

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5216463号
(P5216463)

(45) 発行日 平成25年6月19日(2013.6.19)

(24) 登録日 平成25年3月8日(2013.3.8)

(51) Int.Cl.

F I

G 0 6 F 3 / 0 6 (2006.01)

G 0 6 F 3 / 0 6 3 0 1 T

請求項の数 15 (全 24 頁)

(21) 出願番号	特願2008-195654 (P2008-195654)	(73) 特許権者	000005108
(22) 出願日	平成20年7月30日(2008.7.30)		株式会社日立製作所
(65) 公開番号	特開2010-33393 (P2010-33393A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成22年2月12日(2010.2.12)	(74) 代理人	100093861
審査請求日	平成23年1月18日(2011.1.18)		弁理士 大賀 真司
		(72) 発明者	加納 義樹
			神奈川県川崎市麻生区王禅寺1099番地
			株式会社日立製作所システム開発研究所
			内
		(72) 発明者	杉本 定広
			神奈川県川崎市麻生区王禅寺1099番地
			株式会社日立製作所システム開発研究所
			内
		審査官	古河 雅輝

最終頁に続く

(54) 【発明の名称】 ストレージ装置、その記憶領域管理方法及びフラッシュメモリパッケージ

(57) 【特許請求の範囲】

【請求項1】

一つ以上のホスト計算機から送信されたデータを記憶するストレージ装置において、
複数のフラッシュメモリチップを備えた一つ以上のフラッシュメモリパッケージを有し

、
前記一つ以上のフラッシュメモリパッケージにより提供される記憶領域は、一つ以上の
論理デバイスから構成される実データを記憶する領域である第1の領域と、前記論理デバ
イスに対する前記ホスト計算機のライト命令をログ形式で追記して保管する領域である第
2の領域とを前記一つ以上のフラッシュメモリパッケージ毎に有し、

前記第2の領域へのライト命令の追記を抑止して、前記第2の領域に保管されたすべて
のデータを所定のメモリ上に書き出した後、前記第1の領域に該データを書き出すリクラ
メーション処理を実行するリクラメーション処理部と、

前記リクラメーション処理の発生頻度を監視する監視部と、

前記リクラメーション処理の発生頻度に応じて、前記第2の領域のサイズを変更する変
更部とを備える、

ことを特徴とするストレージ装置。

【請求項2】

前記変更部の前記第2の領域のサイズの変更は、前記論理デバイス毎に変更する、
ことを特徴とする請求項1記載のストレージ装置。

【請求項3】

10

20

所定のリラクメーション頻度の範囲のうちの上限值を超えた場合に前記第 2 の領域の追加の要求を、管理装置へ出力する追加要求出力部を備える、

ことを特徴とする請求項 2 記載のストレージ装置。

【請求項 4】

前記管理装置から前記上限値の設定を変更する指示を取得する取得部を備える、

ことを特徴とする請求項 3 記載のストレージ装置。

【請求項 5】

前記ホスト計算機のいずれかから前記上限値の設定を変更する指示を取得する取得部を備える、

ことを特徴とする請求項 3 記載のストレージ装置。

【請求項 6】

所定のリラクメーション頻度の範囲のうちの下限值を下回った場合に前記第 2 の領域の削減の要求を、管理装置へ出力する削減要求出力部を備える、

ことを特徴とする請求項 2 記載のストレージ装置。

【請求項 7】

前記管理装置から前記下限値の設定を変更する指示を取得する取得部を備える、

ことを特徴とする請求項 6 記載のストレージ装置。

【請求項 8】

前記ホスト計算機のいずれかから前記下限値の設定を変更する指示を取得する取得部を備える、

ことを特徴とする請求項 6 記載のストレージ装置。

【請求項 9】

前記一つ以上のフラッシュメモリパッケージは、R A I D (Redundant Arrays of Inexpensive Disks) グループを構成する、

ことを特徴とする請求項 1 記載のストレージ装置。

【請求項 10】

一つ以上のホスト計算機から送信されたデータを記憶するストレージ装置の記憶領域管理方法において、

複数のフラッシュメモリチップを備えた一つ以上のフラッシュメモリパッケージにより提供される記憶領域は、一つの以上の論理デバイスから構成される実データを記憶する領域である第 1 の領域と、前記論理デバイスに対する前記ホスト計算機のライト命令をログ形式で追記して保管する領域である第 2 の領域とを前記一つ以上のフラッシュメモリパッケージ毎に有し、

前記第 2 の領域へのライト命令の追記を抑止して、前記第 2 の領域に保管されたすべてのデータを所定のメモリ上に書き出した後、前記第 1 の領域に該データを書き出すリクレーション処理を実行するステップと、

前記リラクメーション処理の発生の頻度を監視するステップと、

前記リラクメーション処理の発生頻度に応じて、前記第 2 の領域のサイズを変更するステップと、

を含むことを特徴とするストレージ装置の記憶領域管理方法。

【請求項 11】

前記第 2 の領域のサイズの変更するステップは、前記論理デバイス毎に変更できる、

ことを特徴とする請求項 10 記載のストレージ装置の記憶領域管理方法。

【請求項 12】

所定のリラクメーション頻度の範囲のうちの上限值を超えた場合に前記第 2 の領域の追加の要求を、管理装置へ出力するステップを含む、

ことを特徴とする請求項 11 記載のストレージ装置の記憶領域管理方法。

【請求項 13】

所定のリラクメーション頻度の範囲のうちの下限值を下回った場合に前記第 2 の領域の削減の要求を、管理装置へ出力するステップを含む、

10

20

30

40

50

ことを特徴とする請求項 1 1 記載のストレージ装置の記憶領域管理方法。

【請求項 1 4】

前記一つ以上のフラッシュメモリパッケージは、R A I D グループを構成する、
ことを特徴とする請求項 1 0 記載のストレージ装置の記憶領域管理方法。

【請求項 1 5】

一つ以上のホスト計算機から送信されたデータを記憶するストレージ装置に用いられる、
複数のフラッシュメモリチップを備えたフラッシュメモリパッケージは、

一つの以上の論理デバイスから構成される実データを記憶する領域である第 1 の領域と、
前記論理デバイスに対する前記ホスト計算機のライト命令を ログ形式で追記して保管する領域である第 2 の領域とを有し、

