位のティアを選択して良い。また、例えば、再配置プログラム415は、仮想ページのI/O頻度が、その仮想ページに割り当てられている実ページを有するティアのティアレンジの最大値より高くても、その実ページ内のデータ要素を移行対象としなくて良い。

実施例 4

[0287] 本発明の実施例4では、二つのモニタテーブル409は、交互に使用されることに代えて、所定のタイミングで編集不可能にCMPK119又は他の記憶資源に保存される。例えば、月曜日のモニタ結果を表すモニタテーブル409が月曜日のテーブル409として保存され、土曜日のモニタ結果を表すモニタテーブル409が土曜日のテーブル409として月曜日のモニタテ

ーブル４０９とは別に保存されてよい。このようにすることで、再配置処理の運用の運用方法の選択肢が増える。

- [0288] 実施例４では、例えば、再配置プログラム４１５が、再配置直前に、アクティブのモニタテーブル４０９を保存するかどうかを、管理者に問い合わせる等の方法により決定する。保存する場合には、再配置処理の直前に、プログラム４１５が、モニタテーブル４０９を管理装置１０７又は別の記憶資源に保存する。プログラム４１５は、移行ページテーブル４１１の作成のために参照するモニタテーブル４０９をどれにするかを、保存されている複数のモニタテーブル４０９から選択し、選択したモニタテーブル４０９を用いて、移行ページテーブル４１１を作成することができる。

実施例 ５

- [0289] 本発明の実施例５では、Ｉ／Ｏ頻度とティアレンジとの比較方法として、絶対値比較と相対値比較のいずれかを選択して設定することができる。
- [0290] 「絶対値比較」とは、ページ単位の移行処理について言えば、仮想ページのＩ／Ｏ頻度（例えば、合計Ｉ／Ｏ数、平均Ｉ／Ｏ数、又は最大Ｉ／Ｏ数）を、その仮想ページを有するＶＶＯＬ内の他の仮想ページのＩ／Ｏ頻度を考慮することなく、そのまま、ティアレンジと比較することである。絶対値比較によれば、他の仮想ページ又はＶＶＯＬのＩ／Ｏ頻度を考慮する必要がないので、比較に要する時間を短くすることができ、以って、再配置処理に要する時間を短くすることが期待できる。
- [0291] 「相対値比較」とは、ページ単位の移行処理について言えば、仮想ページのＩ／Ｏ頻度（例えば、合計Ｉ／Ｏ数、平均Ｉ／Ｏ数、又は最大Ｉ／Ｏ数）を、その仮想ページを有するＶＶＯＬ内の他の仮想ページのＩ／Ｏ頻度を基に修正し（正規化し）、修正後のＩ／Ｏ頻度をティアレンジと比較することである。具体的には、例えば、或る仮想ページ又はＶＶＯＬのＩ／Ｏ頻度が低くても、他の仮想ページ又は他のＶＶＯＬのＩ／Ｏ頻度がもっと低ければ、全体として、上記或る仮想ページ又はＶＶＯＬのＩ／Ｏ頻度はそれほど低く無いように修正される。また、例えば、或る仮想ページ又はＶＶＯＬのＩ

／＼頻度が高くても、他の仮想ページ又は他のVVOLのI／O頻度がもっと高ければ、全体として、上記或る仮想ページ又はVVOLのI／O頻度はそれほど高く無いように修正される。相対値比較によれば、モニタ時間帯での全体のI／O状況に応じた最適な配置が期待できる。

[0292] また、本実施例では、比較対象のI／O頻度として、例えば、平均I／O数とするか最大I／O数とするかを選択して設定することができる。平均I／O数によれば、モニタ時間帯全体でのI／O頻度に応じてデータ要素を配置することが期待できる。最大I／O数によれば、I／O頻度が一時的に大きくなってもI／O性能が下がらないようにデータ要素を配置することが期待できる。

[0293] 比較方法を絶対値比較とするか相対値比較とするか、及び、比較対象を平均I／O数とするか最大I／O数とするかは、例えば、前述したプール定義GUI又はVVOL作成GUIを通じて管理者が設定することができる。プールについての比較方法及び比較対象は、例えば図20に示すようにプール管理テーブル4403に登録され、VVOLについての比較方法及び比較対象は、例えば図20に示すようにVVOL管理テーブル4404に登録される（図32及び図20には、図7又は図8に示したカラムが省略されているが、実施例2でも、そのカラムは存在する）。なお、VVOLとそれに割り当てられているプールについて比較方法又は比較対象が異なっている場合には、実施例1と同様に、VVOLについての設定が優先されて良い。

[0294] 図33は、比較方法／比較対象の設定処理のフローを示す。

[0295] 比較方法／比較対象の設定処理は、例えば、プール作成処理及び／又はVVOL作成処理の中で行われてよい。

[0296] UI制御プログラム413が、或るプール又は或るVVOL（図33の説明において「対象プール又は対象VVOL」と言う）について、相対値比較の指定を受けた場合（S4201：NO）、対象プール又は対象VVOLについて、比較方法“相対値比較”をテーブル4403又は4404に設定する（S4202）。

- [0297] プログラム 4 1 3 が、対象プール又は対象 VVOL について、絶対値比較の指定を受けた場合（S 4 2 0 1 : YES）、対象プール又は対象 VVOL について、比較方法“絶対値比較”をテーブル 4 4 0 3 又は 4 4 0 4 に設定する（S 4 2 0 3）。
- [0298] プログラム 4 1 3 が、対象プール又は対象 VVOL について、最大 I/O 数の指定を受けた場合（S 4 2 0 4 : NO）、対象プール又は対象 VVOL について、比較対象“最大 I/O 数”をテーブル 4 4 0 3 又は 4 4 0 4 に設定する（S 4 2 0 5）。
- [0299] プログラム 4 1 3 が、対象プール又は対象 VVOL について、平均 I/O 数の指定を受けた場合（S 4 2 0 4 : YES）、対象プール又は対象 VVOL について、比較対象“平均 I/O 数”をテーブル 4 4 0 3 又は 4 4 0 4 に設定する（S 4 2 0 6）。
- [0300] なお、比較方法及び／又は比較対象は、管理者から手動の設定に代えて、自動で設定又は変更されても良い。例えば、プログラム 4 1 4 又は 4 1 5 は、モニタテーブル 4 0 9 B を基に、VVOL 又はプールの I/O の状況をチェックし、チェックした I/O の状況に応じて、テーブル 4 4 0 3 及び／又は 4 4 0 3 に設定済みの「比較方法」及び／又は「比較対象」の値を変更して良い。例えば、一つの VVOL における全ての仮想ページの I/O 頻度（又は、複数の VVOL の I/O 頻度）のばらつきが所定範囲に収まっていないことが検出された場合には、「比較方法」が、“相対値比較”から“絶対値比較”に変更され、逆に、そのばらつきが所定範囲に収まっていることが検出された場合には、「比較方法」が、“絶対値比較”から“相対値比較”に変更されてよい。また、例えば、一時的に I/O 頻度が急増することが検出された場合には、「比較対象」が、“平均 I/O 数”から“最大 I/O 数”に変更され、I/O 頻度のばらつきが所定範囲に収まっていることが検出された場合には、「比較対象」が、“最大 I/O 数”から“平均 I/O 数”に変更されて良い。
- [0301] 以上、本発明の幾つかの実施例を説明したが、本発明は、これらの実施例

に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

[0302] 例えば、少なくとも一つのPVOLが、仮想的なPVOLであっても良く、PVOLの基になるPDEVが、別のストレージシステムに存在しても良い。この場合、PVOLを構成する実ページに書き込まれるデータ要素は、そのPVOLの基になっている、別のストレージシステム内のPDEVに書き込まれて良い。

[0303] また、例えば、モニタテーブル409は、実ページに対するI/Oが行われたか否かに関わらず、仮想ページに対するI/Oが発生する都度に更新されてもよい。

[0304] また、例えば、ページ単位の移行処理において、移行先実ページの割当て先の仮想ページ又はその仮想ページを有するVVOLについて、移行先実ページにデータ要素が移行されたときにI/O頻度及び最終I/O時刻が更新されなくても良い。移行後に、その仮想ページ又はVVOLに対するI/Oに応じてその仮想ページ又はVVOLに割り当てられている実ページにI/Oが行われるときに、その仮想ページ又はVVOLのI/O頻度及び最終I/O時刻が更新されて良い。

[0305] また、例えば、RVOL種類は、RVOLの基になっているRGを構成するPEVの種類に代えて又は加えて、そのRGのRAIDレベル(RAID1、RAID5(3D+1P)、など)に基づいても良い。このため、例えば、第1のRGと第2のRGをそれぞれ構成するPDEVの種類が同じであっても、第1のRGと第2のRGのRAIDレベルが異なっていれば、第1のRGに基づく第1のRVOLと、第2のRGに基づく第2のRVOLは種類が異なることになり、それ故、第1のRVOLと第2のRVOLが異なるティアに属して良い。ティアは、RVOL種類に限らず、他の観点から定義されても良い。

[0306] また、例えば、一つの実ページの容量は、一つの仮想ページに複数の実ページが割り当てられるような容量であっても良い。この場合、ページ単位の

移行は、例えば、仮想ページに割り当てられている複数の移行元実ページから、複数の移行先実ページにデータ要素がそれぞれ移行されてよい。

[0307] また、例えば、一つの実ページの容量は、一つの実ページが複数の仮想ページに割り当てられるような容量であっても良い。この場合、ページ単位の移行処理では、複数の仮想ページが選択され、それら複数の仮想ページについて、一つの移行元実ページから移行先実ページにデータ要素が移行されてよい。

[0308] また、プールVOLは、外部のストレージシステムが提供するVOLに対応付けられた仮想的な論理ボリュームであっても良い。この場合、プールVOLの実ページに書き込まれるデータ要素は、そのプールVOLに対応する、外部のストレージシステムのVOLに書き込まれる。

[0309] また、I/O頻度の単位は、例えば、IOPS（1秒間当たりのI/Oの数）でも良い。

[0310] また、度数分布におけるI/O頻度幅は、1でも良い（つまり、I/O頻度毎に仮想ページ数がカウントされても良い）。

符号の説明

[0311] 103…ストレージシステム

請求の範囲

- [請求項1] 複数の記憶デバイスから提供される複数の記憶領域を記憶領域の属性に従って分類された複数の記憶領域グループを管理し、
- 論理ボリューム内の複数のアドレスに含まれる少なくとも1つのアドレスへのライト要求に応じて前記複数の記憶領域に含まれる少なくとも1つの記憶領域が割り当てられる前記論理ボリュームの複数の提供し、
- 前記論理ボリューム内の前記少なくとも1つのアドレスのアクセス状況に応じて、前記ライト要求によって前記少なくとも1つのアドレスにライトされているデータを、前記複数の記憶領域グループの1つに含まれる前記少なくとも1つの記憶領域から前記複数の記憶領域グループに含まれる他の記憶領域グループ内の少なくともひとつの記憶領域へ移行する、ことを特徴とするストレージシステムの制御方法。
- [請求項2] 請求項1に記載の方法であって、
- 前記複数の記憶領域グループに含まれる第一の記憶領域グループと第二の記憶領域グループとは、それぞれのグループに相応しい前記アクセス状況の程度をグループ毎に属性として有しており、前記第一と前記第二の記憶領域グループとの前記アクセス状況の程度はグループ毎に異なっているものの一部重複しており、
- 前記移行処理においては、前記複数の記憶領域グループの1つに含まれる前記少なくとも1つの記憶領域が、前記第一又は前記第二の記憶領域グループに含まれており且つ前記少なくとも1つのアドレスの前記アクセス状況が前記アクセス状況の程度のうち前記一部重複している範囲内である場合、前記データの移行は実施されない、ことを特徴とするストレージシステムの制御方法。
- [請求項3] 一以上のホストに接続されるストレージシステムであって、
- 複数種類の物理記憶デバイス群と、
- 記憶制御装置と

を備え、

前記記憶制御装置が、

記憶資源と、

前記複数種類の物理記憶デバイス群、前記記憶資源及び前記ホストに接続されるコントローラと

を有し、

物理記憶デバイス群は、一以上の物理記憶デバイスで構成されており、

前記複数種類の物理記憶デバイス群が、複数種類の実ページ群の基になっており、

実ページ群の種類は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいており、

実ページ群が、複数の実ページを有し、

前記コントローラが、一以上の仮想ボリュームを前記一以上のホストに提供し、

前記仮想ボリュームは、複数の仮想ページで構成される仮想的な論理ボリュームであり、

前記仮想ページは、仮想的な記憶領域であり、

前記実ページは、実体的な記憶領域であり、

前記記憶資源が、

どの仮想ページにどの実ページが割り当てられているかを表す情報である割当て情報と、

仮想ページ毎の I / O (Input/Output) 頻度を表す情報であるモニタ情報と、

実ページ群の種類毎に設定された I / O 頻度レンジを表す情報を含んだ情報である実ページ群管理情報と

を記憶し、

前記コントローラが、前記ホストからライトコマンドを受信し、前

記ライトコマンドからライト先の仮想ページを特定し、特定した仮想ページにフリーの実ページを割り当てよう前記割当て情報を更新し、割り当てた実ページにライト対象のデータを格納し、

前記コントローラが、仮想ページ又はその仮想ページに割り当てられる実ページに対する I / O に応じて、前記モニタ情報における、その仮想ページの I / O 頻度を更新し、

