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

特許第5439581号
(P5439581)

(45)発行日　平成26年3月12日(2014.3.12)　(24)登録日　平成25年12月20日(2013.12.20)

(51)Int.Cl.　F I
G06F 3/06 (2006.01)　G06F 3/06 301Z
G06F 3/06 305C
G06F 3/06 540

請求項の数 13　(全 23 頁)

(21)出願番号　特願2012-501484(P2012-501484)
(86)(22)出願日　平成21年10月15日(2009.10.15)
(65)公表番号　特表2012-533099(P2012-533099A)
(43)公表日　平成24年12月20日(2012.12.20)
(86)国際出願番号　PCT/JP2009/005383
(87)国際公開番号　WO2011/045838
(87)国際公開日　平成23年4月21日(2011.4.21)
審査請求日　平成24年1月11日(2012.1.11)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　100093861
弁理士　大賀　眞司
(74)代理人　100129218
弁理士　百本　宏之
(72)発明者　水野　陽一
神奈川県横浜市戸塚区吉田町２９２番地
株式会社日立製作所　横浜研究所内
(72)発明者　大平　良徳
神奈川県横浜市戸塚区吉田町２９２番地
株式会社日立製作所　横浜研究所内

審査官　木村　貴俊

最終頁に続く

(54)【発明の名称】ストレージシステム、ストレージ装置、ストレージシステムの記憶領域の最適化方法

(57)【特許請求の範囲】
【請求項１】
計算機からのライトアクセスを処理するストレージシステムであって、
前記ライトアクセスに伴うライトデータを記憶する記憶領域を備える記憶装置と、
前記記憶装置の前記記憶領域を、１以上のプール、前記１以上のプールのそれぞれに含まれる１以上のRAIDグループ、前記１以上のRAIDグループのそれぞれに含まれる複数の第１の論理領域、そして、前記複数の第１の論理領域のそれぞれを構成する複数の第２の論理領域、として管理し、前記計算機に複数の仮想ボリュームを提供し、当該計算機から前記ライトアクセスを受信すると、前記１以上のプールの何れかの前記１以上のRAIDグループに含まれる前記複数の第１の論理領域のうちの１つを、前記ライトアクセスの宛先仮想ボリュームに対応付け、前記宛先仮想ボリュームのデータがライトされる領域に、前記対応付けられた第１の論理領域に属する前記複数の第２の領域のうちの１つを割当てるコントローラと、
前記計算機を接続する第１のインターフェースと、
前記記憶装置を接続する第２のインターフェースと、
前記複数の第１の論理領域を管理するための第１の制御情報と、前記複数の第２の論理領域を管理するための第２の制御情報と、前記複数の第１の論理領域と前記複数の仮想ボリュームとの間の対応関係を定める第３の制御情報と、を格納する制御メモリと、を備え、
前記コントローラは、

前記ライトアクセスを受信すると、前記第３の制御情報により前記宛先仮想ボリュームに対応する前記第１の論理領域を判定し、当該第１の論理領域に属する前記複数の第２の論理領域のうちの１つを、前記宛先仮想ボリュームのデータがライトされるべき領域に割当てる第１処理部と、

前記複数の仮想ボリュームのうちの何れかに対応付けられた第１の論理領域を前記対応関係にある仮想ボリュームから解放する第２処理部と、

当該解放の際、前記解放される第１の論理領域に属する１以上の第２の論理領域に格納されているデータを、他の第１の論理領域に属する１以上の他の第２の論理領域に移行し、前記１以上の第２の論理領域を前記対応関係にある仮想ボリュームから解放する第３処理部と、を備える、

ストレージシステム。

【請求項２】

前記第３の制御情報は、前記複数の第１の論理領域のそれぞれを一つの同じ仮想ボリュームに対応させるものであり、

前記制御メモリは、前記複数のRAIDグループから前記プールを形成するための第４の制御情報を格納し、

前記第１の論理領域に属する前記第２の論理領域は、前記第１の論理領域が対応する仮想ボリュームに割当てられ、

前記コントローラは、前記一つ又は複数の仮想ボリュームへの前記第２の論理領域の割当て数を算出する第４処理部と、前記第２の論理領域の割当て数と前記RAIDグループ数に基づいて、RAIDグループ毎に必要な前記第１の論理領域の必要数を算出する第５処理部と、前記RAIDグループに属する、前記仮想ボリュームに割当てられている第１の論理領域の数と前記必要な第１の論理領域の数とを対比し、前者が後者より不足している場合に、同じRAIDグループ又は他のRAIDグループから、前記仮想ボリュームに割当てられていない第１の論理領域を前記第１の論理領域数が不足しているRAIDグループに割当てる第６処理部と、を有し、前記第２処理部は、後者が前者より多い場合に余分な第１の論理領域の前記仮想ボリュームへの割当てを解放し、

さらに、前記コントローラの前記第２処理部は前記第１の論理領域の解放を、当該第１の論理領域の重要度を判定し、その判定の結果、重要度が低いものから実行する、請求項１記載のストレージシステム。

【請求項３】

前記第３の制御情報は、前記複数の第１の論理領域のそれぞれを一つの同じ仮想ボリュームに対応させるものである、請求項１記載のストレージシステム。

【請求項４】

前記１以上のRAIDグループのそれぞれに含まれる前記複数の第１の論理領域のそれぞれが対応する仮想ボリュームが異なる場合がある、請求項３記載のストレージシステム。

【請求項５】

前記第１の論理領域に属する前記第２の論理領域は、前記第１の論理領域が対応する仮想ボリュームに割当てられる、請求項１記載のストレージシステム。

【請求項６】

前記コントローラの前記第２処理部は、前記一つ又は複数の仮想ボリュームへ前記第２の論理領域を複数割当てるのに足る前記第１の論理領域の必要数に基づいて、前記RAIDグループに属する前記第１の論理領域を前記対応関係にある仮想ボリュームへの割当てから解放する、請求項１記載のストレージシステム。

【請求項７】

前記コントローラは、前記一つ又は複数の仮想ボリュームへの前記第２の論理領域の割当て数を算出する第４処理部と、前記第２の論理領域の割当て数と前記RAIDグループ数に基づいて、RAIDグループ毎に必要な前記第１の論理領域の必要数を算出する第５処理部と、前記RAIDグループに属する、前記仮想ボリュームに割当てられている第１の論理領域の数と前記必要な第１の論理領域の数とを対比し、前者が後者より不足している場合に、同

じRAIDグループ又は他のRAIDグループから、前記仮想ボリュームに割当てられていない第１の論理領域を前記第１の論理領域数が不足しているRAIDグループに割当てる第６処理部と、を有し、前記第２処理部は、後者が前者より多い場合に余分な第１の論理領域の前記仮想ボリュームへの割当てを解放する、請求項１記載のストレージシステム。

【請求項８】

前記コントローラの前記第２処理部は前記第１の論理領域の解放を、当該第１の論理領域の重要度を判定し、その判定の結果、重要度が低いものから実行する、請求項１記載のストレージシステム。

【請求項９】

前記仮想ボリュームへの割当てられている前記第２の論理領域の数が多い程、又は、前記第１の論理領域に属する前記複数の第２の論理領域のそれぞれの優先度に基いて決定される、前記第１の論理領域の優先度が高い程、前記コントローラの前記第２処理部は、前記第１の論理領域の重要度を高いと判断する、請求項８記載のストレージシステム。

【請求項１０】

前記５処理部は、前記第２の論理領域の割当て数を前記RAIDグループ数毎にした算出値を求め、この算出値以下の最大の整数値を前記第１の論理領域の必要数とし、前記第６処理部は、前記複数のRAIDグループについて前記第１の論理領域の必要数を合計したものが、前記一つ又は複数の仮想ボリュームへの前記第２の論理領域の割当て数を下回る数分の第１の論理領域を前記複数のRAIDグループの少なくとも一つのRAIDグループに設定する、請求項７記載のストレージシステム。

【請求項１１】

前記５処理部は、前記第２の論理領域の割当て数を前記RAIDグループ数毎にした算出値を求め、この算出値以上の最小の整数値を前記第１の論理領域の必要数とする、請求項７記載のストレージシステム。

【請求項１２】

計算機と記憶装置とに接続され、前記記憶装置の記憶領域を用いて前記計算機からのライトアクセスを処理するストレージ装置であって、

前記記憶装置の前記記憶領域を、１以上のプール、前記1以上のプールのそれぞれに含まれる１以上のRAIDグループ、前記１以上のRAIDグループのそれぞれに含まれる複数の第１の論理領域、前記複数の第１の論理領域のそれぞれを構成する複数の第２の論理領域、として管理し、前記計算機に複数の仮想ボリュームを提供するコントローラと、

前記計算機を接続する第１のインターフェース、

前記記憶装置を接続する第２のインターフェース、

前記複数の第１の論理領域を管理するための第１の制御情報と、前記第１の論理領域に複数の第２の論理領域を管理するための第２の制御情報と、前記複数の第１の論理領域と前記複数の仮想ボリュームとの間の対応関係を定める第３の制御情報と、を格納する制御メモリ、