前記第 2 の領域へのライト命令の追記を抑止して、前記第 2 の領域に保管されたデータを所定のメモリ上に書き出した後、前記第 1 の領域にデータを書き出すリクラメーション処理を実行するリクラメーション処理部と、

前記リクラメーション処理の発生頻度を監視する監視部と、

前記リクラメーション処理の発生頻度に応じて、前記第 2 の領域のサイズを変更する変更部とを備える、

ことを特徴とするフラッシュメモリパッケージ。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、計算機システムに用いられるストレージ装置に関し、特に、記憶媒体としてフラッシュメモリなどの不揮発性半導体メモリを用いたストレージ装置、その記憶領域管理方法及びフラッシュメモリパッケージに関する。

【背景技術】

【0 0 0 2】

ストレージ装置は、一般的に、ランダムアクセス可能な不揮発性記憶媒体を備える。ランダムアクセス可能な不揮発性記憶媒体は、例えば、磁気ディスク（以下ハードディスクとも呼ぶ）、光ディスク等であり、例えば特許文献 1 のように、ハードディスクを多数備えた構成をとる。

【0 0 0 3】

また、近年、従来のハードディスクの代替としてフラッシュメモリなどの不揮発性半導体メモリを記憶媒体としたストレージ装置が注目を集めている。フラッシュメモリはハードディスクに比べて高速動作可能で、かつ消費電力が低いという利点を持つ。特許文献 2 には、ストレージ装置において、複数のフラッシュメモリを備え、S C S I (Small Computer System Interface) 等の従来のハードディスクのアクセス手段でアクセス可能としたフラッシュメモリディスクを、ストレージ装置のハードディスクの代替として利用する技術が開示されている。

【0 0 0 4】

フラッシュメモリの効率的な書き込み方式を実現する方式として、特許文献 3 がある。この方式では、フラッシュメモリのブロック内を、ライト領域とデータ領域とにわけ、ライトデータはライト領域にログ形式で追記し、ライト領域にデータがまとまった段階で、ブロックのデータをメモリに読み出し、ライト領域のデータをメモリ上で上書き後、データ領域にデータを書き出す方式がある。なお、この書き込み方式をリクラメーション方式と呼ぶ。

【特許文献 1】特開 2 0 0 4 - 5 3 7 0 号公報

【特許文献 2】米国特許第 6 5 2 9 4 1 6 号公報

【特許文献 3】米国特許第 7 0 3 9 7 8 8 号公報

【発明の開示】

【発明が解決しようとする課題】

【0 0 0 5】

先に述べたように、フラッシュメモリの領域をデータ領域とライト領域に分けてライトの性能を向上させる方式の場合、ホスト計算機のアプリケーションによっては、ライト命令が少ないがゆえにライト領域のリクラメーションがほとんど発生しない場合があり、割り当てられたライト領域が効率的に利用されないという課題がある。

【0006】

上記の課題を解決するには、ホスト計算機のライト性能にあわせて、該当するデータ領域に対応するライト領域を割り当てられるストレージ装置を実現する必要がある。

【0007】

また、ライト領域は、ライト性能に合わせて削減する場合にライト性能にあわせてフラッシュメモリパッケージの減設することも必要である。

10

【0008】

本発明は、以上の点を考慮してなされたもので、ホスト計算機からのライト命令に応じてライト領域を効率的に割り当てることが可能とするストレージ装置、その記憶領域管理方法及びフラッシュメモリパッケージを提案しようとするものである。

【課題を解決するための手段】

【0009】

本発明は、複数のフラッシュメモリを備えた一つ以上のフラッシュメモリパッケージを有し、一つ以上のホスト計算機から送信されたデータを記憶するストレージ装置において、前記一つ以上のフラッシュメモリパッケージにより提供される記憶領域は、一つの以上の論理デバイスから構成される実データを記憶する領域である第1の領域と、前記ホスト計算機のライト命令を保管する領域である第2の領域とを前記一つ以上のフラッシュメモリパッケージ毎に有し、前記ホスト計算機からのライト命令の頻度を監視し、前記ライト命令の頻度に応じて前記第2の領域のサイズを変更するものである。

20

【0010】

この構成により、ホスト計算機からのライト命令に応じてライト領域を効率的に割り当てることが可能となる。なお、記憶領域単位をアプリケーション単位又はホスト計算機単位で設定することにより、アプリケーション単位又はホスト計算機単位で第2の領域を変更することが可能となる。

【0011】

また、ストレージ装置において、フラッシュメモリパッケージ全体を記憶領域から削減することにより第2の領域を削減する場合、その削減したフラッシュメモリパッケージの位置を表示することにより、不要となったフラッシュメモリパッケージを減設することができる。

30

【発明の効果】

【0012】

本発明によれば、ホスト計算機からのライト命令に応じてライト領域を効率的に割り当てることが可能となる。

【発明を実施するための最良の形態】

【0013】

以下に、図面を参照しながら本発明の実施形態を説明する。尚これにより本発明が限定されるものではない。

40

【0014】

(第1の実施形態)

先ず第1の実施形態について説明する。本第1の実施形態はフラッシュメモリパッケージ(以下、FM-PKとも称する。)上にフラッシュメモリチップを搭載したストレージ装置により、ホストライトI/O(Input/Output)に基づいて、必要なライト領域を、実データを格納するユーザ領域(データ領域とも呼ぶ。)に対して割り当てる場合について説明する。

【0015】

本第1の実施形態には、FM-PK毎に、第1の領域であるユーザ領域、第2の領域で

50

あるライト領域を用意している。この2つの領域を用意するメリットは、フラッシュメモリパッケージの追加と共に、ユーザ領域が増やせることにある。

【0016】

先ず、第1の実施形態における計算機システム1の構成について説明する。図1及び図2は、計算機システム1の構成を示す図である。図3から図9は計算機システム1で使用する管理情報を示す図である。

【0017】

図1は計算機システム1の構成の一例を示す図である。計算機システム1は複数のホスト計算機（以下、ホストとも称する。）100、ストレージ装置200、管理コンソール250を含んでいる。

10

【0018】

ホスト100とストレージ装置200とは、ストレージ装置200に保管するデータが転送できる入出力ネットワークであるSAN（Storage Area Network）を介して接続されている。