前記実ページは、実体的な記憶領域であり、実ページに格納されるデータはその実ページの基になっている物理記憶デバイスに記憶され、

前記コントローラが、以下の (A) 乃至 (C) の処理を含んだ処理である移行処理を行い、

(A) 前記モニタ情報から特定された I / O 頻度が所定の条件に適合していない仮想ボリューム又は仮想ページを特定する、

(B) 特定された仮想ボリューム又は仮想ページに割り当てられている、第 1 種の実ページ群内の第 1 の実ページ、に格納されているデータを、第 2 種の実ページ群内の第 2 の実ページに移行する、

(C) 前記第 1 の実ページの割当て先の仮想ページである第 1 の仮想ページに前記第 1 の実ページに代えて前記第 2 の実ページを割り当てよう前記割当て情報を更新する、

前記所定の条件とは、前記第 1 種の実ページ群の I / O 頻度レンジに収まることであり、

前記第 2 種の実ページ群は、前記 (A) で特定された仮想ボリューム又は仮想ページの I / O 頻度に I / O 頻度レンジが収まっている実ページ群であり、

前記コントローラは、I / O 頻度レンジの決定処理を行い、その決定処理において、前記モニタ情報を基に、各種の実ページ群の I / O 頻度レンジを決定する、

ストレージシステム。

[請求項4] 請求項3記載のストレージシステムであって、
前記コントローラが、前記決定処理において、以下の（a）及び（b）の処理を行う、
（a）前記モニタ情報を用いて、I／O頻度と仮想ページ数との関係を表す情報である関係情報を作成する、
（b）前記関係情報を基に、各種実ページ群のI／O頻度レンジを決定する、
ストレージシステム。

[請求項5] 請求項3記載のストレージシステムであって、
前記コントローラが、前記（b）の処理において、以下の（b1）乃至（b3）の処理を行い、
（b1）前記関係情報を用いて、前記複数種類の実ページ群から実ページが割り当てられている一以上の仮想ボリューム内の全ての仮想ページのI／O頻度の合計であるI／O頻度合計を算出する、
（b2）各種実ページ群の性能値の比と、前記I／O頻度合計とを基に、実ページ群の境界のI／O頻度である境界値を実ページ群の境界毎に算出する、
（b3）各種実ページ群のI／O頻度と実ページ群の境界毎の境界値とを基に、各種実ページ群のI／O頻度レンジを決定する、
実ページ群の前記性能値は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいて決定された、性能に関する値であり、
前記実ページ群の境界とは、実ページ群とその上位及び／又は下位の実ページ群との境であり、
各種実ページ群のI／O頻度レンジの最大値及び／又は最小値は、その実ページ群とその上位及び／又は下位の実ページ群との境の境界値に基づいて決定された値である、
ストレージシステム。

- [請求項6] 請求項5記載のストレージシステムであって、
 実ページ群の前記性能値は、その実ページ群における、前記移行処理での移行先となり得る実ページの数と、その実ページ群内の一つの実ページあたりの性能値との掛け算により算出された値であり、
 実ページ群内の一つの実ページあたりの性能値は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいて決定された、性能に関する値である、
 ストレージシステム。
- [請求項7] 請求項3記載のストレージシステムであって、
 前記 I / O 頻度レンジの最大値は、その I / O 頻度レンジに対応した大きい方の境界値より大きい、及び／又は、その I / O 頻度レンジの最小値は、その I / O 頻度レンジに対応した小さい方の境界値より小さい、
 ストレージシステム。
- [請求項8] 請求項3記載のストレージシステムであって、
 前記コントローラが、前記決定処理において、以下の処理を行い、
 （a）前記モニタ情報を用いて、前記複数種類の実ページ群から実ページが割り当てられている一以上の仮想ボリューム内の全ての仮想ページの I / O 頻度の合計である I / O 頻度合計を算出する、
 （b）各種実ページ群の性能値の比と、前記 I / O 頻度合計とを基に、実ページ群の境界の I / O 頻度である境界値を実ページ群の境界毎に算出し、
 （c）各種実ページ群の I / O 頻度と実ページ群の境界毎の境界値とを基に、各種実ページ群の I / O 頻度レンジを決定し、
 実ページ群の前記性能値は、その実ページ群の基になっている物理記憶デバイス群の属性に基づいて決定された、性能に関する値であり、
 前記実ページ群の境界とは、実ページ群とその上位及び／又は下位

の実ページ群との境であり、

各種実ページ群の I / O 頻度レンジの最大値及び／又は最小値は、その実ページ群とその上位及び／又は下位の実ページ群との境の境界値に基づいて決定された値である、
ストレージシステム。

[請求項9]

請求項 3 記載のストレージシステムであって、

前記コントローラが、前記決定処理において、以下の (d) 及び (e) の処理を行い、

(d) 実ページ群の I / O 頻度レンジに収まる I / O 頻度となっている全ての仮想ページに割り当てられている実ページ内のデータ要素がその実ページ群に移行されたと仮定して、その実ページ群の使用割合がその実ページ群の割合閾値を超えるか否かを判断する、

(e) 前記 (d) での判断の結果が否定的の場合に、決定された I / O 頻度レンジを前記実ページ群管理情報に含める、

実ページ群の前記使用割合とは、その実ページ群を構成する実ページの数に対する、仮想ページに割り当てられている実ページの数の割合であり、

前記実ページ群の割合閾値とは、前記使用割合の閾値である、
ストレージシステム。

[請求項10]

請求項 9 記載のストレージシステムであって、

前記コントローラが、前記決定処理において、更に、以下の (f) の処理を行う、

(f) 前記 (d) での判断の結果が肯定的の場合に、実ページ群の使用割合がその実ページ群の割合閾値を超えないような大きさに I / O 頻度レンジを修正する、
ストレージシステム。

[請求項11]

請求項 10 記載のストレージシステムであって、

前記コントローラが、前記決定処理において、更に、以下の (g)

の処理を行う、

(g) 修正後の I / O 頻度レンジを前記実ページ群管理情報に含める

、

ストレージシステム。

[請求項12]

請求項3記載のストレージシステムであって、

前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I / O 頻度とは、その仮想ボリューム又は仮想ページの、前記モニタ情報に含まれている I / O 頻度である、

ストレージシステム。

[請求項13]

請求項3記載のストレージシステムであって、

前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I / O 頻度とは、前記モニタ情報における、その仮想ボリュームとは別の仮想ボリュームの I / O 頻度、又はその仮想ページとは別の仮想ページの I / O 頻度、との相対的な値である、

ストレージシステム。

[請求項14]

請求項3記載のストレージシステムであって、

前記コントローラは、前記(A)での I / O 頻度を絶対値とするか相対値とするかの指定を受け、

絶対値とすることの指定を受けた場合、前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I / O 頻度とは、その仮想ボリューム又は仮想ページの、前記モニタ情報に含まれている I / O 頻度であり、

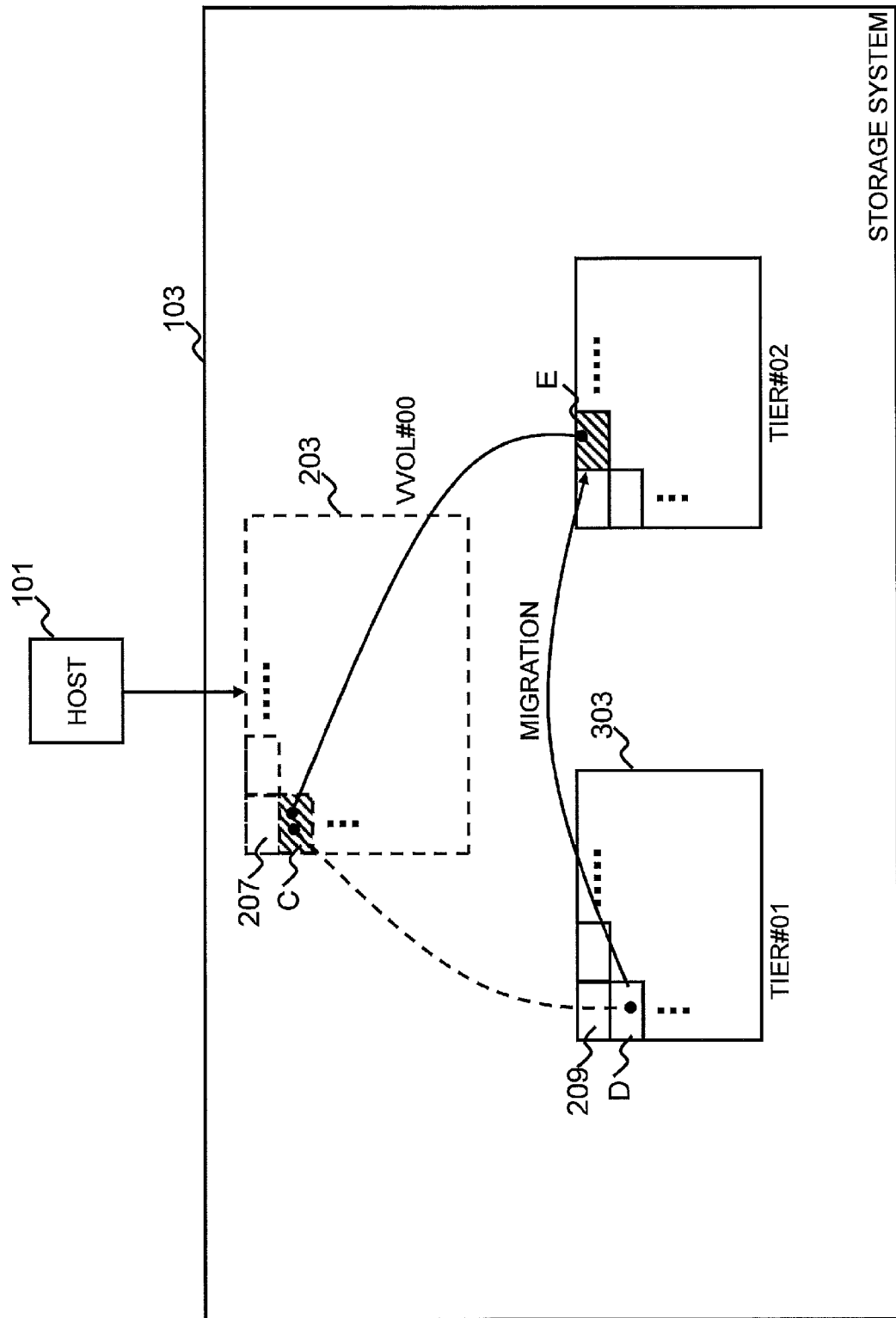
相対値とすることの指定を受けた場合、前記モニタ情報から特定された、仮想ボリューム又は仮想ページの I / O 頻度とは、前記モニタ情報における、その仮想ボリュームとは別の仮想ボリュームの I / O 頻度、又はその仮想ページとは別の仮想ページの I / O 頻度、との相対的な値である、

ストレージシステム。

- [請求項15] 請求項3記載のストレージシステムであって、
前記（A）での I ／ O 頻度は、或る時間における I ／ O 数の平均である、
ストレージシステム。

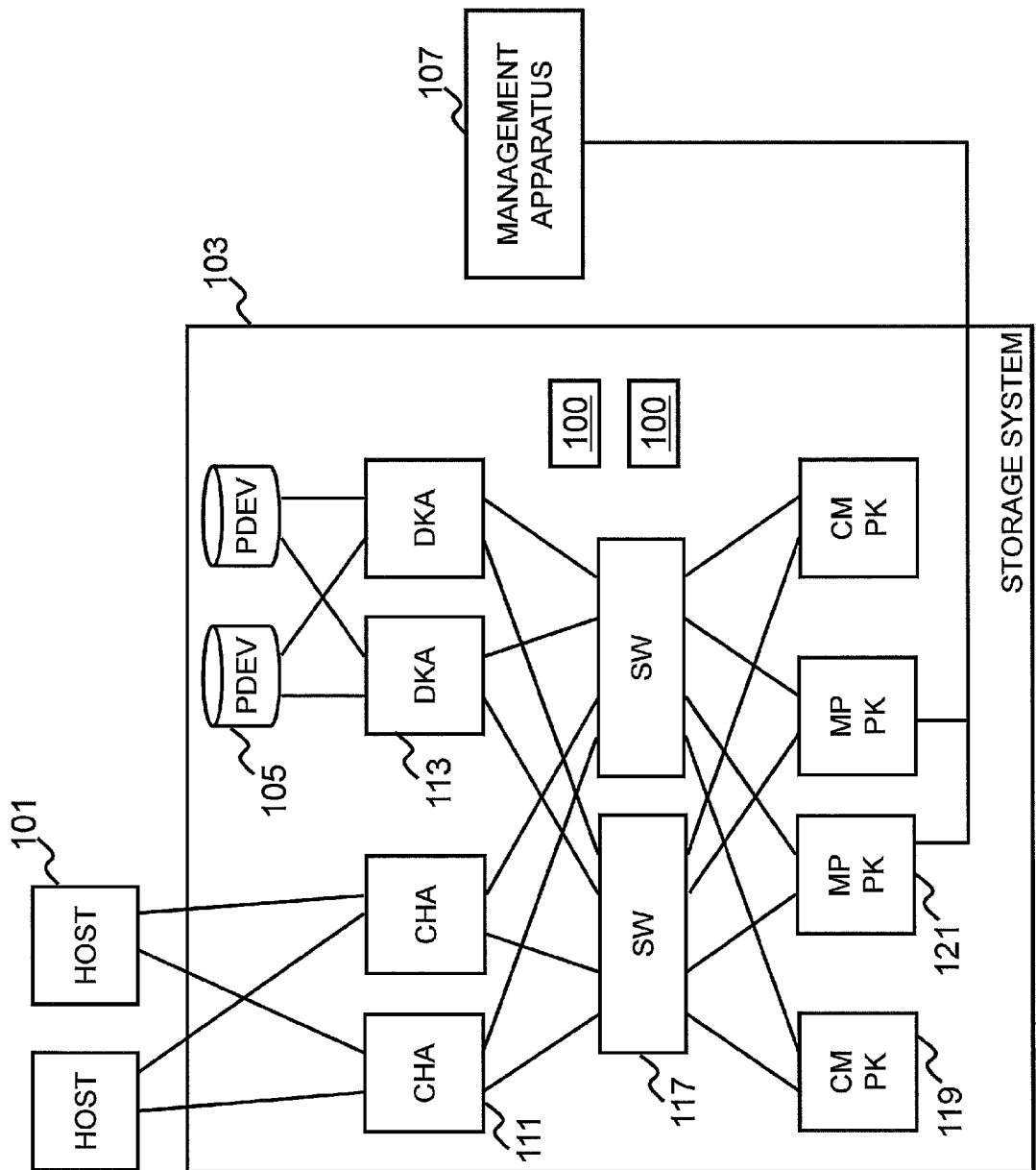
[図1]