前記コントローラは、

前記計算機から前記ライトアクセスを受信すると、前記第３の制御情報により前記宛先仮想ボリュームに対応する前記第１の論理領域を判定し、当該第１の論理領域に属する前記複数の第２の論理領域のうちの１つを、前記宛先仮想ボリュームのデータがライトされるべき領域に割当てる第１処理部と、

前記複数の仮想ボリュームのうちの何れかに対応付けられた第１の論理領域を前記対応関係にある仮想ボリュームから解放する第２処理部と、

当該解放の際、前記解放される第１の論理領域に属する１以上の第２の論理領域に格納されているデータを、他の第１の論理領域に属する１以上の他の第２の論理領域に移行し、前記１以上の第２の論理領域を前記対応関係にある仮想ボリュームから解放する第３処理部と、を備えるストレージ装置。

【請求項１３】

計算機からのライトアクセスを処理するストレージシステムの記憶領域の最適化方法で

あって、
前記ストレージシステムは、前記計算機に複数の仮想ボリュームを提供し、当該計算機から前記ライトアクセスを受信すると、前記ライトアクセスに伴うライトデータを、記憶装置の何れかの記憶領域に格納させるコントローラを有し、
当該コントローラは、
前記記憶装置の前記記憶領域を、1以上のプール、前記1以上のプールのそれぞれに含まれる1以上のRAIDグループ、前記1以上のRAIDグループのそれぞれに含まれる複数の第1の論理領域、前記複数の第1の論理領域のそれぞれを構成する複数の第2の論理領域、として管理し、
前記ライトアクセスを受信すると、前記ライトアクセスの宛先仮想ボリュームに対応する前記第１の論理領域を判定し、当該第１の論理領域に属する前記複数の第２の論理領域のうちの１つを前記宛先仮想ボリュームのデータがライトされるべき領域に割当て、
前記複数の仮想ボリュームのうちの何れかに対応付けられた第１の論理領域を前記何れかの仮想ボリュームから解放し、
前記第１の論理領域を解放する際に、前記第１の論理領域に属する割当て済みの１以上の前記第２の論理領域に格納されているデータを、他の第１の論理領域に属する１以上の他の第２の論理領域に移行し、前記１以上の第２の論理領域を前記何れかの仮想ボリュームから解放する、
各ステップを実行する、
ストレージシステムの記憶領域の最適化方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明はストレージシステムに係わり、特に、ホスト計算機に対して動的に記憶容量を割当てるストレージシステムに関するものである。本発明は、さらに、ストレージシステムのストレージ装置、及び、ストレージシステムの記憶領域の最適化方法に関するものである。
【背景技術】
【０００２】
従来から、ホスト計算機に対して大規模なデータストレージサービスを提供する計算機システムが存在する。この計算機システムは、ホスト計算機と、ホスト計算機が接続するストレージ装置と、ストレージ装置の管理装置と、を備えたストレージシステム或いはコンピュータシステムとして知られている。
【０００３】
ストレージ装置は、複数のハードディスクをＲＡＩＤ（Redundant Array of Independent/Inexpensive Disks）方式で管理する。そして、ストレージ装置は、複数のハードディスクが有する物理的な記憶領域をＲＡＩＤグループとして論理化し、これを論理ボリュームとしてホスト計算機に提供する。ホスト計算機は論理ボリュームにアクセスして、データのリード・ライトをストレージ装置に要求する。
【０００４】
ストレージ装置は、複数のページを論理ボリュームに割当てることによって、ホスト計算機から論理ボリュームに対してライトアクセスされたデータを、記憶装置の記憶領域に保持する。ページとは、記憶装置の物理的な記憶領域が分割されることによって、一定の記憶容量を持たせた論理的な単位である。
【０００５】
この種の論理化技術の一つとして、論理ボリュームの記憶容量を仮想化したシステムが、シン・プロビジョニングとして知られている(例えば、米国特許第６８２３４４２号公報)。
【０００６】
このシステムは、論理ボリュームを仮想化した記憶容量を持つ仮想ボリュームとして設

定するものであって、仮想ボリュームのブロックアドレスに対するライト要求をホスト計算機から受信する都度、ライト要求があったブロックアドレスにページを割当てて、仮想ボリュームの実記憶容量を動的に変化させるものである。

【０００７】

仮想ボリュームに記憶領域を割当てるための、実記憶領域からなるグループはプールと呼ばれ、このプール内にプールボリュームと呼ばれるＲＡＩＤグループの論理ボリュームが複数存在する。ページはプールボリュームから仮想ボリュームに対して割当てられる。

【０００８】

ストレージ装置は、複数のプールボリュームから均等にページを選択しようとするが、長い時間を経過してゆく過程で、この均等が崩れホスト計算機からのＩ／Ｏに対する応答性を低下させるという問題があった。そこで、特開２００８－２３４１５８号公報は、複数のプールボリューム間でページを移動させることを開示している。

【先行技術文献】

【特許文献】

【０００９】

【特許文献１】米国特許第６８２３４４２号公報

【特許文献２】特開２００８－２３４１５８号公報

【発明の概要】

【発明が解決しようとする課題】

【００１０】

米国特許第６８２３４４２号公報は、チャンクレットという管理単位で複数のディスク装置の記憶領域を管理し、チャンクレットからページが仮想ボリュームに割当てることを開示している。

【００１１】

しかしながら、ストレージ装置が同一の管理単位に属する複数のページをそれぞれ異なる仮想ボリュームに割当てようとすると、管理単位が同一のＲＡＩＤグループに設定される以上、共通のパリティによって定義される複数のページについて、パリティ演算が異なるコントローラに跨ってしまい、パリティ演算のためのオーバヘッドが過大になるという問題がある。

【００１２】

一方、ストレージ装置が、一つの管理単位に属する複数のページを同一の仮想ボリュームに割当てると、ある管理単位のページを他の管理単位に移動させる際、割当て先となる仮想ボリュームが異なると、ページの移動が制限される。

【００１３】

この場合、ストレージ装置は、管理単位が有する記憶領域を有効に活用できないばかりか、ページの移動が出来ないことによって、複数のプールボリューム間でのページの割当てを均等に調整できず、ホスト計算機からのＩ／Ｏに対する応答性の低下を十分に解消できないという問題があった。

【００１４】

そこで、本発明は、仮想ボリュームに割当てられる記憶領域を、ＲＡＩＤグループに設定された管理領域によって管理しても、パリティ演算のためのオーバヘッドが過大にならないストレージシステム、ストレージ装置、及び、ストレージシステムの記憶領域の最適化方法を提供することを目的とするものである。

【００１５】

さらに、本発明の他の目的は、仮想ボリュームに割当てられる記憶領域を、ＲＡＩＤグループに設定された管理領域によって管理しても、管理領域の記憶領域を有効に活用することができる、ストレージシステム、ストレージ装置、及び、ストレージシステムの記憶領域の最適化方法を提供することにある。

【００１６】

本発明のさらに他の目的は、論理ボリュームに割当てられる記憶領域をＲＡＩＤグルー

プに設定された管理領域によって管理しても、計算機からのＩ／Ｏに対する応答性の低下を十分に解消できる、ストレージシステム、ストレージ装置、及び、ストレージシステムの記憶領域の最適化方法を提供することにある。

【課題を解決するための手段】

【００１７】

ストレージ装置が一つの管理単位に属する複数のページを同一の仮想ボリュームに割当てるようにすると、管理単位が仮想ボリュームへの割当てに利用されていないページを多く含んでいても、ストレージ装置は、このページを他の仮想ボリュームへの割当てに充当することができない。

【００１８】

そこで、本発明は、ページの割当てに十分活用されていない特定管理単位を仮想ボリュームへの割当てから解放し、この特定管理単位に属する割当て済みページを、他の管理単位に移動させることにより、特定管理単位の記憶領域を他の仮想ボリュームへのホスト計算機からのライトアクセスに供するようにしたものである。

【発明の効果】

【００１９】

以上説明したように、本発明によれば、仮想ボリュームに割当てられる記憶領域を、ＲＡＩＤグループに設定された管理領域によって管理しても、パリティ演算のためのオーバヘッドが過大にならないストレージシステム、ストレージ装置、及び、ストレージシステムの記憶領域の最適化方法を提供することができる。

【００２０】

さらに、本発明によれば、仮想ボリュームに割当てられる記憶領域を、ＲＡＩＤグループに設定された管理領域によって管理しても、管理領域の記憶領域を有効に活用することができる。さらにまた、計算機からのＩ／Ｏに対する応答性の低下を十分に解消できる。