【0019】

SAN300は、一般に、SCSI（Small Computer System Interface）プロトコルを通信可能なスイッチ、例えば、ファイバーチャネルプロトコルの場合ファイバーチャネルスイッチ、また、iSCSI（Internet Small Computer System Interface）の場合にはiSCSIが動作可能なイーサネットスイッチ、さらに、ファイバーチャネルオーバーイーサネットの場合はイーサネットスイッチであり、いずれかのスイッチを介して、ホスト100とストレージ装置200とは接続される。本第1の実施形態では、説明の都合上、ファイバーチャネルプロトコルを用いるファイバーチャネルスイッチを用いた場合で説明する。

20

【0020】

ホスト100は、一般的なサーバを想定しMPU（Micro Processing Unit）101、メモリ（MEM）103、ハードディスク（HDD）102、ネットワークインタフェースサーキット（NIC）103、ホストバスアダプタ（HBA）104を含んで構成されている。

【0021】

NIC103は、TCP/IP（Transmission Control Protocol/ Internet Protocol）によって通信できる、イーサネット（米国Xerox社の登録商標）、IEEE802.11/b/a/g/j/n等の無線LAN（Local Area Network）、FDDI（Fiber-Distributed Data Interface）等の通信制御方式を用いる。本第1の実施形態では説明の都合上、イーサネットの通信制御方式を用いた場合で説明する。

30

【0022】

HBA104は先に記載したSAN300に接続する為に、SAN300のプロトコルに合致した入出力インタフェースを有している。本第1の実施形態の場合は、ファイバーチャネルの入出力インタフェースを有する。

【0023】

ストレージ装置200は、複数のコントローラ210、複数のフラッシュアダプタ220（詳細は後述する。）を含み構成されている。コントローラ210、フラッシュアダプタ220とは、内部ネットワーク230で接続されている。

40

【0024】

内部ネットワーク230は、フラッシュメモリパッケージ内のフラッシュメモリチップにアクセスすることが可能なネットワーク、例えば、PCI-X、PCIエクスプレス等を用いる。この際、インタフェースとしては、ライト時にブロック単位内のページ単位に追記するアクセスを行うインタフェースを内部ネットワーク230上で操作することになる。本第1の実施形態では、フラッシュメモリに対して直接アクセスすることを想定しているが、フラッシュメモリを搭載し、SCSIコマンドを処理できるSSD（Solid State Drive）において、先に示したページ単位に追記、並びに、ブロック単位のアクセスを

50

行なうSSDであれば、ネットワークは上記のネットワークに限定されることは無く、ファイバーチャネルI/F、イーサネット等のSCSIプロトコルが利用できるネットワークを使用しても良い。

【0025】

コントローラ210は、ホスト100から命令されるSCSIコマンドの処理を行う。MPU211、キャッシュメモリ213、NIC214、内部ネットワーク230に接続可能なインタフェースI/F225から構成する。

【0026】

フラッシュメモリアダプタ(FMA)220は、複数のフラッシュメモリパッケージ(FM-PK)221A, 221B, 221C, 221D, …, FM-PKを収納するスロットである。

10

【0027】

FM-PK221A, …, は、先に説明した内部ネットワーク230を介して接続される。FM-PKは先に示したように複数搭載可能で、本第1の実施形態では、“#”の後に1から始まる自然数で、個々のパッケージを指定できるようにしている。

【0028】

FM-PK221上にはフラッシュメモリチップ(図示を省略する。)が搭載され、ホスト100から格納されるデータを保管する。フラッシュメモリチップのブロック領域内は、読み出しは512バイト単位に行え、書き込みは、ブロックを消去した後、ページ単位で追記して行える。なお、本第1の実施形態では、ブロックの単位は256KB、ページの単位は2KBを一例として考える。なお、本実施例では、512バイトの読み出しを行うことを前提に記載したが、512バイトのデータに対する保障コードとして8バイトを別ブロックに保管し管理する形態も考えられる。

20

【0029】

管理コンソール250は、管理LAN(Local Area Network)240経由で、ストレージ装置200に接続される。管理コンソール250は、ストレージ装置200の設定や監視を行う。

【0030】

図2は、計算機システム1の論理構成例を示す図である。ホスト100は、アプリケーション毎に提供されるリクレーション割合設定部121を有している。このリクレーション割合設定部121は、ストレージ装置200にリクレーションの1日のうちの頻度の範囲(上限値、下限値)を提供する。これらの値は、アプリケーションベンダがソフトウェアパッケージのソフトウェア中にデフォルト値として提供してもよいし、管理者がファイル、もしくは、GUI(Graphical User Interface)等の管理インタフェースを用いて設定しアプリケーションがSAN230, LAN240経由の情報伝達手段を用いて通知してもよい。

30

【0031】

リクレーションの頻度を示す頻度情報(前記頻度の範囲(上限値、下限値))の伝達は、SAN230経由でもよいし、LAN240経由でもよい。SAN230経由の場合、例えば、日立製作所が提供しているSAN経由でストレージの制御を行うレイドマネージャ(Raid Manager)に1日のうちのリクレーション頻度の範囲(上限値、下限値)を、0~100%で指定される割合で設定する。このレイドマネージャでは、ストレージ装置200のコマンドデバイスを設定し、そのコマンドデバイスの特定の領域をリード処理・ライト処理することで、ホスト100とストレージ装置200の通信が可能となっている。LAN240経由の場合、TCP/IP経由で接続し、ストレージ装置200のNIC214を通じてホスト100とストレージ装置200間の通信をする。

40

【0032】

ストレージ装置200は、ホスト100から送信されるSCSIコマンドを受領し、FM-PK上のフラッシュメモリチップに対してアクセス(リード、ライト)を行う。主な操作は、コントローラ210上のI/O処理部262にて行われ、I/O処理部262は

50

ポートーLU (Logical Unit) マッピング部258、LUーLDEVマッピング部257、LDEVーRG (RAID GROUP) マッピング部256、RGーFMマッピング部255、フラッシュメモリパッケージ (FMーPK) 管理部251と連係して動作する。

【0033】

ポートーLU (Logical Unit) マッピング部258は、ホスト100からのポートに対してのI/OのSCSIのリードやライトコマンドに関して、IO処理部が、SCSIコマンドで指定された適切なLUを選択する為のLUリストを保管している。LUーLDEVマッピング部257は、I/O処理部が、ポートーLUマッピング部で選択されたLUに対対するLDEV (詳細は後述する。) の管理を行う。LDEVーRGマッピング部256は、LDEVに対応するRAID (Redundant Arrays of Inexpensive (もしくはIndependent) Disks) 構成から構成されるフラッシュメモリパッケージのユーザ領域に対するマッピング関係を保管とその処理を行う。RGーFMマッピング部255は、RAID構成とフラッシュメモリ (FM) パッケージとのマッピング関係を保管とその処理を行う。