FIG.1



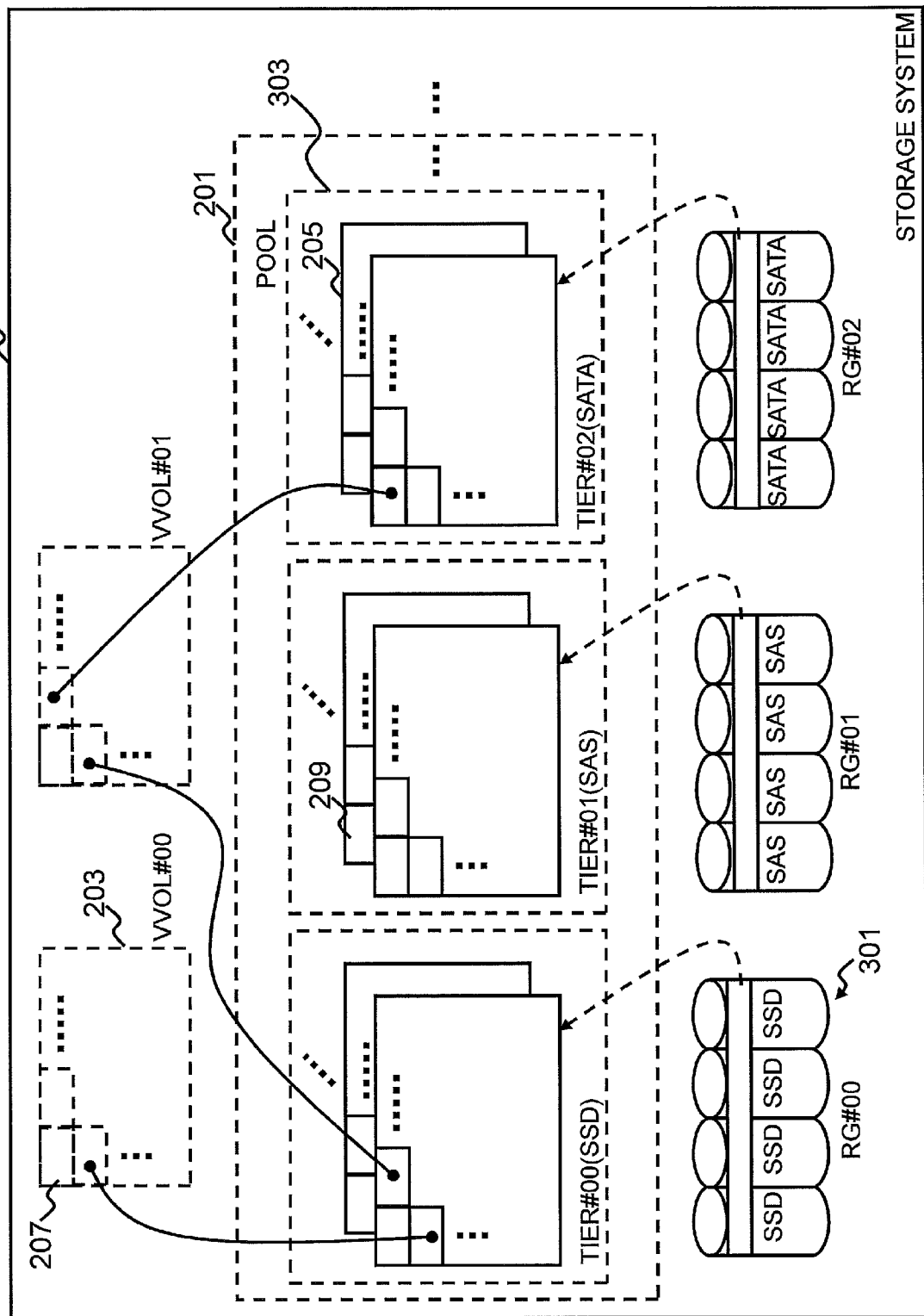
[図2]

FIG.2



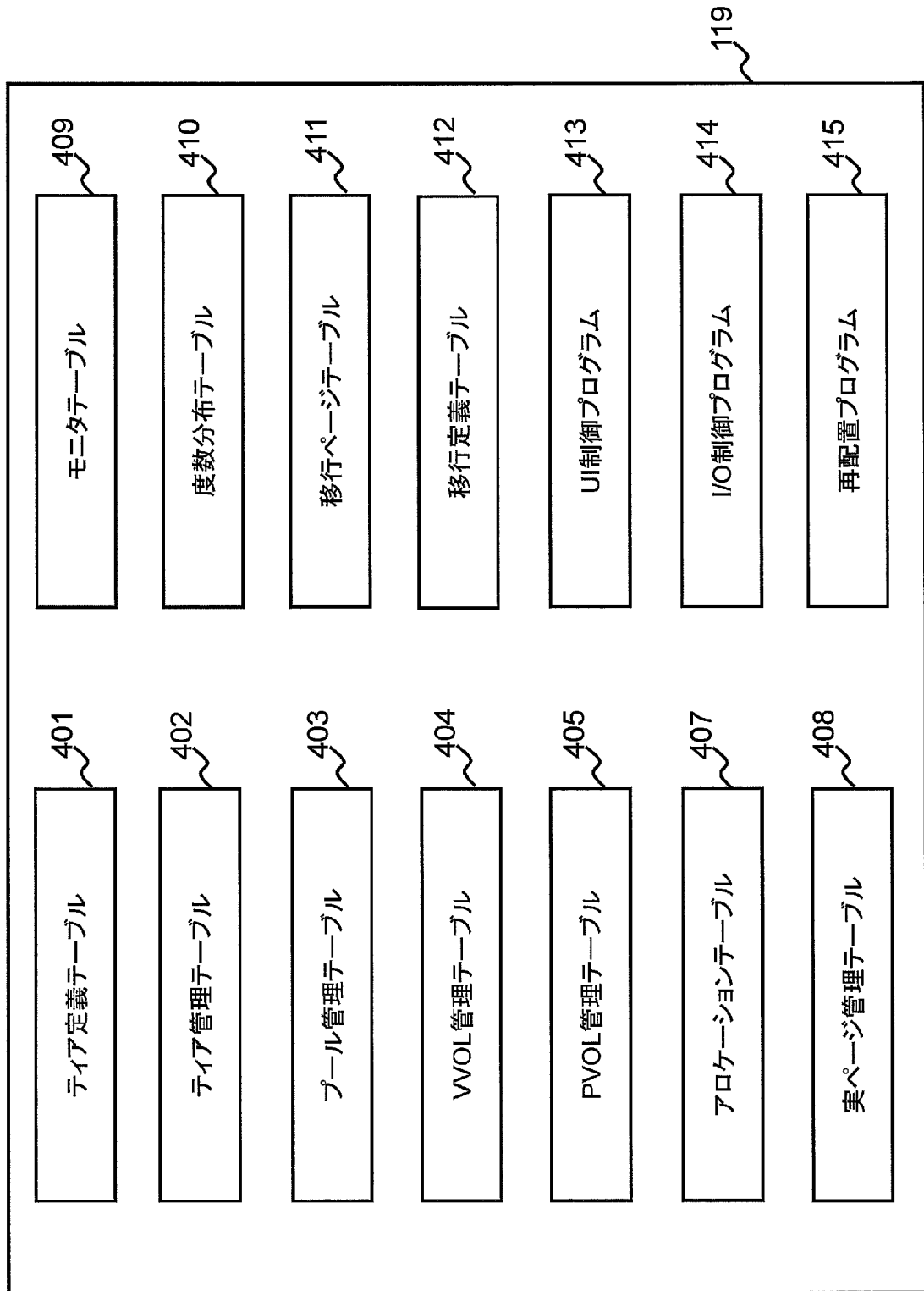
[図3]

FIG.3



[図4]

FIG.4



[図5]

FIG.5

RVOL種類	TIER ID
SSD	00
HDD-SAS(15K)	01
HDD-SAS(10K)	02
HDD-SATA	03
...	...

401

[図6]

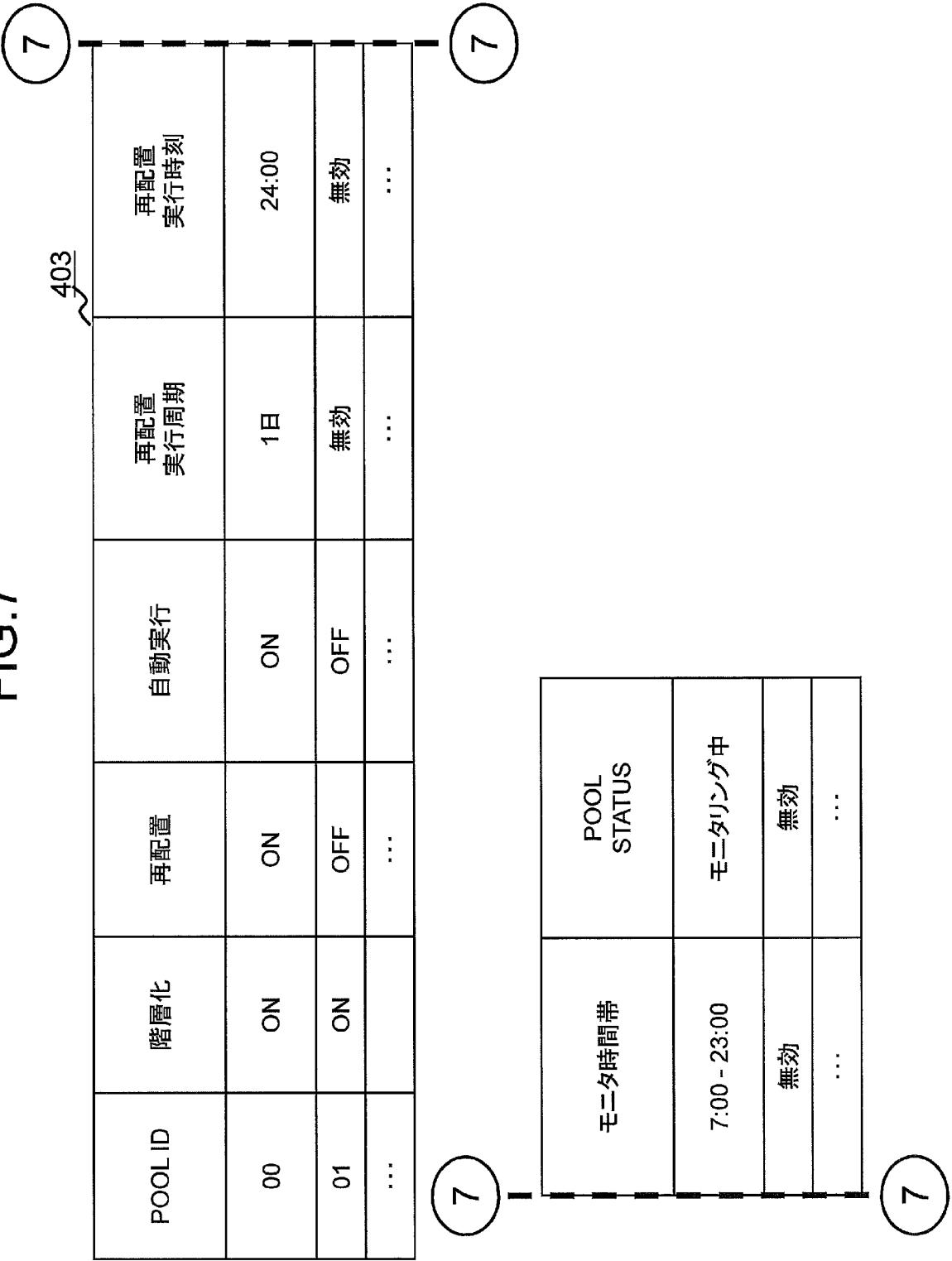
FIG.6

402

POOL ID	TIER ID	TIER RANGE	PIVOL ID	TIER閾値	使用割合	実ページID
00	00	1000-5000	00	90%	75%	00000-00999
			02			
			...			
	01	0-999	04	80%	75%	01000-02999
			...			
			...			
...	...	...	...	...	...	
	...	...	...	...	...	

[図7]

FIG.7



[図8]

FIG.8

8					8				
404									
WVOL ID	POOL ID	ALLOCATED TIER ID	DEFAULT TIER ID	容量	再配置	自動実行	再配置 実行周期	再配置 実行時刻	モ二タ時間帯
00	00	00-01	00	1TB	ON	ON	1日	24:00	7:00 - 23:00
01	00	02	02	500GB	OFF	OFF	無効	無効	無効
...		...	...	...	...	...	...	...	...

[図9]

FIG.9

405

RG ID	PDEV ID	TYPE	PVOL ID	容量
00	01-04	SSD	01	500GB
			02	500GB
01	05-08	HDD-SAS(15K)	03	500GB
			04	500GB
			05	500GB
...	...	...	...	...