【図面の簡単な説明】

【００２１】

【図１】図１は、本発明のストレージシステムの一実施形態に係るハードウエアブロック図である。

【図２】図２は、コントローラの詳細を示すハードウエアブロック図である。

【図３】図３は、ストレージ装置１が、ホスト計算機に提供する、複数の仮想ボリュームと記憶装置の記憶領域との関係を示す機能ブロック図である。

【図４】図４は、ページ管理テーブルの一例を示す。

【図５】図５は、RAIDグループ管理テーブルの一例である。

【図６】図６は、チャンク管理テーブルの一例である。

【図７】図７は、チャンクを仮想ボリュームに対して、解放したり、あるいは新たに割当てるなどの再配置を実行するための管理テーブルの一例である。

【図８】図８は、仮想ボリューム毎に解放すべきチャンクを管理するテーブルの一例である。

【図９】図９は、管理端末によって管理ユーザ・管理クライアントに提供される、設定画面の一例を示すものである。

【図１０】図１０は、ＬＵ設定処理を説明するフローチャートである。

【図１１】図１１は、管理端末が、ページの移動やチャンクの解放をストレージ装置に設定するための処理を説明するフローチャートである。

【図１２】図１２は、容量優先モードでの再配置処理を説明するフローチャートである。

【図１３】図１３は、ワイドストライプモードによる再配置処理を説明するフローチャートである。

【図１４】図１４は、ディスク占有モードの再配置処理を説明するフローチャートである。

【図１５Ａ】図１５Ａは、図1に示すストレージシステムにおけるページの移動の前後のプールの状態を示すブロック図である。

【図１５Ｂ】図１５Ｂは、図1に示すストレージシステムにおけるページの移動の前後のプールの状態を示すブロック図である。
【図１６Ａ】図１６Ａは、比較例におけるページの移動の前後のプールの状態を示すブロック図である。
【図１６Ｂ】図１６Ｂは、比較例におけるページの移動の前後のプールの状態を示すブロック図である。
【発明を実施するための形態】
【００２２】
図１は、本発明のストレージシステムの一実施形態に係るハードウエアブロック図である。ストレージシステムは、ディスクアレイを有するストレージ装置１、上位装置としてのホスト計算機２、及び、管理端末３と、を備える。
【００２３】
ストレージ装置１は、ホスト２と接続され、管理端末３とも接続されている。また、ホスト計算機２は、ネットワークを介して管理端末３と接続している。
【００２４】
ネットワークは、いかなる種類のものでもよい。例えば、ＬＡＮのようなＩＰネットワークである。
【００２５】
ホスト計算機２は、ユーザに種々のサービスを提供する計算機である。ホスト計算機２は、ネットワークインクフェース、ネットワークインターフェースに接続されるプロセッサ、及び、プロセッサに接続されるメモリを備える。
【００２６】
メモリは、アプリケーションプログラムを格納している。アプリケーションプログラムは、ホスト計算機２がユーザにサービスを提供するために、プロセッサによって実行されるプログラムである。
【００２７】
アプリケーションプログラムを実行するプロセッサは、必要に応じて、ストレージ装置１へのアクセス要求（データ書き込み要求、又はデータ読み出し要求）を発行する。
【００２８】
管理端末３は、計算機システムを管理する計算機である。管理端末３は、ネットワークインターフェース、ネットワークインターフェースに接続されるプロセッサ、及びプロセッサに接続されるメモリを備える。
【００２９】
管理端末３のメモリには、管理プログラム３１が格納されている。管理プログラム３１は、論理ボリュームの作成、ストレージ装置１の管理、及び、ストレージ装置１によるページ再配置処理を実行するためにプロセッサによって実行されるプログラムである。
【００３０】
ストレージ装置１は、ホスト計算機２によって書き込まれたデータを記憶装置に格納する。記憶装置は、複数のディスク装置１２からなるディスクアレイとして、計算機システムに存在する。符号１２ａ及１２ｂは、複数のディスク装置１２の一つである。ディスク装置１２は、磁気ディスクドライブ、フラッシュメモリまたはその他の記憶装置である。コントローラ１１は、ホスト計算機２からのライト命令及びリード命令を実行して、記憶装置に対するデータのリード・ライト処理を達成する。
【００３１】
コントローラ１１は、複数のハードディスクが有する物理的な記憶領域をＲＡＩＤグループとして論理化して、論理ボリュームをホスト計算機に提供する。さらに、コントローラ１１は、一つ又は複数のＲＡＩＤグループを、プールとして、仮想ボリュームに対して割当てられるべき記憶領域として管理する。一つ又は複数のプールがストレージ装置に存在する。
【００３２】

図２は、コントローラ１１の詳細を示すハードウエアブロック図である。コントローラ１１は、メモリ１１１、ＣＰＵ１１２、上位インターフェース１１３、下位インターフェース１１４、インタフェース１１５、そして、ブリッジ１１６を備えている。

【００３３】

メモリ１１１は、例えば、半導体記憶装置のようなデータ記憶装置である。メモリ１１１には、ＣＰＵ１１２によって実行されるプログラム、及びＣＰＵ１１２によって参照されるデータ、テーブルが格納されている。

【００３４】

メモリ１１１には、少なくとも仮想ボリューム制御プログラム１１７、配置制御プログラム１１８、及び、コピー処理プログラム１１９が格納されている。

【００３５】

仮想ボリューム制御プログラム１１７は、仮想ボリュームを制御するためにプロセッサ１１２によって実行されるプログラムであり、アドレス管理テーブル群１１７１、ページ管理テーブル１１７２を含む。

【００３６】

アドレス管理テーブル群１１７１は仮想ボリュームとディスク装置１２、ＲＡＩＤグループ１３、プール１５のアドレスの対応関係を管理するテーブル群である。ページ管理テーブル１１７２は、図４を利用して、後に説明される。

【００３７】

配置制御プログラム１１８は、ページをＲＡＩＤグループ（以後、ＲＡＩＤグループを「ＲＧ」と略称する。）に配置する際の管理を実行するために、プロセッサによって実行されるプログラムである。

【００３８】

配置制御プログラムは、ＲＧ管理テーブル１１８１、チャンク管理テーブル１１８２、再配置管理テーブル１１８３、及び解放チャンク管理テーブル１１８４を含む。これらのテーブルは、後に、図５乃至図８を用いて説明される。

【００３９】

コピー処理プログラム１１９は、後述のページの移動に伴うコピー処理を制御するためにプロセッサによって実行されるプログラムである。

【００４０】

ＣＰＵ１１２は、メモリ１１１に格納されている各プログラムを実行するプロセッサである。上位インターフェース１１３は、コントローラ１１をホスト計算機２に接続するためのインターフェースである。

【００４１】

なお、ページの移動について、出願人は、PCT/JP2009/058533 PCT/JP2009/058593において明らかにしている。

【００４２】

下位インターフェース１１４は、コントローラ１１をディスク装置１２に接続するためのインターフェースである。下位インクフェース１１４は、例えば、ＳＣＳＩアダプタ又はＦＣアダプタである。

【００４３】

インターフェース１１５は、コントローラ１１を管理端末３と接続するためのインターフェースである。コントローラ１１と管理端末３とを接続するネットワークがＬＡＮである場合、インターフェース１１５は、いわゆるネットワークインターフェースカードである。

【００４４】

ブリッジ１１６は、メモリ１１１、ＣＰＵ１１２、上位インターフェース１１３、下位インターフェース１１４、及びインターフェース１１５を互いに接続し、これらの間で行われる通信を制御する。

【００４５】

図３は、ストレージ装置１が、ホスト計算機に提供する、複数の仮想ボリューム１４（１４ａ，１４ｂ・・・）と記憶装置の記憶領域との関係を示す機能ブロック図である。複数のディスク装置１２が一つのＲＡＩＤグループ１３を構成し、複数のＲＡＩＤグループ１３が一つのプール１５に属している。図３においては、二つのＲＡＩＤグループによってプールが定義されている。

【００４６】

仮想ボリュームは、記憶容量を仮想化し、自身では物理的な記憶容量を持たない論理ボリュームであって、仮想的な記憶容量をホスト計算機に対して提供するものである。

【００４７】

ホスト計算機が仮想ボリュームにライトアクセスすると、コントローラ１１は、プールとして確保されているＲＧの複数のページの中から所定のページを仮想ボリュームのライト領域に割当てる。この割当ては、仮想ボリュームのブロックアドレスとページのブロックアドレスとの対応関係を含み、これはアドレス管理テーブル群１１７１に記録されている。

【００４８】

仮想ボリュームは、ＬＵＮ０及びＬＵＮ１の如く複数存在する。複数の仮想ボリュームがプールに対応付けられている。

【００４９】

プール１５によって管理されている複数のページは、チャンク１３１単位で仮想ボリューム１４に割当てられる。この割当てに相当する情報は、ページ管理テーブル（図４）、ＲＧ管理テーブル（図５）、そして、チャンク管理テーブル（図６）に登録されている。

【００５０】

コントローラ１１は、この情報に基づいてページの割当てなど所定の制御を実行し、又は、管理テーブルに対して、情報を登録又は更新する。ストレージ管理者は、プールの実記憶領域が不足した場合は、プールにＲＧ１３を追加する。ストレージ管理者は、必要に応じて記憶デバイスをストレージ装置に追加する。コントローラ１１は、ＲＡＩＤグループ１３を複数のチャンク１３１に分割して管理する。