10

【0034】

FMーPK管理部251は、ユーザデータを保管するユーザ領域252、ライトデータを一時的に保管するライト領域253、フラッシュメモリチップ上の障害ブロックを管理する代替領域254を含み構成する。

【0035】

ライト領域253の領域は、ホスト100のライト性能の変化、もしくは、先に説明を行ったリクラメーション割合設定部121で指定された割合によって、同一RAIDグループ内のLDEV毎に変化する。

20

【0036】

LDEVーライト領域管理部261は、個々のLDEVのライト性能、もしくは、リクラメーション割合設定部121で指定された値を元にライト領域の大きさを管理する。

【0037】

次に個々の管理テーブルの詳細を説明する。図3は、FMーアドレステーブル300の例を示している。FMーアドレステーブル300は、FMーPK番号フィールド301、開始アドレス位置フィールド302、終了アドレス位置フィールド303、ブロック (BLK) サイズフィールド304、ページサイズフィールド305を有している。

【0038】

30

FMーPK番号フィールド301は、FMーPK番号を保存する。開始アドレス位置フィールド302は、FMーPK番号フィールド301に指定されるFMーPK番号毎に搭載されている複数のフラッシュメモリチップから構成される記憶領域の開始アドレス位置を保存する。終了アドレス位置フィールド303は、その記憶領域の終了アドレス位置を保存する。ブロック (BLK) サイズフィールド304は、フラッシュメモリチップのブロックサイズを保存する。ページサイズフィールド305は、ページサイズを保存する。

【0039】

本第1の実施形態では、開始アドレス位置、終了アドレス位置は、64ビットの16進のアドレス空間で定義する。また、本第1の実施形態で用いるフラッシュメモリチップは、例えば、MLC (Multi-level cell) であり、データの書き込み時には、ブロック内のアクセスはアドレスの開始位置から終了位置の一方に、ページ単位で2KB単位の書き込みをする。データの読み込みは512KB単位で行う。

40

【0040】

図4は、RGーFMマッピング部255で用いるFMーPKを用いたRAID構成テーブル400の例を示している。RAID構成テーブル400は、RG番号フィールド401、FMーPKフィールド402、RAID構成フィールド403を有している。

【0041】

RG番号フィールド401は、RAIDグループの番号を保存する。FMーPKフィールド402は、RAIDグループを構成する複数のFMーPK番号を保存する。RAID

50

構成フィールド403は、RAIDグループのRAID構成を保存する。

【0042】

例えば、RG（RAIDグループ）1は、FM-PKが#1～#4で構成され、RAID5の分散パリティを用いて保護されている。この分散パリティの生成の仕方は、ストレージ装置200が管理しているキャッシュメモリ213の単位でもよいし、フラッシュメモリチップのブロックサイズ等の値でもよい。本第1の実施形態では、フラッシュメモリチップのブロックサイズを用いることにする。また、FM-PKの管理として、複数のFM-PKを連結し、連結されたFM-PK間でRAID構成する形態をとってもよい。この一例として示すRG2では、FM-PK#5とFM-PK#9の間が連結(Concatination)しており、また同様にFM-PK#6、#7、#8がFM-PK#10、#11、#12と連結しており、これらによりRAID5（3Data + 1Parity）が構成されている。

10

【0043】

図5は、FM-PK管理部251で行われる領域管理で利用する、FM-PKに対するユーザ・ライト・代替ブロック領域テーブル500の例を示している。ユーザ・ライト・代替ブロック領域テーブル500は、FM-PK番号フィールド501、ユーザ領域フィールド502、ライト領域フィールド503、代替領域フィールド504を有している。

【0044】

FM-PK番号フィールド501は、FM-PK番号を保存する。ユーザ領域フィールド502は、ユーザの記憶領域として利用するユーザ領域の開始位置のアドレスから終了位置のアドレスを保存する。ライト領域フィールド503は、ライトを一次受けしページ単位で記録するライト領域の開始位置のアドレスから終了位置のアドレスを保存する。代替領域フィールド504は、ユーザ領域、もしくは、ライト領域で故障したブロックを代替する代替領域の開始位置のアドレスから終了位置のアドレスを保存する。

20

【0045】

本第1の実施の形態では、FM-PKのユーザ領域、ライト領域、代替領域を64ビットの16進のアドレス空間として管理するが、実際のアクセスは、フラッシュメモリチップのブロックが指定され、読み込みは512バイト単位、ライトは2Kバイト単位で行われる。

【0046】

また、後述するフラッシュメモリチップへのライト性能の向上のために、FM-PKを導入する際、ライト領域のみの構成も可能である。この場合は、ユーザ領域を設けない構成となるので、一例として、FM-PK#9のユーザ領域に示すように、“-”を用いてユーザ領域を設定しないようにする。

30

【0047】

図6Aは、FM-PK管理部251によって管理された、FM-PK群の領域のイメージを示す図である。例として、図5のRG#1の構成を示している。

【0048】

FM-PK管理部251は、RG#1をFM-PK#1、FM-PK#2、FM-PK#3、FM-PK#4によりRAID5（3データと1パリティ構成）によるデータ保護を行い、一つの記憶空間を作成している。FM-PK管理部251は、RAIDにより一つの記憶空間を作成した後、その中に、ユーザ領域610、ライト領域620、代替領域630を構成する。

40

【0049】

その後、ユーザ領域610は、LDEV-RGマッピング部256により、個々のLDEVに対して記憶領域が分割される。作成されたLDEVに対しては、ライト領域が割り当てられる。ライト領域の割り当ては、LDEV-ライト領域管理部261が行い、図6Bに示すように、LDEV1用ライト領域610A、LDEV2用ライト領域610B、…、LDEVN用ライト領域610Nなど、個々のLDEVに対して割り当てを行う。この割当動作の詳細については後述する。

【0050】

50

図7はLDEV-RGマッピング部256が使用する論理デバイス(LDEV)の構成と性能情報を示すテーブル700の例を示している。テーブル700は、LDEV番号フィールド701、RG番号フィールド702、開始アドレスフィールド703、容量フィールド704、ライト性能フィールド705を有している。