[図10]

FIG.10

407

WVOL ID	仮想ページ ID	POOL ID	実ページ ID	TIER ID
00	00000	00	00010	01
	00001	00	00100	00
	...	00	...	...
01	00000	00	00020	01
	00001	-	-	-
	...	...	...	...
...	...	...	...	...

[図11]

FIG.11

408

POOL ID	実ページ ID	STATUS
00	00000	ON USE
	00001	FREE
	...	...
...	...	...

[図12]

FIG.12

409

VVOL ID	仮想ページID	合計I/O数	平均I/O数	最大I/O数	最終I/O時刻
00	00000	1000	100	200	17:00
	00001	20	2	5	12:00
	...	...	...	...	...
01	00000	—	—	—	—
	00001	—	—	—	—
	...	...	...	...	...
...	...	...	...	...	...

[図13]

FIG.13

410

VVOL ID	平均I/O数	ページ数
00	0	561
	1	27
	2	141
	3	57
	4-7	72
	8-15	236
	16-31	0
...	...	...

[図14]

FIG.14

411

仮想ページ ID	移行元 TIER ID	移行先 TIER ID	移行 STATUS
10001	00	01	移行済
10002	01	02	移行済
10005	02	03	移行中
10008	03	01	移行中
10014	01	00	未移行
...	...		...

[図15]

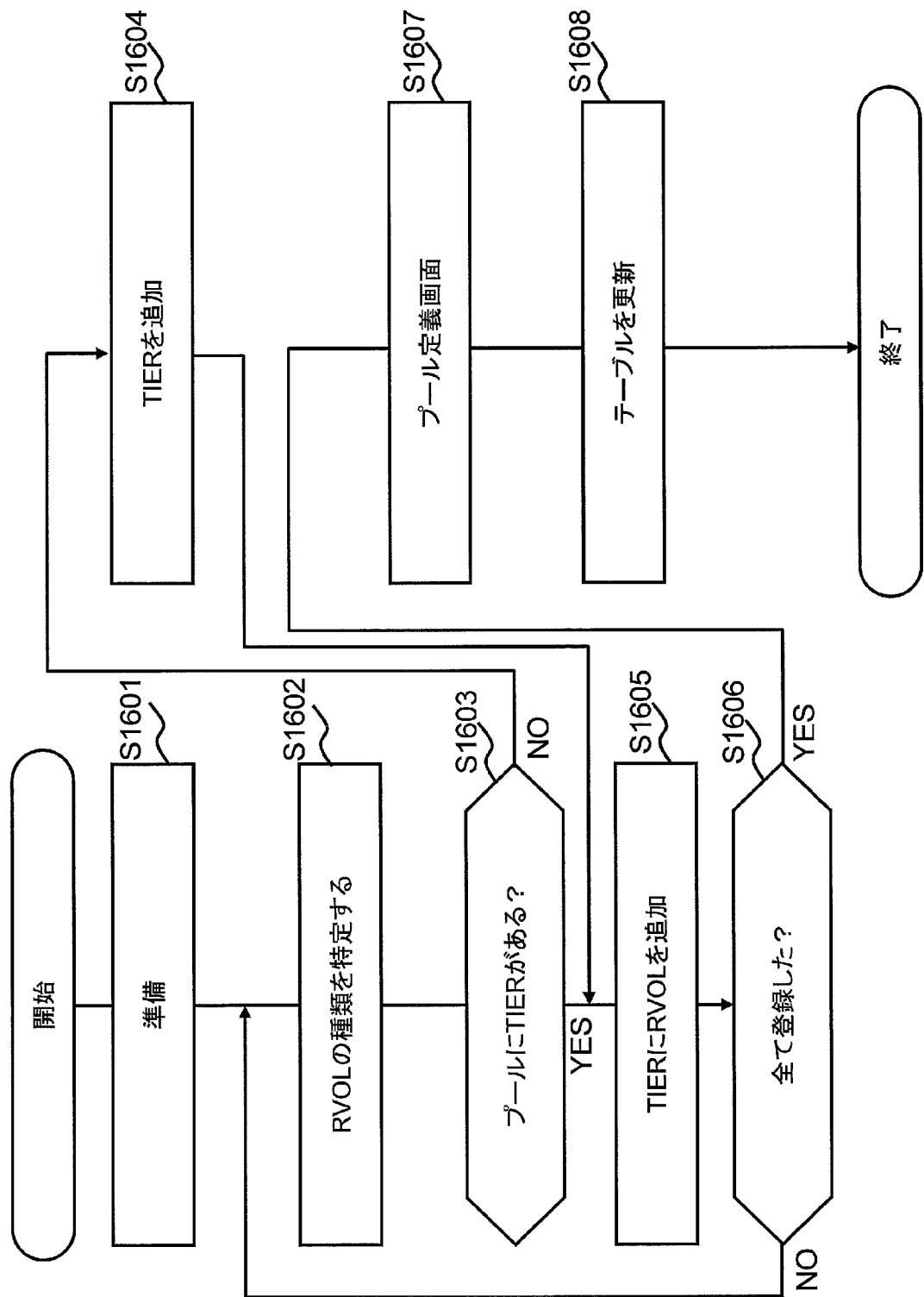
FIG.15

412

VVOL ID	TIER維持	ALLOCATED TIER ID	格納閾値
00	ON	00	80%
		01	無効
		03	無効
01	OFF	01	70%
		02	無効
02	ON	02	無効
		03	70%
...	...	...	...

[図16]

FIG.16



[図17]

FIG.17

1700

プール定義

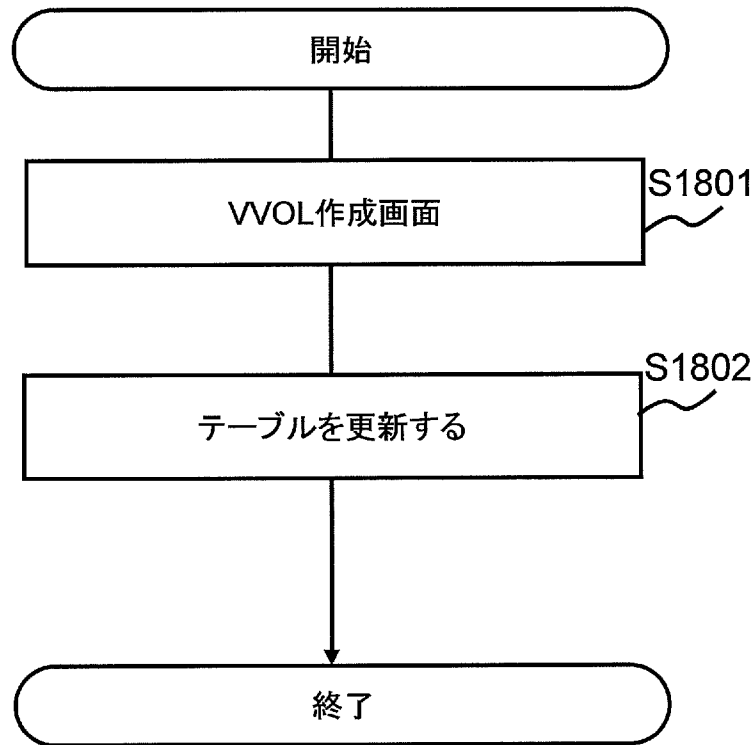
POOL ID	00	
階層化	☒ ON	☐ OFF
再配置	☒ ON	☐ OFF
自動実行	☐ ON	☒ OFF
実行周期	1 DAY	
実行時刻	24:00	
モニタ時間帯	FROM 07:00	TO 23:00
TIER閾値	☐ 共通 ☒ 個別 90% TIER 00 01 03 90% 80% 70%	

適用

中止

[図18]

FIG.18



[図19]

FIG.19

1900

WOL作成

WOL ID

00

容量

1TB

POOL ID

00

再配置

☒ ON ☐ OFF

自動実行

☒ ON ☐ OFF

実行周期

1 DAY

実行時刻

24:00

モニタ時間帯

FROM 07:00 TO 23:00

ALLOCATED TIER

TIER 00

☒

TIER 01

☒

TIER 03

☒

格納閾値

80%

☐

無効

☐

無効

☐

デフォルトTIER

00

適用

中止

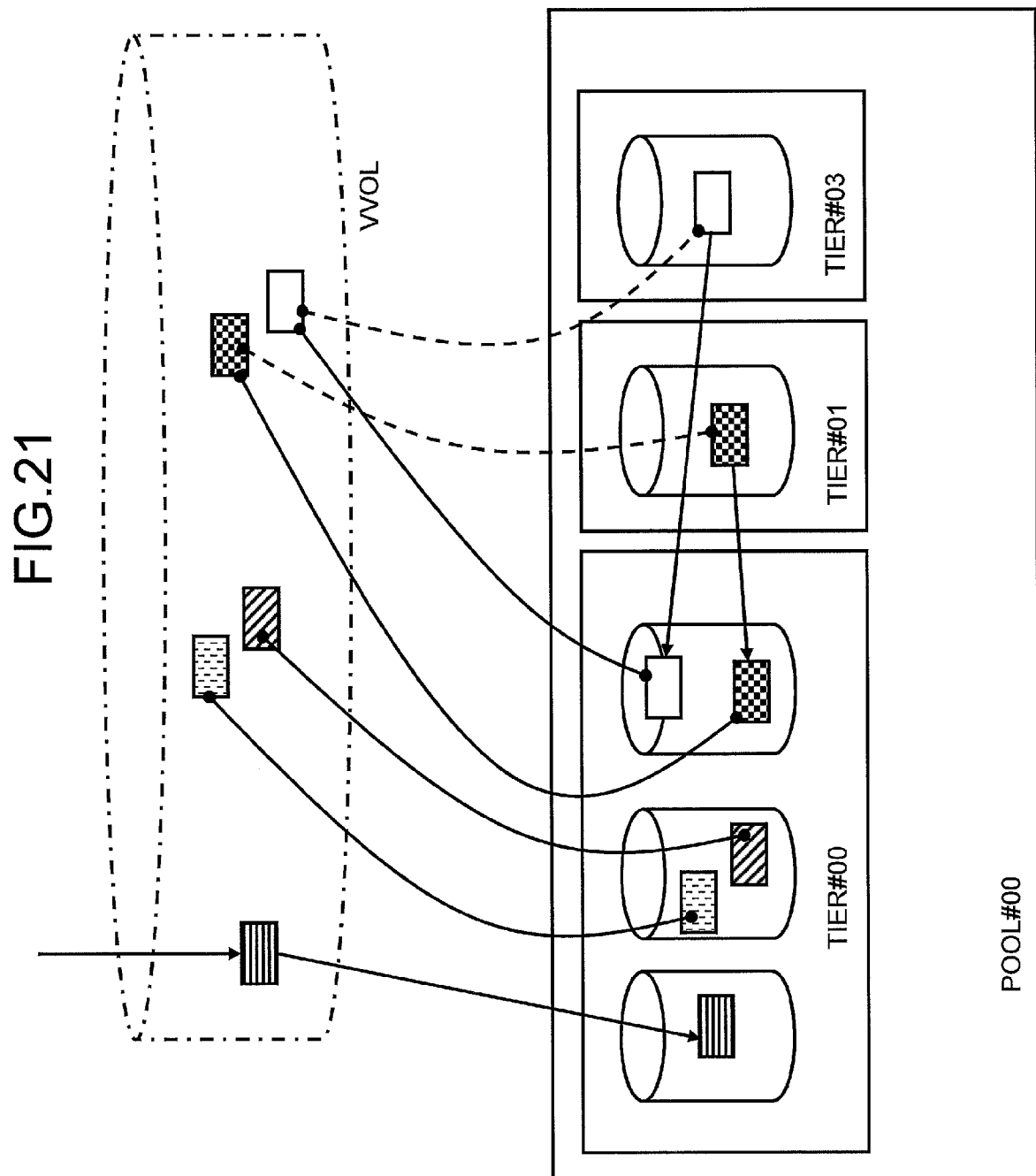
[図20]

FIG.20

4404

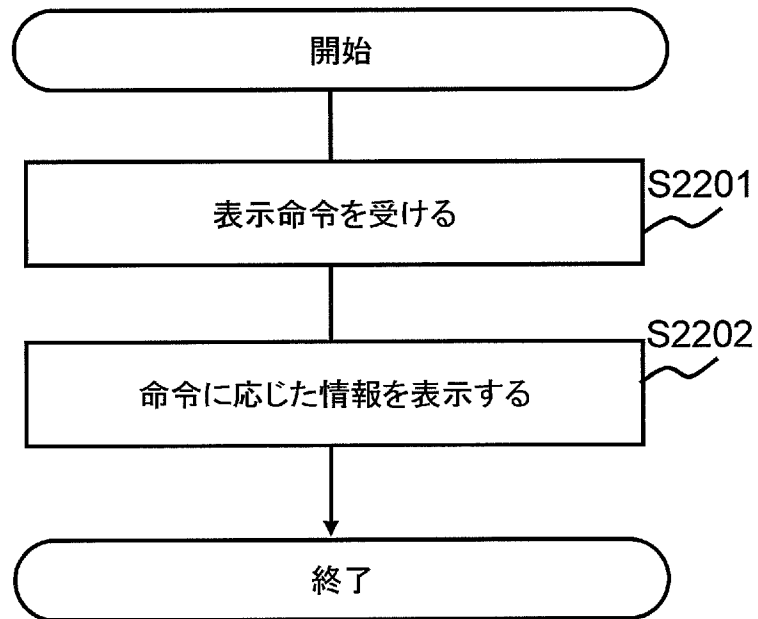
VVOL ID	...	比較方法	比較対象
00	...	絶対値	平均I/O数
01	...	相対値	最大I/O数
...	...	...	...