【００５１】

チャンク１３１（１３１ａ，１３１ｂ）は、連続する複数のページ１３２から構成される。１つのチャンクは、１つの仮想ボリュームに対応付けられる。１つのチャンク内に異なる仮想ボリュームに割当てられるページが混在することがないようにしている。

【００５２】

１つのストライプ列内に複数の仮想ボリュームに割当てられるページが混在してしまうと、パリティ生成のための処理が複数のコントローラに跨るなど、パリティ生成などの処理が複雑化し、コントローラが受けるオーバヘッドが大きくなってしまう。

【００５３】

そこで、ストレージ装置１は、チャンク管理テーブル（図６）から明らかなように、同一ＲＧ内の同一のストライプ列に合わせて、複数ページから構成されるチャンクを定義し、チャンク単位で一つの仮想ボリュームにページを割当てることにしたものである。

【００５４】

ストレージ装置１は、チャンク１３１をＲＧのストライプ列に合わせて一定のサイズで管理する。すなわち、チャンクとは、パリティが揃った一連の記憶領域としての論理ボリューム（ＬＤＥＶ）である。

【００５５】

図３においては、ＲＧ０のチャンク１３１ａが仮想ボリューム（ＬＵＮ０）１４ａに対応し、チャンク１３１ｂが仮想ボリューム（ＬＵＮ１）１４ｂに対応し、ＲＧ１のチャンク１３１ｃが仮想ボリューム（ＬＵＮ０）１４ａに対応している。

【００５６】

仮想ボリューム１４ａに対するホスト計算機からのアクセスはコントローラＣＬ０によって処理され、仮想ボリューム１４ｂに対するアクセスはコントローラＣＬ１によって処

理される。

【００５７】

なお、ストレージ装置１は、チャンク単位で、ホスト計算機からのＩ／Ｏを処理する担当コントローラを設定するようにしてもよい。例えば、仮想ボリューム１４ａのチャンク１３１ａにコントローラ０を割当て、チャンク１３１ｂにコントローラ１を割当てるようにである。この場合は、一つのチャンク内の複数のページがそれぞれ異なる仮想ボリュームに割り当てられるようにしてもよい。

【００５８】

管理端末３がチャンク管理テーブル（図６）の定義を変更することによって、チャンクと仮想ボリュームの対応関係を変更することが可能である。仮想ボリューム（ＬＵＮ０）１４ａと仮想ボリューム（ＬＵＮ１）１４ｂとによってプール１５が設定されているために、単一プールを異なるコントローラが共用する。

【００５９】

仮想ボリューム１４は、ストレージ装置１がホスト計算機２に提供する記憶領域である。コントローラ１１は、ホスト計算機２からの仮想ボリュームへのデータ書き込みに応じて、この仮想ボリュームに割当てられたチャンクから未割当てのページを選択してこの仮想ボリュームのライト領域に割当てていく。仮想ボリューム１４の記憶領域は、ページ１３２に対応付けられている。初期状態では、仮想ボリュームの全て領域は初期設定用の一つのページ（例えば、全てのビットが“０”からなるページ）に割り当てられている。

【００６０】

ストレージ装置１は、チャンク内の未割当てページにデータを格納し、仮想ボリューム１４の記憶領域とページ１３２とのマッピングを更新していく。この更新は、コントローラ１１が、ページ管理テーブル（図４）とチャンク管理テーブル（図６）の情報を更新することによって達成される。

【００６１】

仮想ボリュームのライトアクセス領域に割当てられたページは、これらページ管理テーブルによって割当済みページとして管理される。ストレージ装置１は、仮想ボリュームに割当てたチャンクに未割当てのページがなくなると、プール１５内のＲＧの未使用チャンクをこの仮想ボリュームに割当てる。このとき、コントローラ１１は、仮想ボリューム－ＲＧ－チャンクの対応関係をチャンク管理テーブル（図６）に登録する。未使用チャンクは使用済みチャンクとして管理される。

【００６２】

仮想ボリュームに割当てられたチャンクは、チャンク管理テーブル（図６）に、使用中チャンクとして登録される。仮想ボリュームに新たに割当てようとするチャンクは、複数のＲＧから均等に選択されることが望ましい。これにより、ストレージ装置１は、各ＲＧ間で負荷を分散させることができる。

【００６３】

ホスト計算機２が、データの消去をストレージ装置に要求する等の場合、消去対象のデータの論理アドレス（仮想ボリュームの）に対応するページは仮想ボリュームへの割当てから解放され、未割当てページとなる。ページ及びチャンクの解放は、コントローラ１１によって、管理テーブルにおいて、エントリーが削除されるか、あるいは、未割当てページ又は未割当てチャンクとして登録されることによって達成される。

【００６４】

未割当てページ・未割当てチャンクは、仮想ボリュームに対する新たなデータ書き込みに利用される。ストレージ装置１は、チャンク内の全てのページを解放した場合は、このチャンクを、仮想ボリュームへの割当てから解放された未使用チャンクに分類する。

【００６５】

ページが随時解放されるものの、チャンク内に割当て済みページが１つでも存在する場合、ストレージ装置はチャンク自体を解放できない。このようなチャンクが多数存在すると、これらチャンクに未割当てである未使用ページが存在していても、この未使用ページ

を、チャンクが対応する仮想ボリューム以外の仮想ボリュームの割当てに利用することができない。これでは、記憶容量の利用効率を低下させ、また、ページを複数のＲＧ間で均等にできないために、ホスト計算機からのＩ／Ｏに対する処理性能を低下させる問題がある。

【００６６】

そこで、コントローラ１１の再配置化制御プログラムは、特定のチャンクの使用済みページを他の使用済みチャンクの未割当て領域に移動させることにより、特定チャンクの使用済みページを解放して、特定チャンクを使用済みページが存在しない空のチャンクとして生成し、この空のチャンクを仮想ボリュームへの割当てから解放するようにした。なお、コントローラ１１は、使用済みページを移動する際には、これを割当て先仮想ボリュームが同じチャンクに移動する。

【００６７】

図１５Ａ及び図１５図Ｂは、ページの移動の前後のプールの状態を示すブロック図である。既述したとおり、符号１３はＲＧを示し、符号１５はプールを示し、符号１３１はチャンクを示し、符号１３２は使用済みのページを示す。図１５Ａはページの移動前であり、図１５Ｂはページの移動後のものである。

【００６８】

図１５Ａと図１５Ｂとを比較すると、チャンク１３１ｅ，１３１ｈ，１３１ｇが、当初使用済みページを有するものであったが、使用済みページを他の使用済みチャンクに移動することによって、未使用のチャンクとして解放されたものになる。

【００６９】

コントローラ１１が使用済みページをプール内でリバランス（再配置）することによって、複数のＲＧ全体に対する使用済みページの割当てが均等になる。複数のＲＧそれぞれに対するＩ／Ｏのレートは同程度になり、ストレージ装置がチャンクに基づいてページを仮想ボリュームに割当てるようにしても、ホスト計算機からのアクセスに対するストレージ装置の処理性能が良好に維持されることになる。

【００７０】

図１６Ａ、図１６Ｂは比較例に係るものである。コントローラが使用済みページを移動する際、チャンクの解放を考慮しないため、ページの移動前（図１６Ａ）よりページの移動後（図１６Ｂ）においてページのリバランスは達成されるものの使用チャンクが増えている（１３１ｆ）分、記憶容量効率が低下する。

【００７１】

図４は、ページ管理テーブル１１７１の一例を示す。チャンクＩＤはチャンクを識別するための識別子、ページＩＤはページを識別するための識別子である。“Class”（種別）はページの種別を表す。例えば、インデックス情報やメタデータ等の格納用ページは“Index”、データ格納用ページは“Data”というように分類される。

【００７２】

“Page Priority”（ページ優先度）は、ページの重要度をあらわし、数値が大きいほど重要性が高いことを示している。“Status”（状態）とは、ページの状態を示し、これには、“Allocated”（割当て済み）、“Unallocated”（未割当て）、“Formatting”（フォーマット中）がある。

【００７３】

ストレージ装置１、あるいは管理端末３は、ページ優先度を、ページ種別やアクセス頻度によって決定する。例えば、インデックス情報やメタデータ等を格納したページには、通常のデータに比べ重要度が高いため、最も優先度が高い“５”が付与される。

【００７４】

通常データのうち、高頻度にアクセスされるページには“３”が付与されており、アクセス頻度の低いページには“１”が付与されている。一方、未割当のページには“０”が付与されている。

【００７５】

複数ページに同じ値のページ優先度が付与されてもよい。このように、ページ毎に重要度が付与されることにより、重要度が高いページがなるべく性能低下しないように移動ページを決定することかできる。すなわち、ストレージシステムは、重要度の高いページ(チャンク)を確保し、残りの重要度の低いものを解放しようとするので、重要度の高いページは移動せずにすむ確率が高くなる。