【0051】

LDEV番号フィールド701は、0から始まる論理デバイスの番号を保存する。RG番号フィールド702は、RAIDグループの番号を保存する。開始アドレスフィールド703は、記憶空間の開始アドレスを保存する。容量フィールド704は、記憶空間の容量を保存する。ライト性能フィールド705は、1秒あたりのライト数であるライト性能(Input/Output Per Second)を保存する。

10

【0052】

つまり、LDEV-RGマッピング部256は、論理デバイスを0から始まる論理デバイス番号で指定し、その記憶空間をRAIDグループ番号と対応する開始アドレスとLDEVの容量で構成する。なお、開始アドレスは、RAIDグループのアドレス空間に一致しており、このアドレスはブロック単位で開始される。また、容量に関しても同様に、ブロック単位で増加する。I/O処理部262は、個々のLDEVのライト性能をモニタリングしており、その値をライト性能フィールド705に保存する。

【0053】

図8は、LDEV-ライト領域管理部261が使用する論理デバイスに対するライト領域の割当て済みページ数と使用済みページ数を示すテーブル800の例を示している。テーブル800は、LDEV番号フィールド801、RG番号フィールド802、開始アドレスフィールド803、割当てページ数フィールド804、使用済みページ数フィールド805、リクレーション頻度フィールド806を有している。

20

【0054】

LDEV番号フィールド801は、LDEV番号を保存する。RG番号フィールド802は、RAIDグループ番号を保存する。開始アドレスフィールド803は、LDEVの開始アドレスを保存する。割当てページ数フィールド804は、LDEVに割り当てられたページ数を保存する。使用済みページ数フィールド805は、割当てられたページのうち使用されたページのページ数を保存する。リクレーション頻度フィールド806は、リクレーションの頻度、例えば、図8に示すように、1日に何回行なわれたかを保存する。

30

【0055】

LDEV-ライト領域管理部261は、LDEV番号に対するライト領域を開始アドレス、割当てページ数によって構成する。また、LDEV-ライト領域管理部261は、割当てられたライト領域の管理を行っており、使用済みページ数、並びに、日々のリクレーション回数をモニタリングしており、それぞれ、使用済みページ数フィールド804、リクレーション頻度フィールド805に保管している。本第1の実施形態では、ライト領域の終了アドレスを開始アドレスと割り当て済みページ数のページ単位から計算される値で利用することになるが、便宜先の計算した終了アドレスを本テーブル800に保存してもよい。また、ライト領域の割り当ての初期値は、ライトI/Oが無いことから、LDEVを割当ての際に、LDEV容量の決められた割合、例えばLDEV容量の10%を追加されるLDEVにライト領域として割当ててることを決めてもよい。

40

【0056】

図9は、ポート-LUマッピング部258、LU-LDEVマッピング部257が使用するLUポート、LU、LDEV間のマッピングテーブル900の例を示している。マッピングテーブル900は、WWNフィールド901、ポート番号フィールド902、LUNフィールド903、LDEV番号フィールド904を有している。

【0057】

WWNフィールド901は、ストレージ装置200が有するワールド・ワイド・ネームを保存する。ポート番号フィールド902は、ポート番号を保存する。LUNフィールド

50

903は、ロジカルユニットナンバ（LUN）を保存する。LDEV番号フィールド904は、LDEV番号を保存する。

【0058】

なお、ポート-LUマッピングテーブル900は、ワールド・ワイド・ネームとポート番号をストレージ装置200の出荷時に対応付けてある。管理者は、管理コンソール250経由で、ポート-LUマッピング部258へ、指定するポートに対してLUを作成し、LUN（Logical Unit Number）を割り当てる同時にLDEV番号をLU-LDEVマッピング部257へ、LDEV番号を割当てる。

【0059】

図10は、LDEV-ライト領域管理部261が使用するライト領域上の書き込みデータの保管先を示すテーブル1000の例を示している。テーブル1000は、ブロック番号フィールド1010、ページ番号フィールド1011、対応ブロック番号フィールド1012、サイズフィールド1013を有している。

10

【0060】

ブロック番号フィールド1010は、ブロック番号を保存する。ページ番号フィールド1011は、ページ番号を保存する。対応ブロック番号フィールド1012は、ブロック番号と対応する対応ブロック番号を保存する。サイズ（バイト）フィールド1013は、格納するデータのサイズを保存する。

【0061】

なお、本テーブル1000は、FM-PK番号1004で指定されるFM-PK毎に管理情報を保持し、LDEV毎にページの書き込み先を保存している。

20

【0062】

I/O処理部262は、ホスト100のライトI/O受領時に、書き込み対象となるFM-PKの番号1004中のLDEV番号の記憶領域に対して、次の書き込み領域を示すポインタブロック番号（ポインタBLK）1002、ポインタページ番号（Page）1003の2つから示される領域にデータの書き込みを行い、その管理情報として、テーブル1000のブロック番号（BLK#）1010、ページ番号（Page#）1011、ユーザ領域中の対応フラッシュメモリブロック（BLK番号）、そして、格納したデータのサイズ1013を保管する。なお、“Page”は、ページ単位で増加する。

【0063】

30

以上で、本第1の実施形態における計算機システム1の構成と計算機システム1で利用する管理情報の説明を終える。

【0064】

次に、計算機システム1における動作について説明する。具体的には、FM-PKを導入後、各種領域の割り当てを行い、割当てられた領域に対するホスト100からのライトアクセスの動作について説明する。また、その際に必要となる管理の動作についても説明する。

【0065】

図11は、コントローラ210がFM-PKに対して、RAID構成、並びに、各種領域を割当てる手順を示すフローチャートである。以下にその手順を説明する。

40

【0066】

ストレージ装置200に対して、FM-PKが追加された後、管理コンソール250で後述の図17に示される管理GUIを用いて管理者により、RAIDタイプ1710と各種領域割合（ユーザ領域1733、ライト領域1734、代替領域1735）1730が選択され、必要なFM-PK1720から選択、追加ボタン1753が押下され、領域管理部にRAIDグループが追加される。ステップS1101において、コントローラ210は、管理者によってなされた設定を取得する。