[図21]



[図22]

FIG.22



[図23]

FIG.23

2300

度数分布表示

WVOL ID 00

度数分布

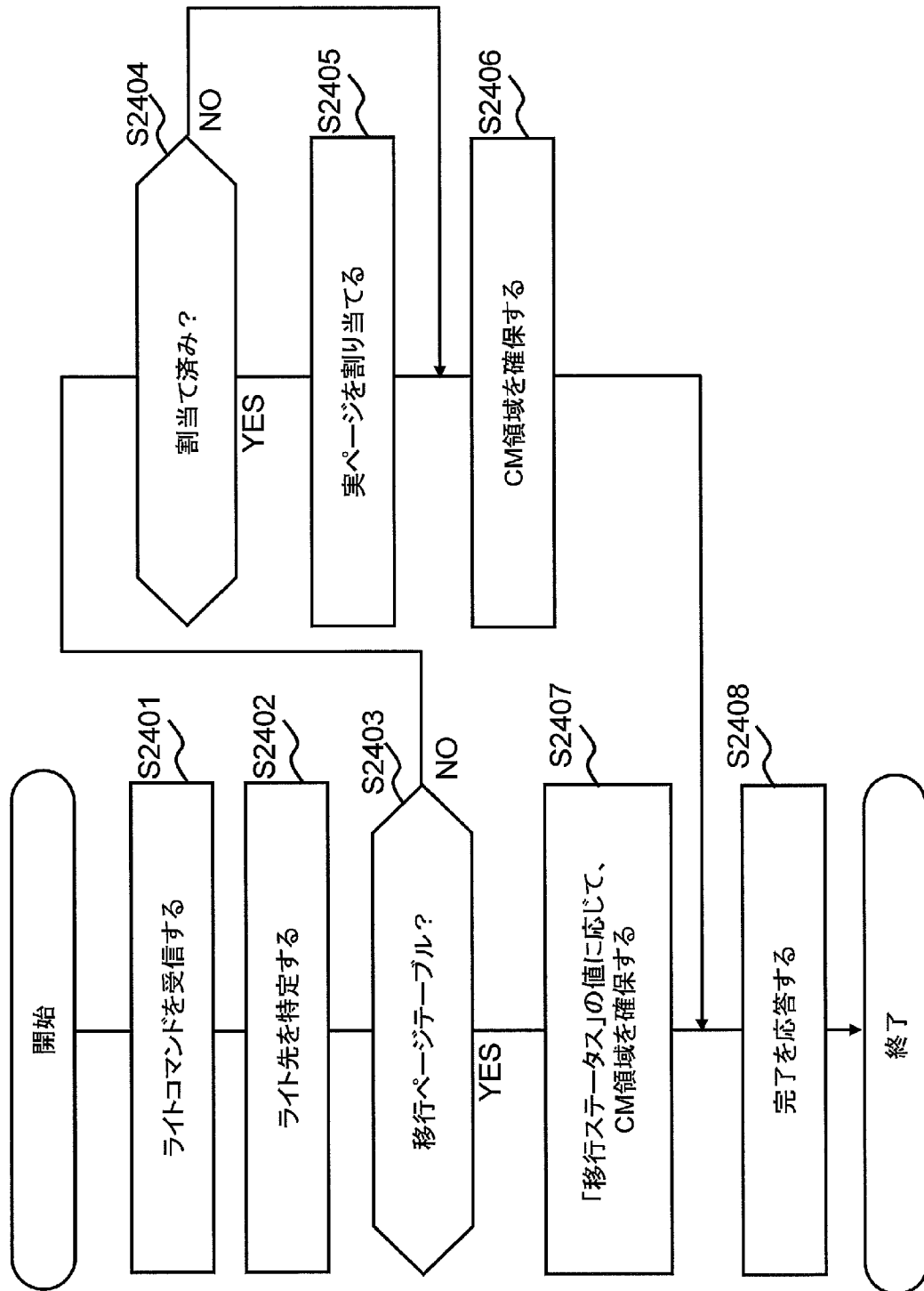
平均I/O数	ページ数
0	561
1	27
2	141
3	57
4-7	72
8-15	236
16-31	0

移行実行

中断

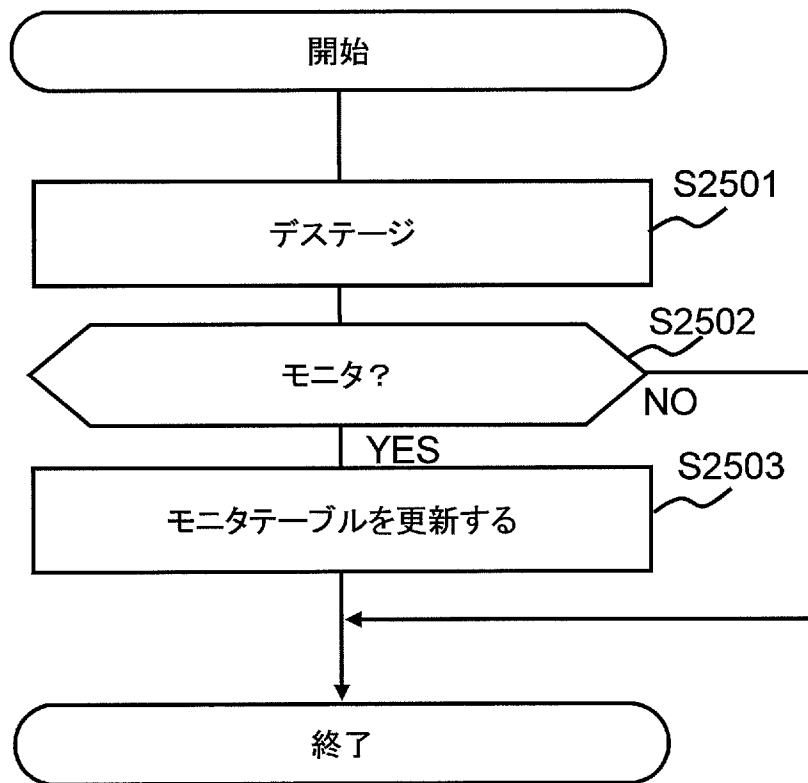
[図24]

FIG.24

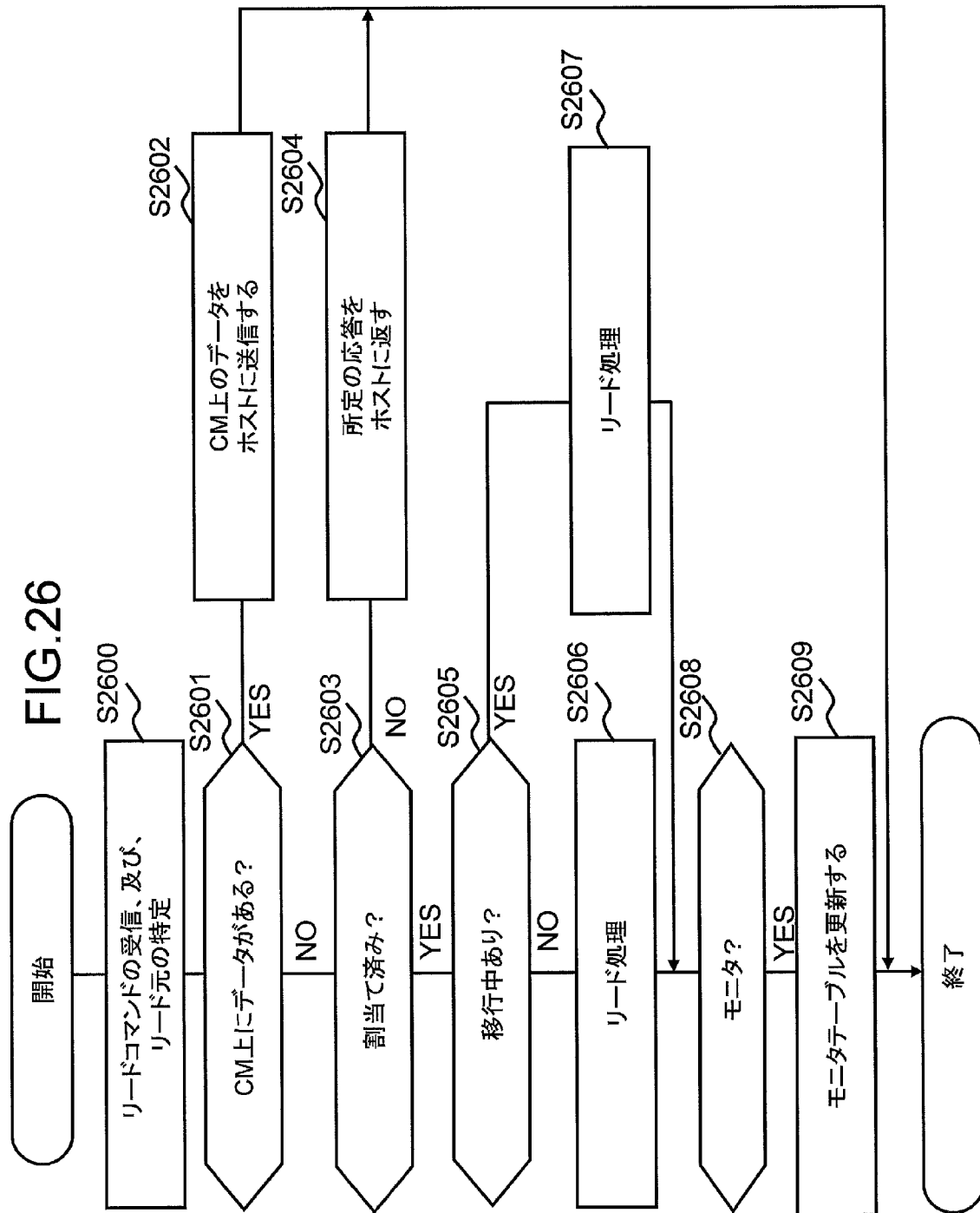


[図25]

FIG.25

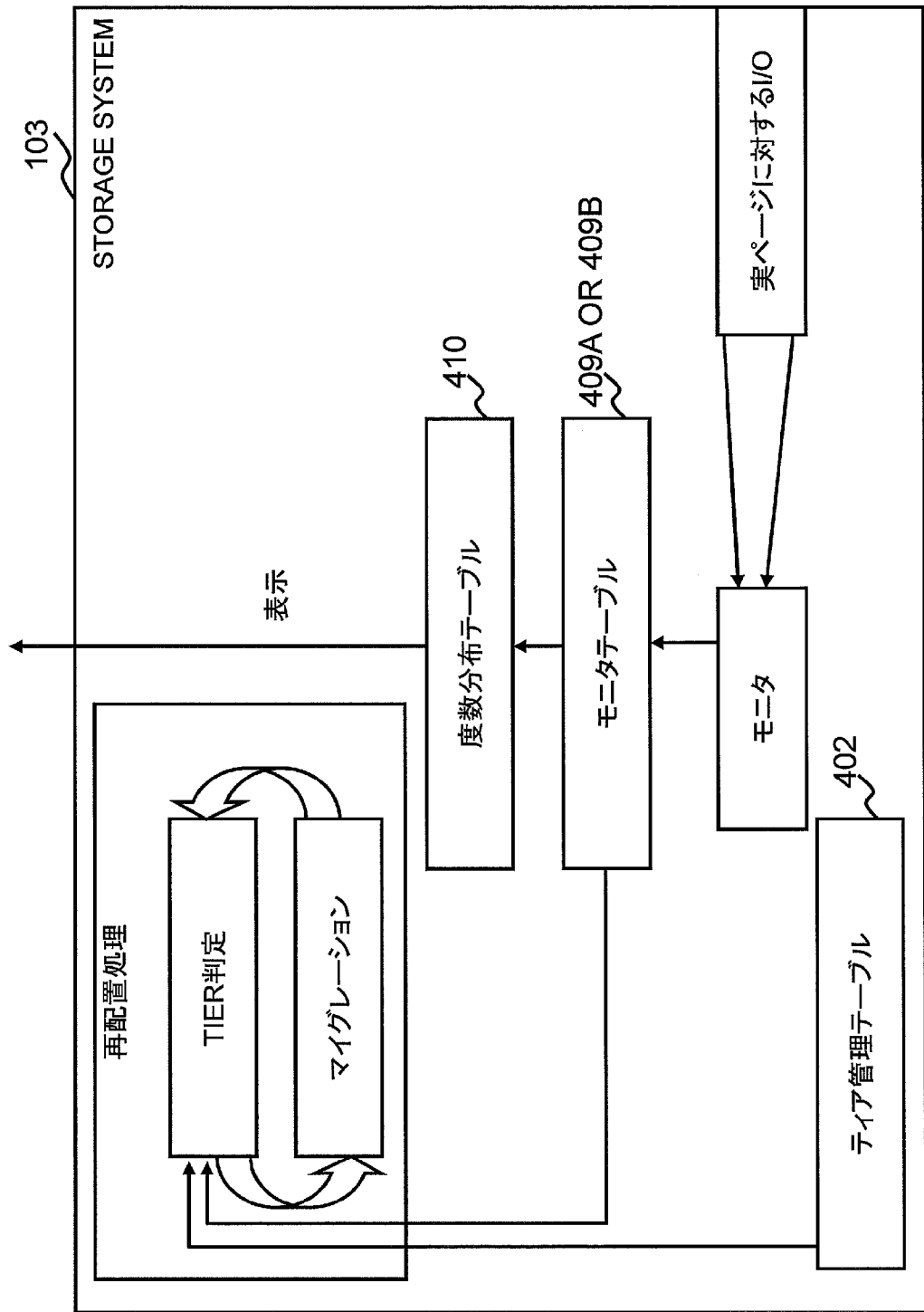


[図26]