【００７６】

ページの優先度はチャンクの優先度を判定する際のファクターとなる。チャンクの優先度は、解放チャンクを判定する際のファクターとなる。

【００７７】

図５は、ＲＧ管理テーブルの一例である。この管理テーブルは、仮想ボリューム毎に、各ＲＧのチャンク数やページ数を管理する。“Virtual Volume”は仮想ボリュームを識別するための識別子、“RAID Group”はＲＡＩＤグループを識別するための識別子である。

【００７８】

“Chunk Size”は１チャンクあたりの容量を表す。“Page Size”は１ページあたりの容量を表す。“Used Chunks”は、仮想ボリューム毎、ＲＧ毎に、仮想ボリュームに割当てられているチャンク数を表す。“Total Allocated Pages”は、仮想ボリュームに割当てられているＲＧ毎で、割当て済みのページ数の合計を表す。

【００７９】

図６は、チャンク管理テーブルの一例である。チャンク管理テーブルは、仮想ボリュームとＲＧ毎に、各チャンクのページ数や優先度を管理する。“Virtual Volume”は仮想ボリュームを識別するための識別子、“RAID Group”はＲＧを識別するための識別子、“Chunk ID”はチャンクを識別するための識別子である。“Allocated Pages”は、そのチャンク中の割当て済みページの数を表す。

【００８０】

“Unallocated Pages”は、チャンク中の未割当てのページの数を表す。このテーブルの例では、１チャンクあたりのページの数は“５”であるとしている。実環境においては、１チャンクあたりのページ数は数千個以上に及ぶこともありえる。

【００８１】

“Chunk Priority”はチャンクの優先度を表し、数値が大きいほど重要度が高いことを示している。チャンク優先度は、例えば、チャンクに属するページの優先度によって決定される。チャンクに属する複数のページの優先度を合計した値がチャンクの優先度とされる。

【００８２】

図７は、チャンクを仮想ボリュームに対して、解放したり、あるいは新たに割当てるなどの再配置を実行するための管理テーブルの一例である。“Virtual Volume”は仮想ボリュームを識別するための識別子、“RAID Group”はＲＧを識別するための識別子である。

【００８３】

“mode”は、再配置方法の種別（後述）と、実行契機が自動であるか手動であるかを表す特性を示すものである。“確保チャンク”は、仮想ボリュームに対して、解放されることなく、対応するＲＧにおいてそのまま維持されるべきチャンクのＩＤを表す。“不足チャンク数”は、ＲＧから対応する仮想ボリュームに対して新たに割当てが必要なチャンクの数を表す。

【００８４】

図８は、仮想ボリューム毎に解放すべきチャンクを管理するテーブルの一例である。解放予定のチャンクがこの管理テーブルに登録される。解放されるチャンクに属するページが他のチャンクに移行すべきページとなる。

【００８５】

図６にも示したように、“Virtual Volume”は仮想ボリュームを識別するための識別子、“Chunk ID”はチャンクを識別するための識別子、“Allocated Pages”はチャンクの割当て済みページ数、“Unallocated Pages”はチャンクの未割当てページ数、“Chunk P

riority”はチャンクの優先度を表す。

【００８６】

図４において、Chunk ID他の項目は、仮想ボリューム制御プログラム１１１によって登録される。図５において、仮想ボリュームとＲＧの対応関係、チャンクサイズ、ページサイズは、仮想ボリューム制御プログラム１１７によって登録され、その他の項目は配置制御プログラム１１８によって登録される。

【００８７】

図６において、仮想ボリュームとＲＧの対応関係、チャンクＩＤは、仮想ボリューム制御プログラム１１７によって登録され、その他の項目は配置制御プログラム１１８によって登録される。

【００８８】

図7において、仮想ボリュームとＲＧの対応関係、チャンクＩＤは、仮想ボリューム制御プログラム１１７によって登録され、その他の項目は配置制御プログラム１１８によって登録される。

【００８９】

図８において、仮想ボリュームとチャンクＩＤの対応は、仮想ボリューム制御プログラム１１７によって登録され、その他の項目は配置制御プログラム１１８によって登録される。

【００９０】

図９は、管理端末３によって管理ユーザ・管理クライアントに提供される、設定画面３１１の一例を示すものである。管理ユーザは設定画面３１１を用いて、チャンクの解放のための制御情報をコントローラ１１の再配置制御プログラム１１８に対して設定する。

【００９１】

設定画面３１１は、ＬＵＮ設定部３１１１、実行契機選択部３１１２、モード選択部３１１３及び解放ＲＡＩＤグループ選択部３１１４を備えている。ＬＵＮ設定部３１１１は、仮想ボリューム１４のＬＵＮ（Logical Unit Number）を指定する設定部である。この設定部は、チャンクの解放が実行されるべき対象の仮想ボリュームを設定するためのものである。

【００９２】

実行契機選択部３１１２は、チャンクの解放と、チャンクの解放と一体に行われるページの再配置のための解放処理の実行契機を自動とするか手動とするかを選択する設定部である。

【００９３】

この契機には、チャンク数の偏り（後述）、未使用チャンク数の減少ある。チャンク数の偏りが生じるケースには、プールボリュームの追加、仮想ボリュームの容量縮小、ページの重複排除等がある、また、ページのリバランス契機でチャンクの解放を行ってもよい。

【００９４】

モード選択部３１１３は、解放処理のモードを管理ユーザが選択するための設定部である。管理ユーザは、通常、記憶容量の効率的な利用を図る容量優先モードを選択すればよい。

【００９５】

一方、ホスト計算機が、ストレージ装置をデータベースとして利用する場合など、ストレージ装置の性能をより優先したい場合には、管理ユーザは、性能優先モードを選択すればよい。

【００９６】

性能優先モードは、さらに細かい設定を含んでいる。例えば、管理ユーザが、複数の仮想ボリュームに平均的に高性能を実現させたい場合には、ワイドストライピングモードを選択する。

【００９７】

また、ある仮想ボリュームの性能が、他の仮想ボリューム１４によって影響を受けるのを抑止させたい場合には、管理ユーザは、ディスク占有モードを選択する。

【００９８】

仮想ボリュームの性能とは、この仮想ボリュームに対するホスト計算機からのＩ／Ｏに対する、ストレージ装置の処理性能のことである。

【００９９】

ここで説明した“設定”は、仮想ボリューム毎に可能である。また、操作を簡単にするために、プール単位にまとめて指定するようなインターフェースを設けても良い。また、他の選択肢を追加してもよい。

【０１００】

次に、解放RAID Groupe選択部３１１４について説明する。これは、解放処理を適用したいＲＧを管理ユーザが選択するためのものである。

【０１０１】

管理ユーザは、プールＩＤとプールに属するＲＧのＩＤを指定する。管理ユーザがＲＧを指定すると、そのＲＧに属するチャンクの全てが解放処理を適用可能なものとなる。

【０１０２】

図１０は、ＬＵ設定処理を説明するフローチャートである。管理クライアントはＬＵＮを指定して実行契機、及び、モードを選択する（１０００）。

【０１０３】

管理プログラム３１は、入力及び選択された情報を、ストレージ装置１のコントローラ１１へ送信する（１００２）。コントローラ１１の配置制御プログラム１１８は、受信した情報を、図７に示す再配置管理テーブル１１８３に登録する（１００４）。具体的には、実行契機が自動であるか手動であるかを表す情報と、再配置モードを表す情報を登録する。

【０１０４】

再配置制御プログラムは、実行契機が自動であるか手動であるかを判断し(１００６)、手動の場合は、指定されたモードの再配置処理を実行する(１００８)。自動の場合は設定処理を一旦終了し、後ほど適切なタイミングで再配置処理を実行する。

【０１０５】

適切なタイミングとは、例えば配置制御プログラムが定期的にＲＧ管理テーブルを参照し、使用チャンク数や合計割当て済みページ数に、仮想ボリュームに対して、偏りが生じた場合である。使用チャンクとは、ページの割り当て以前に仮想ボリュームに割り当てられているチャンクをいう。

【０１０６】

図１１は、管理端末３が、ページの移動やチャンクの解放をストレージ装置１に設定するための処理（ＲＧ解放設定処理）を説明するフローチャートである。管理ユーザは管理プログラム３１に、プールＩＤとＲＧを指定する（１１００）。管理プログラム３１は、管理端末３に入力及び選択された情報をストレージ装置１のコントローラ１１へ送信する（１１０２）。

【０１０７】

そして、コントローラ１１の再配置制御プログラム１１８は、プールのＩＤ及び／又はＲＧのＩＤをＲＧ管理テーブル（図５）、チャンク管理テーブル（図６）、再配置管理テーブル（図７）に登録して、指定されたＲＧの解放処理を実行する。

【０１０８】

図１２は、容量優先モードでの再配置処理を説明するフローチャートである。“容量優先”とは、プール１５において、仮想ボリュームに割当て済みのチャンク数又はその容量を極力低減させるモードである。