【0067】

ステップS1102において、コントローラ210は、追加されたFM-PKは、各種領域割合に従い、FM-PKの終了位置のアドレスから決められる終了位置を100%と

50

して、ユーザ領域、ライト領域、交代領域で指定された割合を元に、領域を分割する。FM-PKの終了位置が異なる場合、FM-PK間で最も小さい記憶領域に合わせて、終了位置をFM-PK間で共通にしてもよい。この場合、残領域が生じるが、それを代替領域用に用いてもよい。領域分割の手順に戻るが、領域の分割の際には、フラッシュメモリチップで規定されるブロックサイズフィールド304のブロックサイズが、割合で指定された領域内に収まるように終了位置が決められる。ユーザ領域フィールド502に続く、ライト領域503は、ユーザ領域で指定された割合の領域から0番目のフラッシュメモリチップのブロック、すなわち、ユーザ領域の終了位置が割合計算で、128KBとして、フラッシュメモリチップのブロックサイズ(256KB)内に重なった場合には、重なりがあるブロックの開始アドレスからスタートする。代替領域1735はライト領域1734と同様にして考える。なお、この配置は一例であり、管理情報を含め、領域間の区切りを明確にするために、空き領域を意図的に用意してもよい。

10

【0068】

ステップS1103において、コントローラ210は、選択されたFM-PKに対して指定されたRAID構成を作成する。つまり、図4で示すように、RAIDグループ番号フィールド401のRAID番号、FM-PKパッケージ番号フィールド402のFM-PK番号、RAID構成フィールド403のRAID構成が共にテーブル400に格納される。この設定以降、指定されたRAID構成により、ストレージ装置200は動作する。

【0069】

20

ステップS1104において、コントローラ210は、図7を用いて説明したLDV-RGマッピングテーブル700を参照し、作成したRAIDグループ番号に対応する領域にLDVが生成されているか否かを調べる。もし、生成されている場合は(S1104:YES)、処理はステップ1005に進み、生成されていない場合には(S1104:NO)、処理は終了する。

【0070】

ステップS1105において、コントローラ210は、リクラメーション処理を実施する。なお、リクラメーション処理の詳細については、後述する。以上でこの処理を終了する。

【0071】

30

次に、I/O処理について説明する。図12は、コントローラ210内のI/O処理部262が行うライト処理である。以下にその手順を説明する。

【0072】

ステップS1201において、I/O処理部262は、ホスト100から受けたSCSIのライトI/Oのトランスファーレングスのデータサイズを確認し、フラッシュメモリのブロックサイズで割り切れないことを確認する。割り切れないときにはステップ1202に進み、割り切れる場合はステップS1211に進む。

【0073】

ステップS1202において、I/O処理部262は、ホスト100から受けたライトI/Oのデータサイズがページサイズ未満であることを確認する。ページサイズ未満であるのであれば(S1202:YES)、処理はステップS1203に進み、ページサイズと同じサイズであれば(S1202:NO)、処理はステップS1206に進む。

40

【0074】

ステップS1203において、I/O処理部262は、対象LDVに対する残ページ数を確認する。ページが残っているのであれば(S1203:YES)、処理はステップS1205に進み、残っていなければ(S1203:NO)、処理はステップS1204に進む。

【0075】

ステップS1204において、I/O処理部262は、図14のリクラメーション処理を行う。

50

【0076】

ステップS1205において、I/O処理部262は、ホスト100から指定された開始アドレスと容量をデータが格納されているFM-PKのライト領域から2KB分のデータをキャッシュメモリ213にリードする。もしデータが無ければ、ユーザ領域からデータを読み出す。同様に、パリティデータについても、2KB分のデータをキャッシュメモリ213にリードする。

【0077】

ステップS1206において、I/O処理部262は、ステップS1205でリードしたキャッシュメモリ213上のデータとパリティデータをXORすることにより、キャッシュメモリ213上のデータを除いたパリティデータを生成する。

10

【0078】

ステップS1207において、I/O処理部262は、キャッシュされた2KBデータに対してホストデータを上書きする。

【0079】

ステップS1208において、I/O処理部262は、キャッシュメモリ213上のデータとステップS1206で生成したパリティデータをXORすることにより、新パリティデータを生成する。

【0080】

ステップS1209において、I/O処理部262は、ステップS1206で更新したデータを元のFM-PKのテーブル1000上のポインタブロック番号1002とポインタページ番号1003で示されるページに書き出す。そしてI/O処理部262は、図10のテーブル1000のテーブルのポインタブロック番号1002、ポインタページ番号1003に該当するブロック番号1010、ページ番号1011列に対応する対応ブロック番号フィールド1012、サイズフィールド1013を更新する。ステップS1208で生成されたパリティデータも同様に、読み出したFM-PKのテーブル1000上のポインタブロック番号1002とポインタページ番号1003で示されるページにデータを書き出し、フラッシュメモリチップ上の書き出すユーザ領域のブロック番号を対応ブロック番号フィールド1012に、ページに書き出した量をサイズフィールド1013に更新する。

20

【0081】

ステップS1210において、I/O処理部262は、ポインタページ番号1003をインクリメントし、次のブロックに行くのであれば、ブロックもインクリメントする。

30

【0082】

一方、ステップS1211において、I/O処理部262は、ホスト100から指定された開始アドレスと容量をデータが格納されているFM-PKのユーザ領域から1ブロック分のデータをキャッシュメモリ213にリードする。

【0083】

ステップS1212において、I/O処理部262は、ステップS1211で読み込まれたキャッシュメモリ213上のデータとパリティデータをXORすることにより、キャッシュメモリ213上のデータを除いたパリティデータを生成する。

40

【0084】

ステップS1213において、I/O処理部262は、キャッシュメモリ213上データとステップS1206で生成したパリティデータをXORすることにより、新パリティデータを生成する。

【0085】

ステップS1214において、I/O処理部262は書き込むブロックのフラッシュメモリチップのブロックを消去する。

【0086】

ステップS1215において、I/O処理部262はFM-BLKにデータ、パリティブロックに書き込む。以上で、ライト処理を終了する。

50

【0087】

図13は、コントローラ210内のI/O処理部262が行うリード処理である。以下にその手順を説明する。

【0088】

ステップS1301において、I/O処理部262がホスト100のSCSIのリードコマンドで指定された開始アドレスと指定されたサイズのデータに対応するLDEVにおけるフラッシュメモリチップのブロックをキャッシュメモリ213上に読み込む。