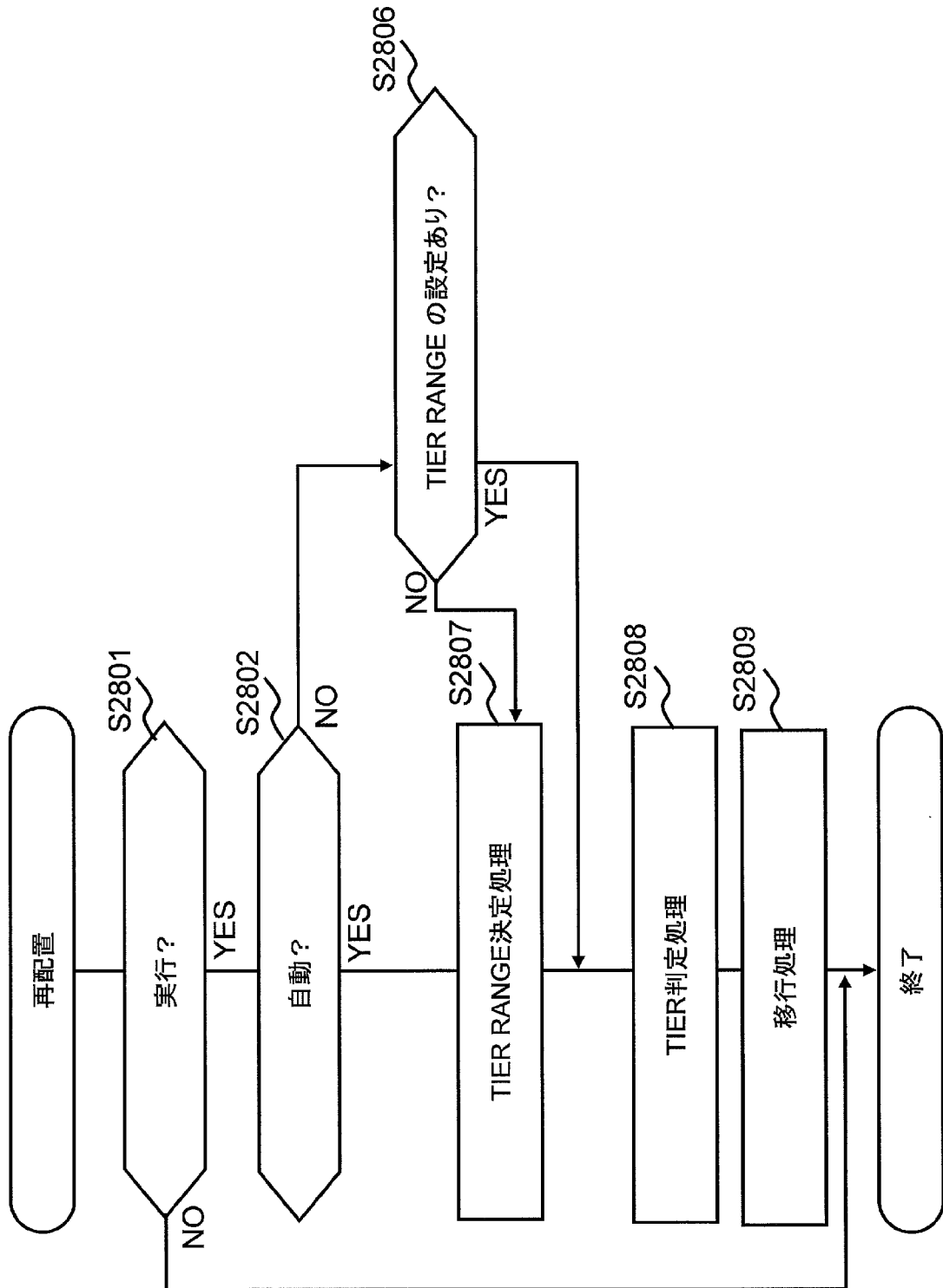
[図27]

FIG.27

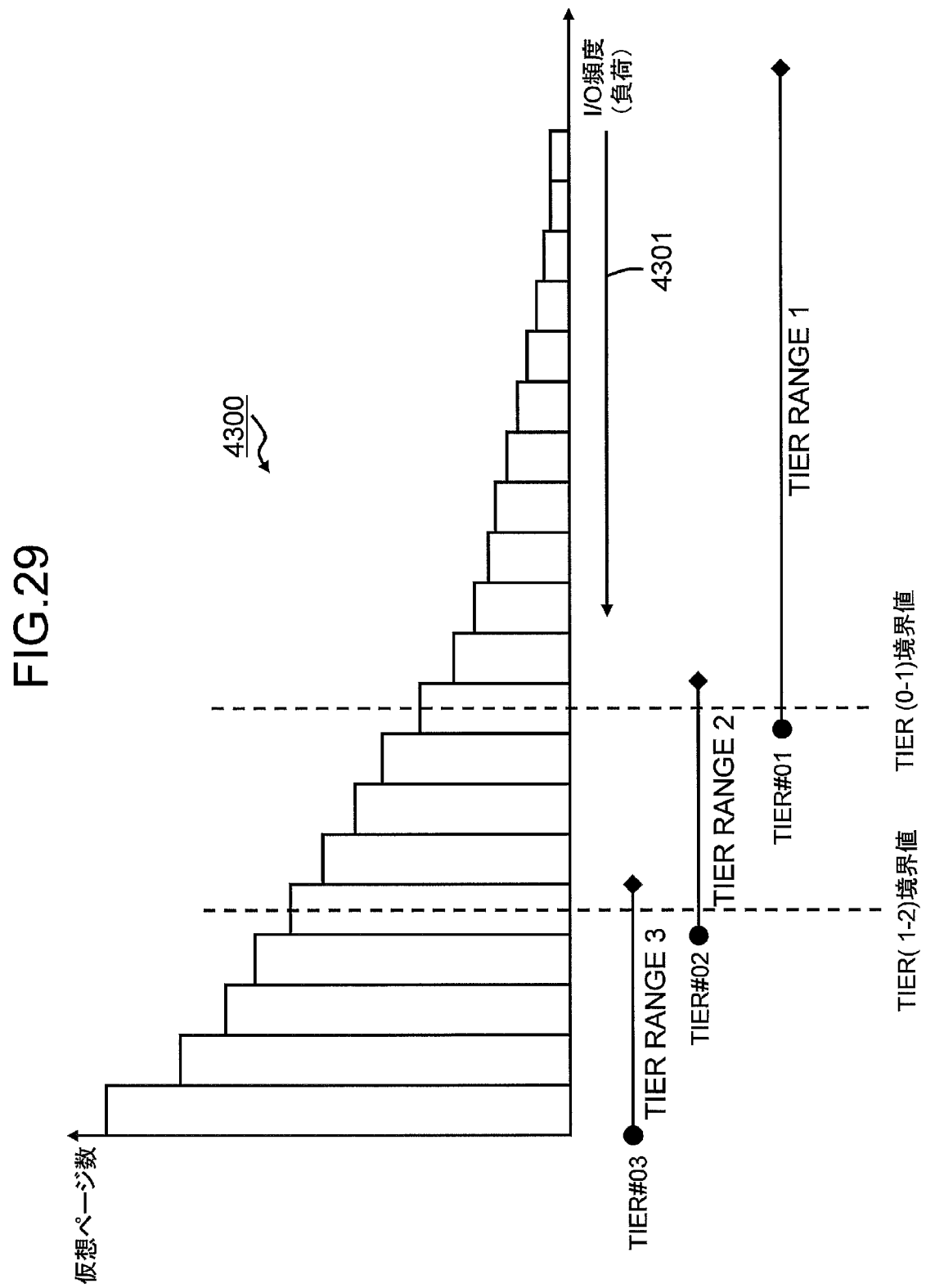


[図28]

FIG. 28

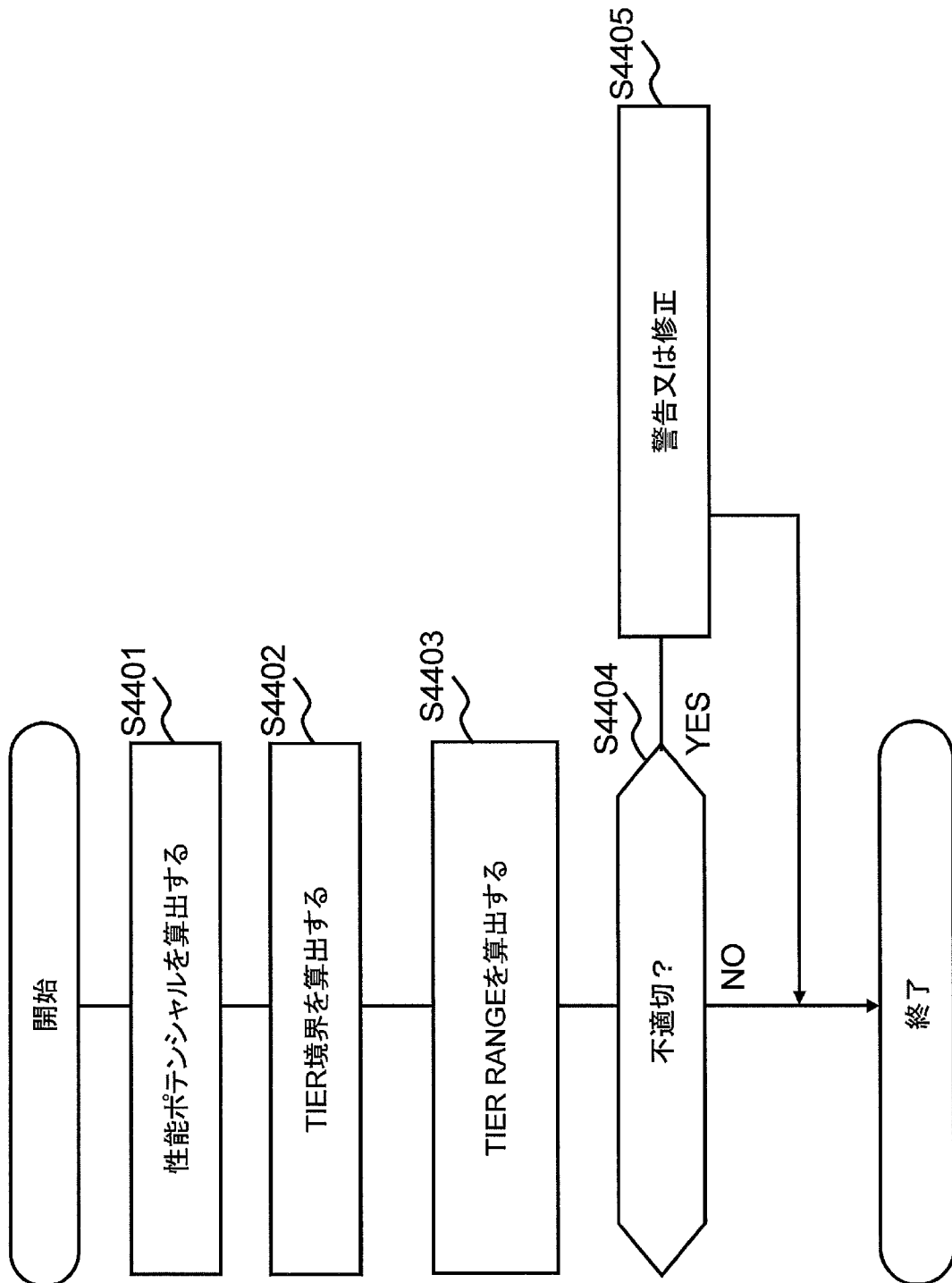


[図29]