【０１０９】

配置制御プログラム１１８は、プールのチャンクを管理するチャンク管理テーブル（図６）を参照し、管理テーブルに登録されている、複数の仮想ボリュームについて、割当済

みページの合計容量を算出する（１２００）。以下、この複数の仮想ボリュームを対象仮想ボリュームと呼ぶ。

【０１１０】

なお、ストレージ装置１は、容量優先モードの再配置処理を一つ又はいくつかの仮想ボリュームを対象として適用してもよい。チャンクの解放処理の形態には、例えば、プールの指定の他に仮想ボリュームの指定がある。

【０１１１】

配置制御プログラム１１８は、算出されたページ容量をチャンクサイズの平均値で割ることによって、ページの割当てに必要なチャンク数を算出する。そして、必要チャンク数をプールに属するＲＧ数で割ることによって、ＲＧ毎に最低限のチャンク数を算出する（１２０２）。なお、ページ容量は、各ページサイズとページ割当量から算出される。配置制御プログラム１１８は、チャンクサイズの平均値によって、必要なチャンク数を算出してこれを確保することとするが、その後、それぞれのチャンクサイズを精査・確認して、足りなければチャンクを追加し、多ければチャンクを減らすようにしてもよい。

【０１１２】

配置制御プログラム１１８は、図６に示す、チャンク管理テーブルを参照し、各ＲＧについて、最低チャンク数分だけ、重要度の高いチャンク、例えば、割当て済みページ数の多いチャンク、その中でも、優先度の高いチャンクを選出する（１２０４）。

【０１１３】

すなわち、前者による重要度が同じ場合は、最も優先度が高いページを有するチャンクを確保する。そして、確保したチャンクを、再配置管理テーブル（図７）に確保チャンクとして、そのＩＤを記録する。

【０１１４】

配置制御プログラム１１８は、最低チャンク数に満たないＲＧの存在を判定すると（１２０５）、不足しているチャンク数を再配置管理テーブル（図７）の不足チャンク数に登録する。

【０１１５】

配置制御プログラム１１８は、再配置管理テーブルを随時参照して、不足チャンク数を確認すると、最低チャンク数に満たないチャンクのＩＤを確認し、このチャンクの残りの使用済みチャンクの中から重要度が高チャンクを解放せず維持するチャンクとして確保する（１２０６）。

【０１１６】

再配置制御プログラム１１８は、次いで、再配置管理テーブル（図７）の不足チャンク数を０に更新する。これにより、各ＲＧに最低チャンク数分のチャンクが確保される。確保したチャンクのチャンクＩＤを再配置管理テーブルの確保チャンクとして記録する。

【０１１７】

ステップ１２０６の一例を具体的に説明すると次のとおりである。配置制御プログラム１１８は、最低チャンク数に端数が生じた場合には、小数点以下を切り捨てた分のチャンク数を必要チャンク数として確保する。例えば、最低チャンク数の計算値が４．７である場合には、小数点以下を切り捨てて最低チャンク数を“４”にする。

【０１１８】

次に、配置制御プログラム１１８は、各ＲＧの最低チャンク数を全てのＲＧについて合計した値が、対象仮想ボリュームについての必要チャンク数を満たしているかを判定する。

【０１１９】

再配置制御プログラム１１８は、これを肯定判定すると、次のステップ（１２０８）に進む。但し、既述のとおり、端数が切り捨てられたことによって、必要チャンク数は実際より少なく見積もられているために、再配置制御プログラム１１８は、否定判定を行う。

【０１２０】

配置制御プログラムは、不足分のチャンクを、確保チャンク以外から選択する。配置制

御プログラムは、確保されたチャンク数が少ないＲＧを優先して、複数のＲＧから順番に、同じＲＧが重複しないように、重要度がより高いチャンクを選択する。不足分のチャンクが選択された段階で、最低チャンク数の確保の処理を終了する。配置制御プログラムは、選択したチャンクのＩＤを再配置管理テーブル（図７）の確保チャンクとして記録する。

【０１２１】

次いで、配置制御プログラム１１８は、確保チャンクとして、配置管理テーブル（図７）に記録されなかった、使用済みチャンクのＩＤを解放チャンク管理テーブル（図８）に登録する（１２０８）。このチャンクは、仮想ボリュームへの割当てから解放されるべきチャンクであって、便宜上、これを解放チャンクと呼ぶこととした。

【０１２２】

コピー処理プログラム１１９は、解放チャンク管理テーブル（図８）を参照し、登録されているチャンク内のページを、確保チャンク内の未割当ページへ移行する（１２１０）。配置制御プログラム１１８は、解放チャンク内の全てのページ移行が完了したら、解放チャンク管理テーブル（図８）からこのチャンクのエントリを削除する。

【０１２３】

解放されたチャンクはフリーキューに接続され、以降、仮想ボリューム制御プログラム１１７は、仮想ボリュームへの割当てにこの解放チャンクを充当する。ページの解放についても同様である。ページの割当ては、仮想ボリューム制御プログラム１１７によって実行される。

【０１２４】

ページの移行は、コピー処理プログラム１１９が移行元ページのデータを移行先ページにコピーし、移行元ページを仮想ボリュームへの割当てから解放することによって実行される。

【０１２５】

コピー処理プログラム１１９は、ページの移行の都度、解放チャンク管理テーブルを参照し、登録されていた全ての解放チャンクのエントリが、解放チャンク管理テーブルから削除されたか否か、すなわち、全ての解放チャンクについての処理が完了したか否かを判定する。これを肯定すると、一連の処理を終了する。そうでなければ、管理テーブルに登録されている解放チャンクについてページの移動の処理を繰り返す。

【０１２６】

配置制御プログラム１１８は、解放チャンクにおいて、全ての使用済ページについて、仮想ボリュームへの割当てが解放されたか否かを判定し、これを判定すると、解放チャンク管理テーブル（図８）での解放チャンクのエントリを削除する。

【０１２７】

図１２に説明したフローチャートは、ＲＧ毎でのページ配置の偏りの解消によるＩ／Ｏ処理性能の改善と、過剰チャンクを解放することによる、ストレージ容量の効率向上を可能にする。

【０１２８】

なお、図１２の説明において、最低チャンク数の端数を切り捨てたが、これを切り上げることも考えられる。このとき、各ＲＧグループの最低チャンク数の合計が仮想ボリュームの必要チャンク数を超えるが、再配置制御プログラムは、超える分のチャンクを解放チャンクとして、解放チャンク設定テーブルに登録する。再配置制御プログラムは、ページの割当て数が少なく、優先度が低いチャンクを解放チャンクとして設定する。

【０１２９】

図１３は、図１２の変形例であって、ワイドストライプモードによる再配置処理を説明するフローチャートである。ワイドストライプモードとは、チャンク解放に伴うページ再配置処理に対するストレージ装置のオーバヘッドを低減しようとするモードである。

【０１３０】

図１２のフローチャートと異なる点は、図１２の実施形態では、最低チャンク数に端数

が生じた場合、再配置プログラムは少数点以下の端数を切り捨てた値を最低チャンク数とし、次いで、不足分のチャンク数を使用済みチャンクから確保するようにしたが、図１３では少数点以下の端数を切り上げて、複数のＲＧのそれぞれについて、最低チャンク数を上回る、同数のチャンクを確保するようにした。

【０１３１】

すなわち、図１３のフローチャートは、図１２のフローチャートには存在していた、端数分のチャンクを新たに確保しなければならないステップ（１２０５,１２０６）を削除できる分、コントローラのオーバヘッドを解消することができる。

【０１３２】

一方、図１３の実施形態は、必要チャンク数を超えるチャンクの仮想ボリュームへの割当てを解放しないために、記憶容量の効率的な活用の点において、図１２の実施形態に及ばないことになる。

【０１３３】

図１４は、ディスク占有モードの再配置処理を説明するフローチャートである。このモードは、特定の仮想ボリュームに特定のＲＧを対応させることにより、このＲＧを他の仮想ボリュームへの割当てから開放して当該他の仮想ボリュームからの影響が及ばない様にして、解放処理のオーバヘッドを低減しようとするものである。

【０１３４】

配置制御ブログラム１１８は、チャンク管理テーブルを参照し、チャンク管理テーブルにエントリがある全ての仮想ボリュームについて、合計の割当済みページ数を算出する（１４００）。

【０１３５】

次いで、配置制御プログラムは、算出されたページ数を基に、必要チャンク数を算出する（１４０２）。配置制御プログラムは、チャンク管理テーブルを参照し、特定の仮想ボリュームについて、割当て済ページ数の多いＲＧを選出する。割当てページ数が同じ場合には、最も優先度が高いチャンクが多く存在するＲＧを選出する（１４０４）。

【０１３６】

配置制御プログラム１１８は、選出されたＲＧに必要チャンク数を確保する（１４０６）。このとき、割当て済みページ数の多いチャンクを選択する。割当て済みページ数が同じチャンクが存在する場合には、優先度が高い方を選択する。確保したチャンクを、再配置管理テーブル（図７）の確保チャンクに登録する。