【0089】

ステップS1302において、I/O処理部262が、テーブル1000の各LDEVのライト領域を参照し、ホスト100のSCSIのリードコマンドで指定された領域（開始位置、サイズ）で指定されたブロック内に更新データがないか、対応BLK番号フィールド1012、サイズフィールド1013を参照して、判定する。データが有る場合（S1302：YES）、処理はステップS1303に進む。データがない場合（S1302：NO）、処理はステップS1304に進む。

10

【0090】

ステップS1303において、I/O処理部262は、ライト領域からキャッシュメモリ213にデータをリードし、既にリードされたブロックのデータに対して上書きする。

【0091】

ステップS1304において、I/O処理部262は、キャッシュメモリ213上の上書き済みのデータを戻り値としてホスト100に戻す。以上でリード処理を終了する。

20

【0092】

次に、ライト時や管理者がLDEVの容量を変更したときに行うリクラメーション処理について説明を行う。図14は、I/O処理部262が行うリクラメーション処理のフローチャートを示している。以下にその手順を説明する。

【0093】

ステップS1401において、I/O処理部262は、該当LDEVのフラッシュメモリチップのライト領域の追記を抑止する。

【0094】

ステップS1402において、I/O処理部262は、テーブル1000を参照し、全ライト領域内のページデータを対応BLK番号フィールド1012とサイズフィールド1013を元に、各LDEVのユーザ領域に上書きする。また、全ページが書き出された段階で、開始アドレスフィールド802の開始アドレスに該当するブロックをポインタBLKとして設定し、ページを“0”としてリセットする。

30

【0095】

ステップS1403において、I/O処理部262は、LDEV間のライト割合が変更されたか否かを判定する。その割合が変更されている場合は（S1403：YES）、処理はステップS1402に進み、割合が変更されていない場合は（S1403：NO）、処理はステップS1406に進む。

【0096】

ステップS1404において、I/O処理部262は、LDEV間のライト割合から、ライト領域の割り当て領域の開始位置と割り当てページ数を再計算する。

40

【0097】

ステップS1405において、I/O処理部262は、再計算した値を、図8の開始アドレスフィールド802の開始アドレスと割り当てページ数フィールド803の割り当てページ数に代入する。

【0098】

ステップS1406において、I/O処理部262は、該当LDEVのフラッシュメモリチップのライト領域の追記を再開する。以上でリクラメーション処理を終了する。

【0099】

ステップS1402でI/O処理部262がLDEVのライト割合の変更を確認したが

50

、その変更を行う為のGUIを図15に示す。図15は、論理デバイス（LDEV）に対するライト領域の管理GUI 1500の例を示す図である。

【0100】

管理GUI 1500は、RAIDグループ番号（RG#）を表示するフィールド1501、割り当て対象となるLDEVを表示する1502、現在の割り当て率（割合）を表示するフィールド1503、推奨される割り当て率（推奨割合）を表示するフィールド1504、割り当て率1503に対するアラートを表示するフィールド1511、日々のリクレーション回数を表示するフィールド1505、1日における最大リクレーション回数に対するリクレーション割合を示すフィールド1506、アプリケーション（AP）個別設定割合を表示するフィールド1507、アラートを表示するフィールド1508から構成される。

10

【0101】

現在の割り当て率と推奨割合の間で差が大きい場合、例えば±10%を超える場合、管理コンソール250は、SNMP（Simple Network Management Protocol）、メール、SMS（Short Message Service）等の通信手段で管理者に通知し、割り当て率1503を変更することを求める。

【0102】

LDEV番号に対して、推奨割合に従って、管理者が、現在の割当率を定義する。推奨割合は、図4の各LDEVの平均iopsの合計のうち各LDEVのiopsの割合をパーセンテージ表記する。現在の割り当て率は、RAIDグループ内のLDEV全体で100%になるように設定する。本第1の実施形態では、割合を管理者が入力しているが、推奨値をそのまま当てはめ、リアルタイムに割り当てを変えてもよい。

20

【0103】

リクレーション回数1505は、図8のリクレーション頻度フィールド806から参照される。リクレーション割合1506は、LDEV毎で1日のうち最大リクレーション可能な回数のうち、日々のリクレーション回数の割合がどの程度になるかをパーセント表記している。この計算の仕方は、次の（1）式から計算される。

（リクレーション割合）＝（リクレーション回数）／（24時間／（個々のLDEVに対する全ライト領域を書き込む時間＋対象ライト領域のページ数分のブロックが消去されたとき時間））…（1）

30

【0104】

AP設定割合フィールド1507は、図2のホスト100上に用意されるリクレーション割合設定部211が、下限値、上限値を設定した場合に表示される。何も設定しない場合、“－”として表記され、その場合には、システムデフォルト値である、リクレーション閾値（下限値）1509、リクレーション閾値（上限値）1510を使用することになる。もし、この閾値を超える場合、アラート1508に表示され、アラートボックスをクリックすると、図17の領域管理GUI（後述）が表示される。

【0105】

次に、リクレーション割合の閾値を超えたときの動作を説明する。リクレーション割合の上限値を超えた場合、リクレーション回数が頻度が高く、性能影響があると判断し、FM-PKを増設し、ライト領域を増やす必要がある。一方で、下限を下回った場合、余分なライト領域があるということなので、ライト領域を削減し、開放する。開放した場合、その領域を他の用途、ユーザ領域や代替ブロック用として使ってもよい。

40

【0106】

図16は、管理コンソール250が、リクレーション回数からライト領域の増減設をアラートする処理を示すフローチャートである。以下にその手順を説明する。

【0107】

ステップS1601において、管理コンソール250は、リクレーション割合が、上限値を超えてないか確認する。上限値は、AP設定割合フィールド1507に上限値が定義されている場合その値を用い、定義されていない場合はリクレーション閾値（上限値

50

部1510のリクラメーション閾値を使用する。上限値を超えている場合（S1601：YES）、処理はステップ1602に進み、設定値を超えていない場合は（S1601：NO）、処理はステップ1603に進む。

【0108】

ステップS1602において、管理コンソール250は、管理者に増設要求をSNMP、メール、SMS等の通信手段で管理者に通知する。この際、GUIには、増設を示すボタンがアラート1508に表示される。

【0109】

ステップ1603において、管理コンソール250は、リクラメーション割合が、下限値を超えていないか確認する。下限値は、AP設定割合部1507の下限値が定義されている場合その値を用いて、定義されていない場合はリクラメーション閾値（下限値）1509を使用する。下限値を下回っている場合、処理はステップS1603に進み、設定値を下回っていない場合は、処理を終了する。