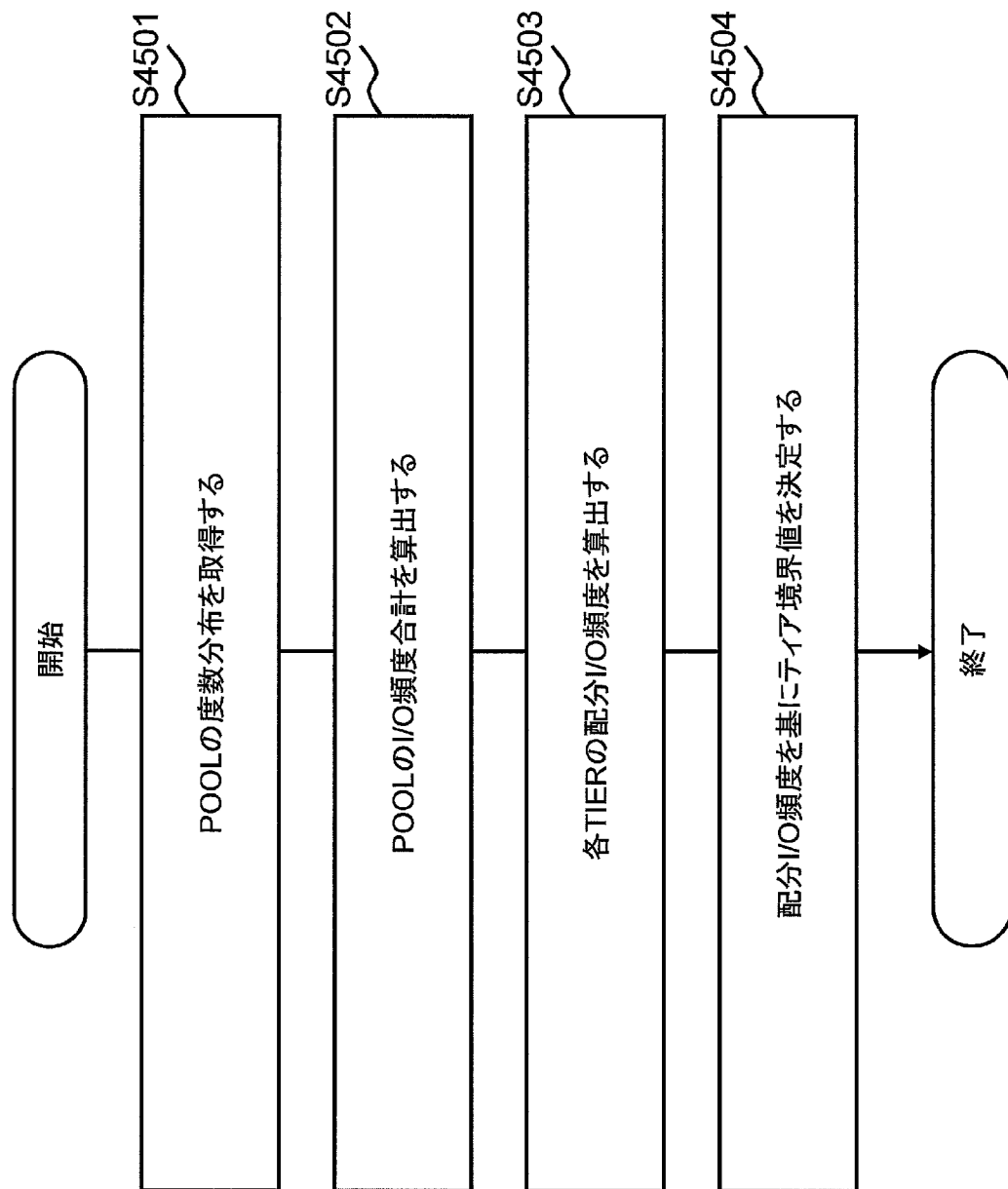
[図30]

FIG.30



[図31]

FIG.31



[図32]

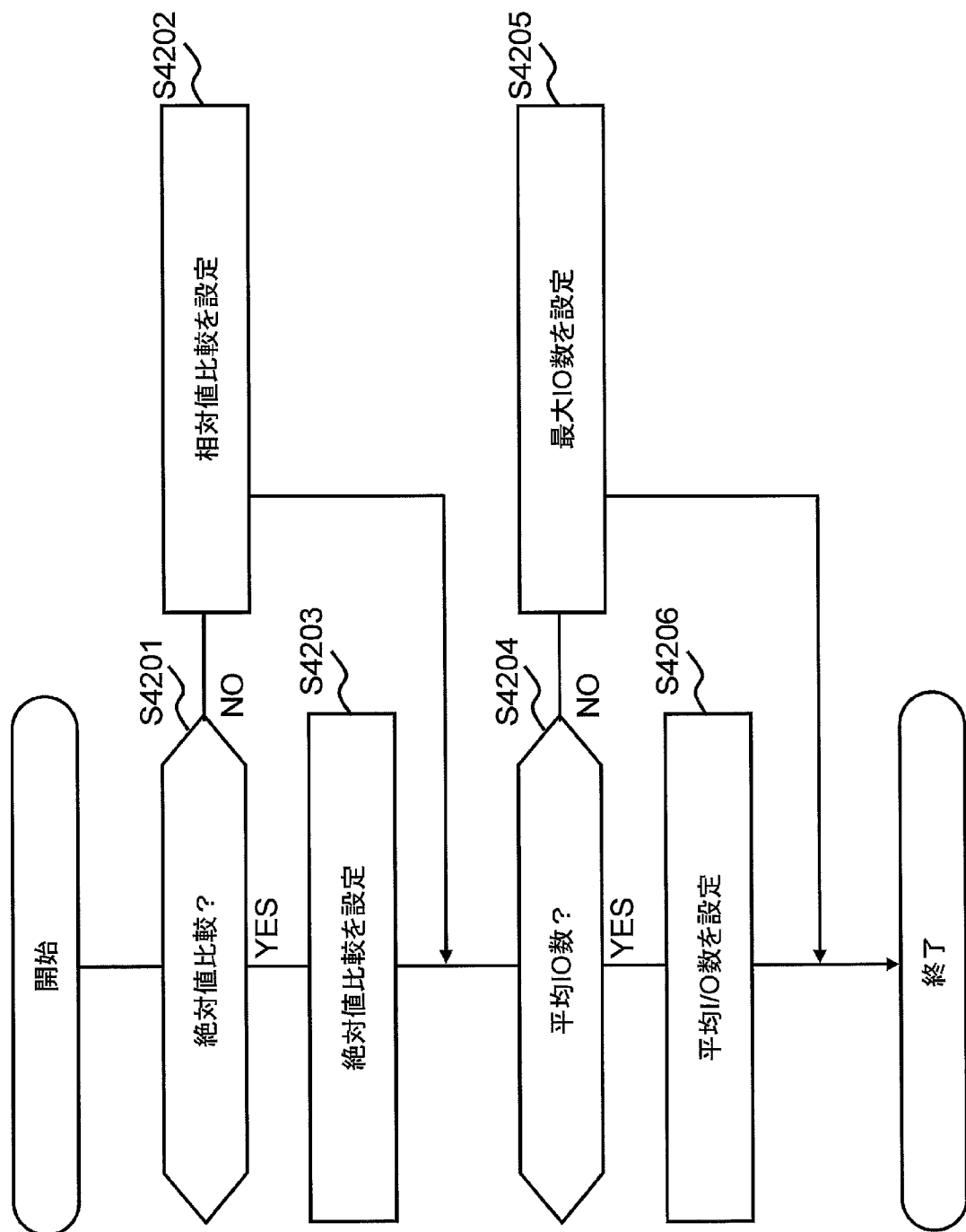
FIG.32

4403

POOL ID	...	比較方法	比較対象
00	...	絶対値	平均I/O数
01	...	相対値	最大I/O数
...	...	...	...

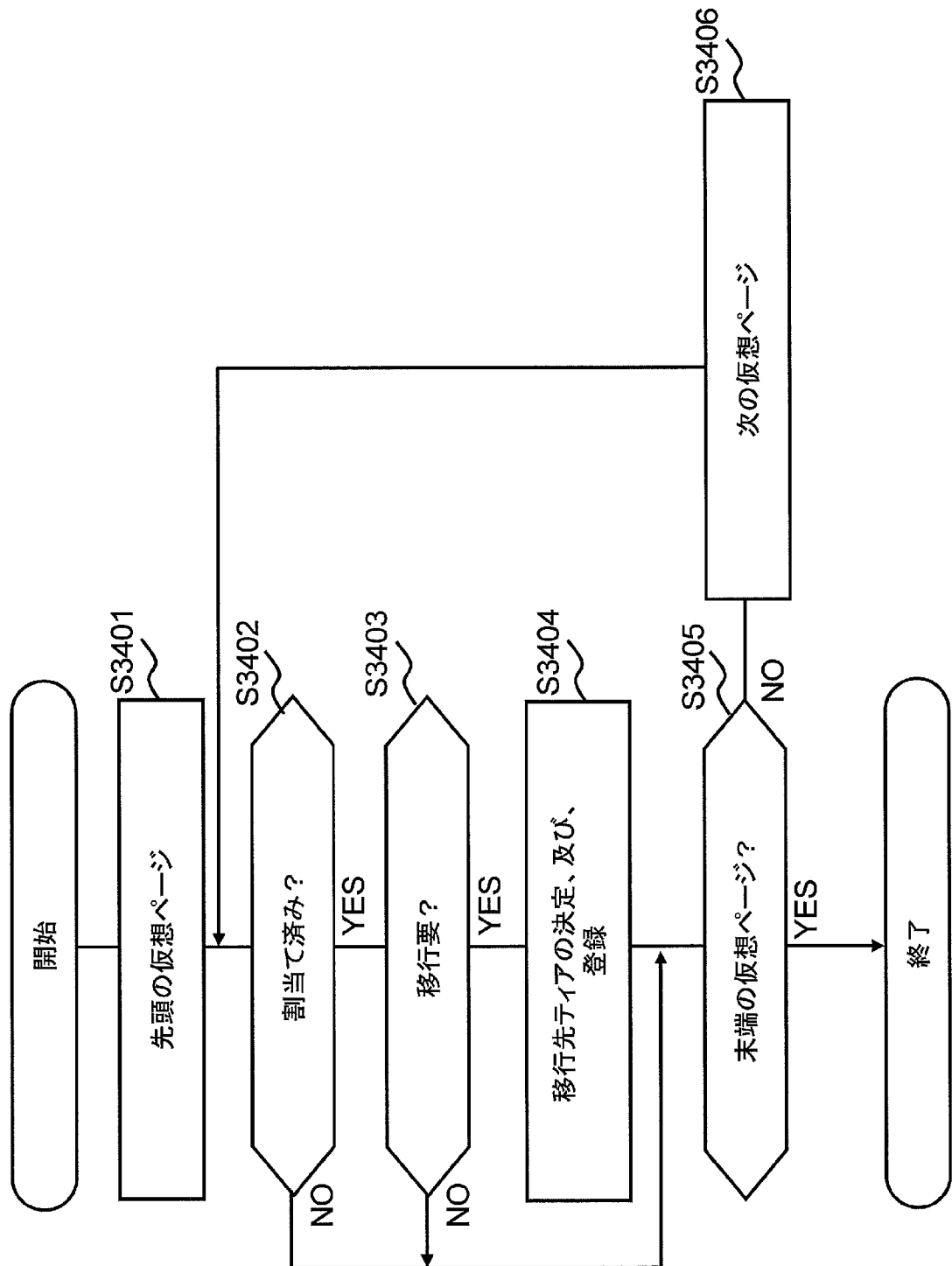
[図33]

FIG. 33



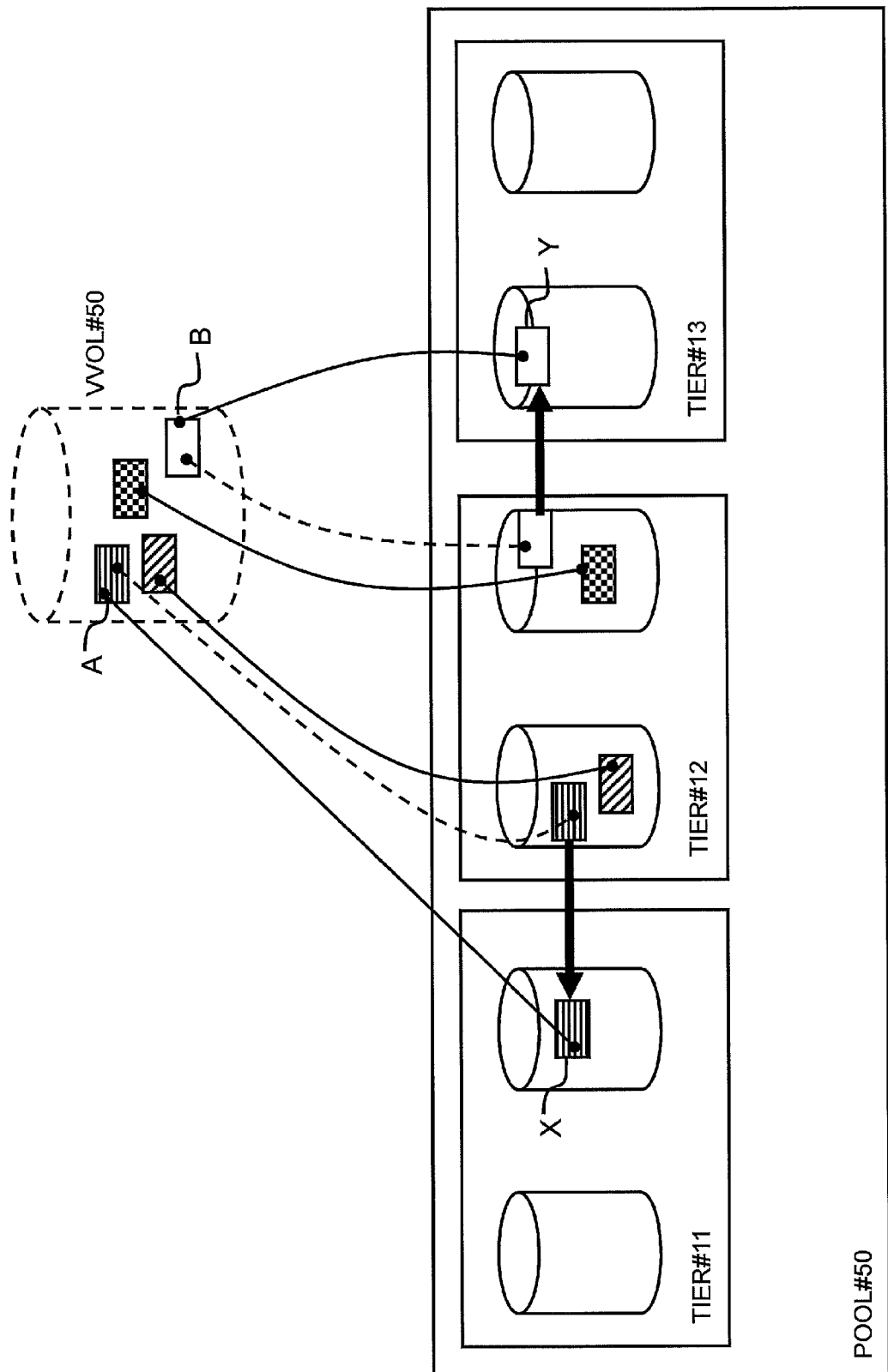
[図34]