【０１３７】

さらに、配置制御プログラムは１１８、選択したＲＧの使用中チャンク数が必要チャンク数に満たないと判定した場合は、不足しているチャンク数を再配置管理テーブルの不足チャンク数に登録し、不足チャンク数分のチャンクを同じＲＧから割当てる。不足チャンクの割当てが完了したら再配置管理テーブル（図７）の不足チャンク数を０に更新する。これによりＲＧに必要チャンク数分のチャンクを確保する。

【０１３８】

配置制御プログラム１１８は、使用中チャンク数が必要チャンク数を超えると判断すると、使用中チャンクのうち確保されなかったチャンクを解放チャンク管理テーブル（図８）に登録する。さらに、配置制御プログラム１１８は、選出されなかった他のＲＧのチャンクを解放すべきチャンクとして、解放チャンク管理テーブルに登録する。

【０１３９】

コピー処理プログラム１１９は、解放チャンク管理テーブル（図８）を参照し、登録されているチャンク内のページを確保チャンク内の未割当ページへ移行する（１４１０）。再配置制御プログラム１１８は、解放チャンク内の全てのページ移行が完了したらこのチャンクを解放し、解放チャンク管理テーブルからチャンクのエントリを削除する。

【０１４０】

再配置制御プログラム１１８は、解放チャンク管理テーブル（図８）を参照し、登録されていた全ての解放チャンクが完了したかを判定する（１４１２）。解放チャンク管理テ

ーブルにエントリが存在しなければ、完了したと判断する。完了していなければ、未実施の解放チャンクに対して処理を繰り返す。

【０１４１】

次に、再配置制御プログラム１１８が、チャンク管理テーブル（図６）を参照して、他の仮想ボリュームに割当てられたチャンクが、選出されたＲＧに存在すると判断した場合は、このチャンクを解放予定のチャンクとして開放チャンク管理テーブル（図８）に登録する（１４１４）。

【０１４２】

再配置制御プログラム１１８は、他の仮想ボリュームに割当てられ、選出されたＲＧに存在するチャンクを解放チャンクとする（１４１４）。再配置制御プログラム１１８は、解放チャンクのページが移行されるチャンクを、他のＲＧに必要チャンクとして確保する。

【０１４３】

以上により、特定の仮想ボリュームに、特定のＲＧのページが占有されて割当てられるために、ストレージ装置は、特定の仮想ボリュームに対して他の仮想ボリュームによる影響を抑止して、特定仮想ボリュームに対するＩ／Ｏ処理の効率を維持、向上することができる。

【産業上の利用可能性】

【０１４４】

本発明は、ページの割当てに十分活用されていない特定管理単位を仮想ボリュームへの割当てから解放し、この特定管理単位に属する割当て済みページを、他の管理単位に移動させることにより、特定管理単位の記憶領域を他の仮想ボリュームへのホスト計算機からのライトアクセスに供するようにしたので、ストレージシステム及びストレージ装置の記憶領域を最適化する際に利用されるものである。

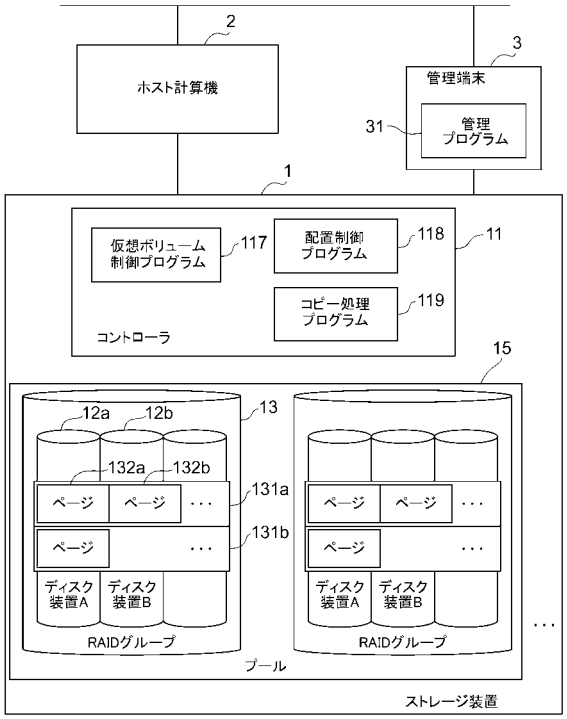
【符号の説明】

【０１４５】

１　ストレージ装置
２　ホスト計算機
３　管理端末
１１　コントローラ
１２　ディスク装置
１３　RAIDグループ
１４　仮想ボリューム
１５　プール
１３１　チャンク
１３２　ページ

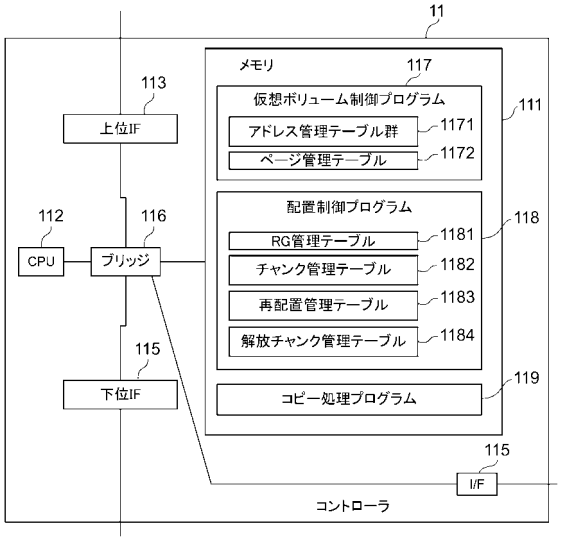
【図１】

図１



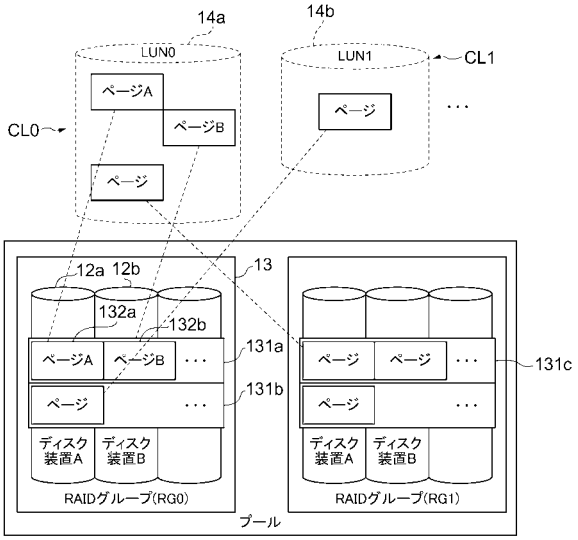
【図２】

図２



【図３】

図３



【図４】

図４

ページ管理テーブル

CHUNK ID	PAGE ID	STATUS	CLASS	PAGE PRIORITY
0	0	ALLOCATED	DATA	3
	1	ALLOCATED	DATA	1
	2	UNALLOCATED	-	0
	3	UNALLOCATED	-	0
	4	UNALLOCATED	-	0
1	10	ALLOCATED	INDEX	5
	11	UNALLOCATED	-	0
	12	ALLOCATED	DATA	3
	13	UNALLOCATED	-	0
	14	ALLOCATED	DATA	1
2	20	ALLOCATED	DATA	1
	21	ALLOCATED	DATA	1
	22	UNALLOCATED	-	0
	23	ALLOCATED	DATA	3
	24	UNALLOCATED	-	0
3	31	FORMATTING	-	0
	32	・・・	・・・	・・・
	・・・	・・・	・・・	・・・
・・・	・・・	・・・	・・・	・・・

【図５】

図５

RG管理テーブル

VIRTUAL VOLUME	RAID GROUP	CHUNK SIZE	PAGE SIZE	USED CHUNKS	TOTAL ALLOCATED PAGES
0	0	100MB	20MB	3	8
	1	100MB	20MB	3	9
	2	100MB	20MB	1	1
1	0	1GB	64KB	180	2110
		2GB	・・・	・・・	・・・
	1	・・・	・・・	・・・	・・・
	・・・	・・・	・・・	・・・	・・・
・・・	・・・	・・・	・・・	・・・	・・・

【図６】

図６

チャンク管理テーブル

VIRTUAL VOLUME	RAID GROUP	CHUNK ID	ALLOCATED PAGES	UNALLOCATED PAGES	CHUNK PRIORITY
0	0	0	2	3	4
		1	3	2	9
		2	3	2	5
		-	-	-	-
	1	10	2	3	2
		11	3	2	7
		12	4	1	8
		-	-	-	-
	2	20	1	4	1
		-	-	-	-
1	0	4	・・・	・・・	・・・
		9	・・・	・・・	・・・
		-	・・・	・・・	・・・
	・・・	・・・	・・・	・・・	・・・
・・・	・・・	・・・	・・・	・・・	・・・

【図７】

図７

再配置管理テーブル(仮想ボリューム毎)

VIRTUAL VOLUME	MODE	RAID GROUP	確保 チャンク	不足 チャンク数
0	自動, 容量優先	0	1	0
		1	11,12	0
		2	20	0
		-	-	-

【図８】

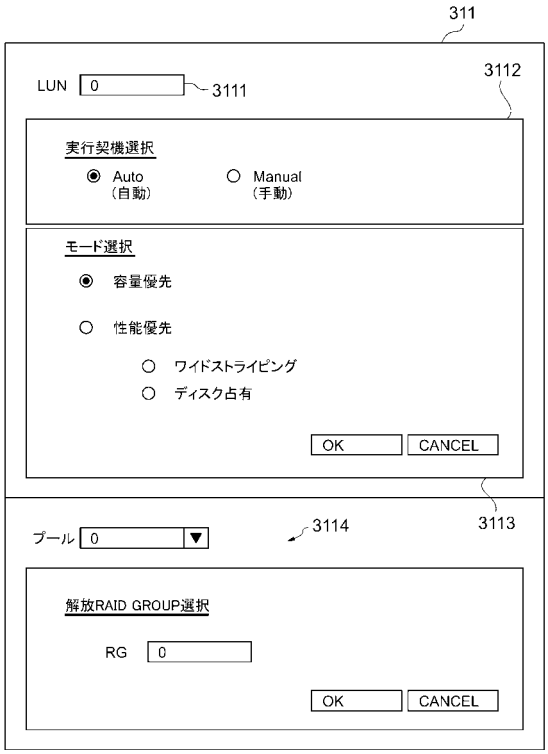
図８

解放チャンク管理テーブル(仮想ボリューム毎)

VIRTUAL VOLUME	CHUNK ID	ALLOCATED PAGES	UNALLOCATED PAGES	CHUNK PRIORITY
0	0	2	3	4
	10	3	2	2
	1	3	2	9
	-	-	-	-

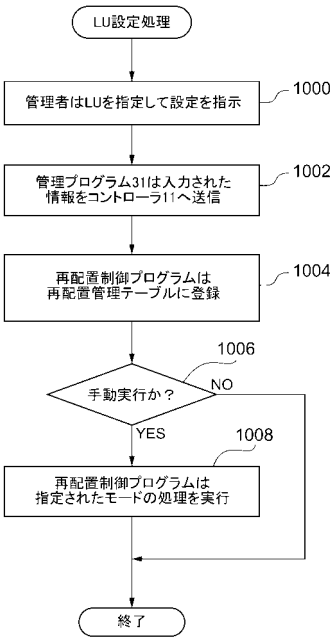
【図９】

図９



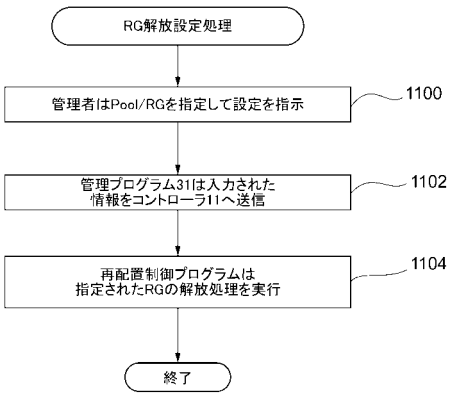
【図１０】

図１０



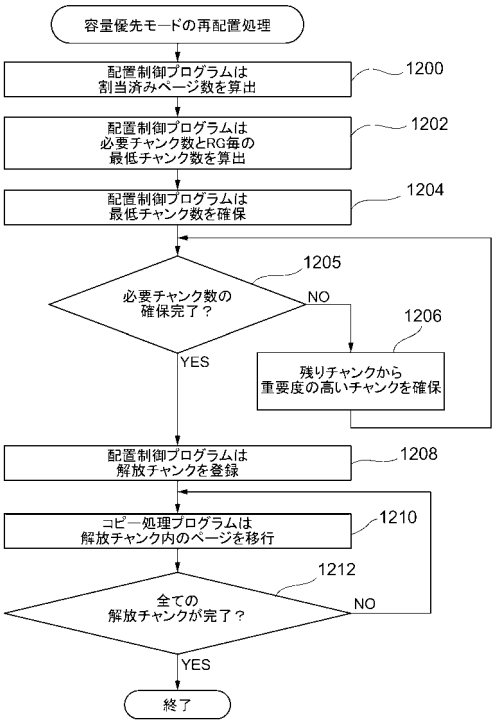
【図１１】

図１１



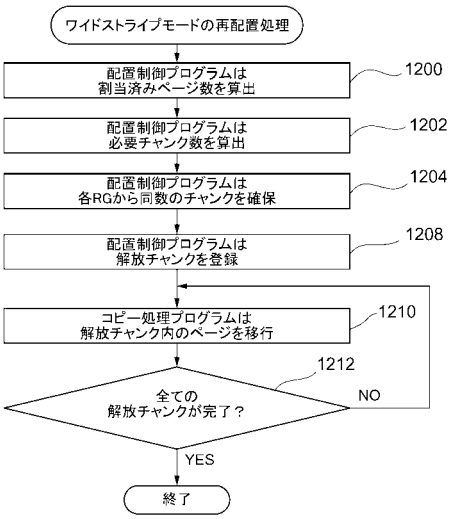
【図１２】

図１２



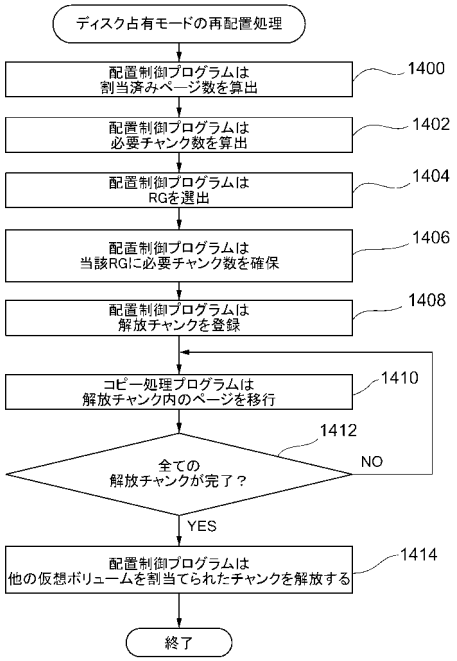
【図１３】

図１３



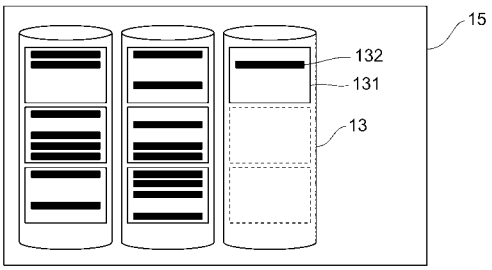
【図１４】

図１４



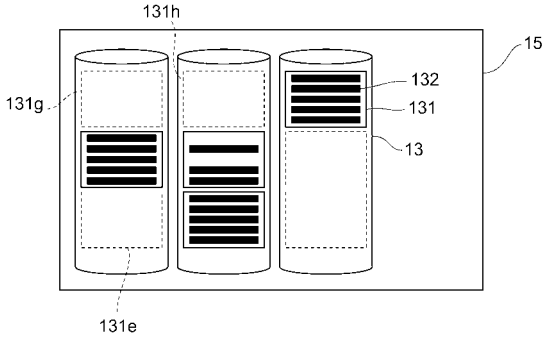
【図１５Ａ】

図１５Ａ



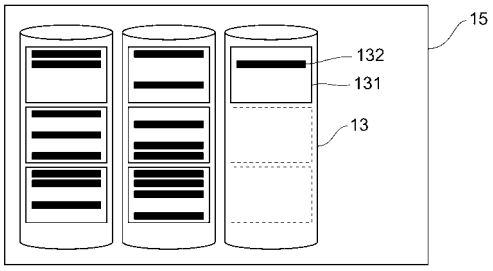
【図１５Ｂ】

図１５Ｂ



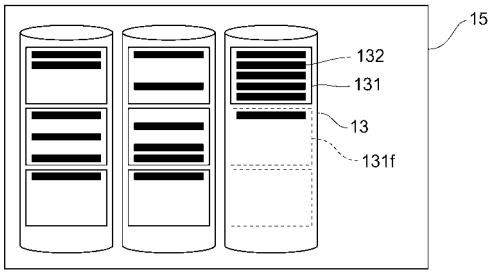
【図１６Ａ】

図１６Ａ



【図１６Ｂ】

図１６Ｂ



フロントページの続き

(56)参考文献 特開２００５－２８４６３２（ＪＰ，Ａ）
特開２００８－０８４０５３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ 3／０６－ 3／０８
Ｇ０６Ｆ １２／００－１２／１６
Ｇ０６Ｆ １３／００－１３／４２