FIG.34



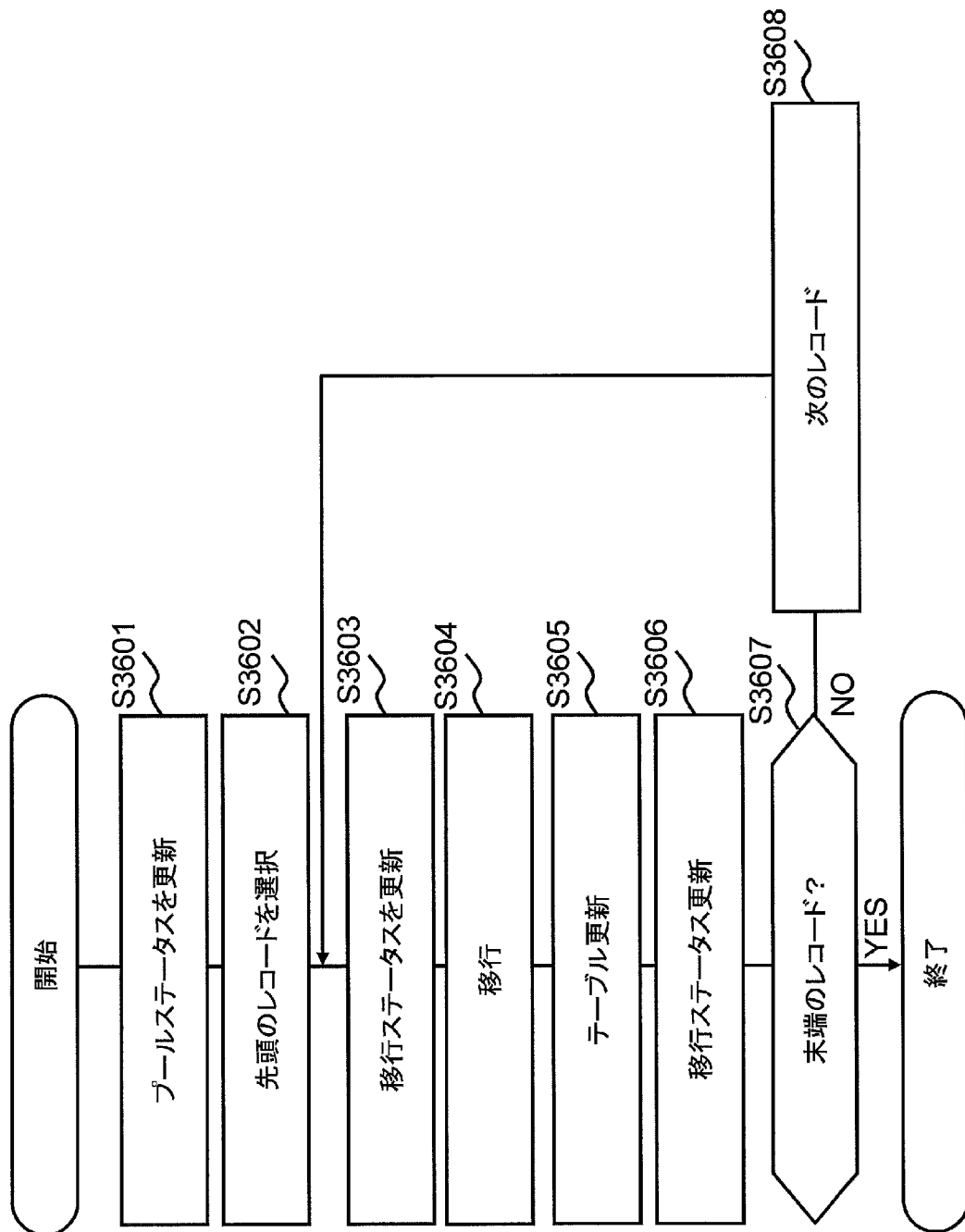
[図35]

FIG.35



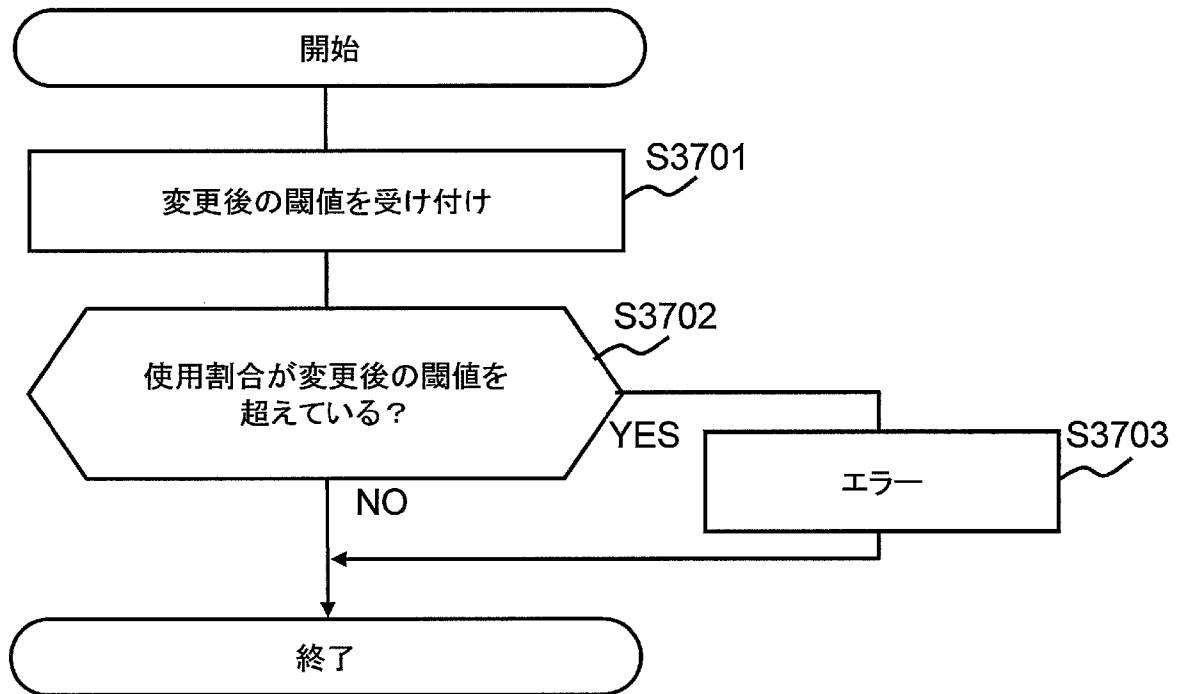
[図36]

FIG.36



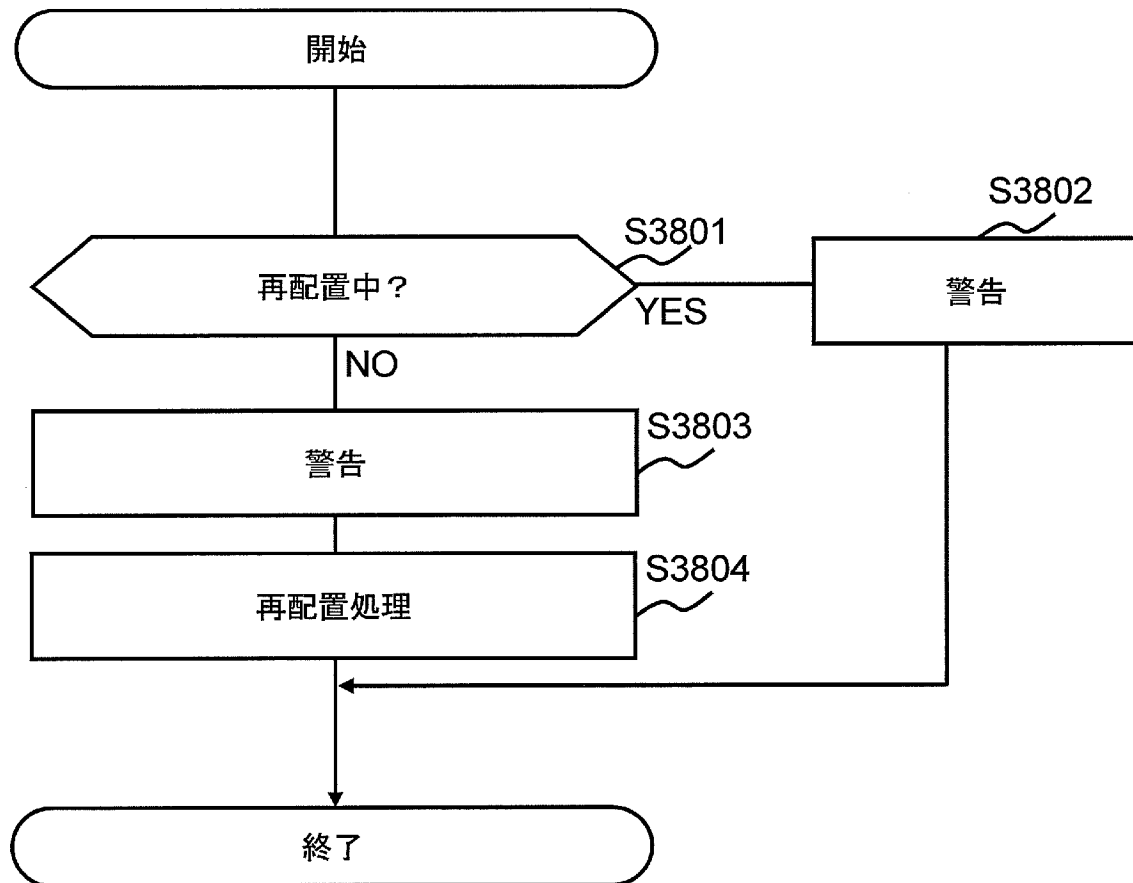
[図37]

FIG.37



[図38]

FIG.38



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/007207

A. CLASSIFICATION OF SUBJECT MATTER

G06F3/06(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F3/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2010

Kokai Jitsuyo Shinan Koho 1971-2010 Toroku Jitsuyo Shinan Koho 1994-2010

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP 2009-43055 A (Hitachi, Ltd.), 26 February 2009 (26.02.2009), paragraphs [0339] to [0443]; fig. 22, 23 & US 2009/0043942 A1	1 2-15
Y	JP 2007-234006 A (International Business Machines Corp.), 13 September 2007 (13.09.2007), paragraph [0059] & US 2007/0226631 A1 & CN 101030900 A	2-15
Y	JP 2001-519563 A (Seagate Technology L.L.C.), 23 October 2001 (23.10.2001), paragraphs [0146] to [0166]; fig. 34 to 36 & US 2002/0059539 A1 & GB 2345366 A & WO 1999/018507 A1 & DE 19882723 T & CN 1281560 A	3-15

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
11 March, 2010 (11.03.10)Date of mailing of the international search report
23 March, 2010 (23.03.10)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/007207

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2007-200333 A (Seagate Technology L.L.C.), 09 August 2007 (09.08.2007), claims 10, 11; paragraph [0033] & US 2007/0185902 A1	1-15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/007207

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The invention in claim 1 is described in the document 1 (JP 2009-43055 A (Hitachi, Ltd.), 26 February 2009 (26.02.2009), paragraphs 0339 - 0443, fig. 22, fig. 23, & US 2009/0043942 A1). Consequently, there is no "special technical feature" common to the invention in claim 1 and the inventions in claims 2 - 15.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

A. 発明の属する分野の分類 (国際特許分類 (I P C))

Int.Cl. G06F3/06(2006.01) i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (I P C))

Int.Cl. G06F3/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2010年
日本国実用新案登録公報	1996-2010年
日本国登録実用新案公報	1994-2010年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X Y	JP 2009-43055 A (株式会社日立製作所) 2009.02.26, 段落 0339-0443、図 22、図 23 & US 2009/0043942 A1	1 2-15
Y	JP 2007-234006 A (インターナショナル・ビジネス・マシーンズ・ コーポレーション) 2007.09.13、段落 0059 & US 2007/0226631 A1 & CN 101030900 A	2-15
Y	JP 2001-519563 A (シーゲイト テクノロジー エルエルシー) 2001.10.23、段落 0146-0166、図 34-36 & US 2002/0059539 A1 & GB	3-15

☒ C 欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」 口頭による開示、使用、展示等に言及する文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

11.03.2010

国際調査報告の発送日

23.03.2010

国際調査機関の名称及びあて先

日本国特許庁 (I S A/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

菅原 浩二

電話番号 03-3581-1101 内線 3565

5 I

9460

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	2345366 A & WO 1999/018507 A1 & DE 19882723 T & CN 1281560 A JP 2007-200333 A (シーゲイト テクノロジー エルエルシー) 2007.08.09, 請求項 10, 11, 段落 0033 & US 2007/0185902 A1	1-15

第Ⅱ欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求項 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。
つまり、
2. ☐ 請求項 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求項 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第Ⅲ欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるところの国際調査機関は認めた。

請求項1に係る発明は文献1（JP 2009-43055 A（株式会社日立製作所）2009.02.26, 段落0339-0443、図22, 図23 & US 2009/0043942 A1）に記載されている。してみると、請求項1に係る発明と請求項2～15に共通する「特別な技術的特徴」はない。

1. ☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求項について作成した。
2. ☒ 追加調査手数料を要求するまでもなく、すべての調査可能な請求項について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったため、この国際調査報告は、手数料の納付のあった次の請求項のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったため、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求項について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料及び、該当する場合には、異議申立手数料の納付と共に、出願人から異議申立てがあった。
- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあったが、異議申立手数料が納付命令書に示した期間内に支払われなかった。
- ☐ 追加調査手数料の納付はあったが、異議申立てはなかった。