【0110】

なお、アラート処理は、日々、数時間毎など定期的に起動され、計算機システム1の正常な稼動を確認する。また、リクラメーション動作時に確認してもよい。

【0111】

図17は、管理コンソール250で表示されるGUIを示す図である。

【0112】

領域管理の画面1700は、RAIDグループ番号（RG#）を表示するフィールド1701、FM-PK番号を表示するフィールド1702、ユーザ領域を表示するフィールド1703、ライト領域を表示するフィールド1704、代替領域を表示するフィールド1705を示している。先に示したように、もし、FM-PK間で容量の差異が有る場合、本第1の実施形態では、容量が最小な物にあわせてアドレス領域を管理する。

【0113】

また、RAIDグループ作成時には、RG作成ボタン1751を押すと、新規のRG番号が作成される。RAIDグループを削除する場合には、各RG番号のRG削除ボタン1752を押すことによって削除する。RAIDグループに対して、FM-PKを割当てする場合、希望するRAIDType1710を選択することで必要なFM-PK（例えば、RAID5で3D+1Pの場合、4つのFM-PKが選択される）が空きFM-PK1720リストから選択チェックボックス1721で選択される。

【0114】

空きFM-PKのリストは、選択されるFM-PKを表示するフィールド1721、FM-PK番号（FM-PK#）を表示するフィールド1722、FM-PKの開始アドレス位置を表示するフィールド1723、終了アドレス位置を表示するフィールド1724から構成される。

【0115】

領域割合1730のスライドバー1731（ユーザ領域とライト領域を区切る）、1732（ライト領域と代替領域を区切る）で選択後、管理者が追加ボタン1753を押すことで、チェックボックス1701で選択したRAIDグループに追加される。

【0116】

なお、スライドバー1731と1732を重ね、ライト領域を0%にした場合、管理コンソール250は、エラーとして処理を戻す。理由は、ライト領域を0%とした場合、ライト領域を保管する領域がなくなるためである。もし、0%を実現する場合、ライト時の動作がブロック単位のリード・モディファイ・ライトにすることでデータにアクセスは可能だが、ライト時の動作はブロック単位でキャッシュメモリ213上にリード後、ホスト100からのライトデータでキャッシュメモリ213上のデータを上書き後、フラッシュメモリチップの元ブロックを消去、データをライトすることになる。

【0117】

次に、図16のステップ1603で管理者に通知したライト領域開放要求時の手順と動

10

20

30

40

50

作を示す。管理者が、図17の各FM-PKの列を選択することで、割合1730がその列の割合を表示する。管理者が、スライドバー1731、1732をライト領域1734に関して縮小する方向で動かしたときのライト領域の割合に従って、開放の動作を行う。図18は、管理コンソール250が開放処理を行うときのフローチャートを示している。

【0118】

ステップS1801において管理コンソール250は、削減調整した後のライト領域のアドレス空間内に、未処理分のリクラメーションデータが残っている場合、すなわち、更新後のライト領域が、ブロック番号1002とポインタブロック番号1003間に収まっている場合（S1801：YES）、処理はステップS18043進む。収まっていない場合（S1801：NO）、処理はステップS1802に進む。

10

【0119】

ステップS1802において、管理コンソール250は、リクラメーション処理を実施する。具体的には、図14の処理のうちステップS1402、S1404、S1405を処理する。ステップS1802のリクラメーション処理内でライト停止運用をしているため、ライト領域の追記停止の処理を別に設ける必要はない。

【0120】

ステップ1803において、管理コンソール250は新領域のライト領域を該当するFM-PK番号のライト領域を更新する。各LDEV番号フィールド801も、開始アドレスフィールド803、割当てページ数フィールド804を更新する。以上で、処理を終了する。開放された領域は、空き領域として管理され、ユーザ領域、すなわちLDEV用に用いてもよいし、代替領域として利用してもよい。

20

【0121】

この第1の実施形態による計算機システム1によると、アプリケーションやホスト計算機単位でライト領域を効率的に割り当てることが可能となる。

【0122】

また、リクラメーション時の頻度を上限値の閾値でモニタリングすることで、消去回数が頻度に発生していることから、ライト領域不足を判断し、フラッシュメモリの増設を指示、フラッシュメモ리를増設後、ライト領域を拡大することで消去回数の適正化し性能劣化を抑えることができる。さらには、フラッシュメモリ特有の消去回数から起因するメディアの劣化を、リクラメーション回数を抑えたことで、延命することができる。

30

【0123】

また、リクラメーション時の頻度を下限値の閾値でモニタリングすることで、消去回数が余り発生していない場合、ライト領域過剰として判断し、ライト領域を縮小することで、消去回数の適正化、無駄なライト領域を削減することが可能となる。

【0124】

（第2の実施形態）

次に第2の実施形態について説明する。第1の実施形態は、FM-PK内に、必ずユーザ領域を割り当てる構成を考えたが、本第2の実施形態は、FM-PK毎にユーザ領域、ライト領域とを分けて利用する形態を考える。この構成のメリットは、ユーザ領域とライト領域が分離することによって、ライト性能に応じて、FM-PK単位で追加・削除できる点である。なお、本第2の実施形態では、代替領域は、ユーザ領域のFM-PK、ライト領域のFM-PKそれぞれに準備してあるが、独立したFM-PKに準備してもよい。

40

【0125】

次に計算機システムの構成について説明する。第2の実施形態における計算機システムの構成は、第1の実施形態の計算機システム1と同一の構成については同一の符号を付して説明は省略することとし、以下では第1の実施形態と第2の実施形態との構成及び作用の差異を中心に説明する。この第2の実施形態は、既述の図6で説明したメモリ（FM-PK）の使い方が異なっている。

【0126】

図19は、メモリ（FM-PK）の各領域のイメージを示す図である。第1の実施形態

50

の場合と同様にFM-PK#1～#8で一つのRAIDグループを形成し、FM-PK#1, #5間の連結を行うことで、一つのデバイスとして扱っている。異なる点は、ユーザ領域1901が、FM-PK#1～FM-PK#4の列でまとめられ、ライト領域1903はFM-PK#5～FM-PK#9でまとめられている点である。

【0127】

次に本第2の実施形態における計算機システム1の動作について説明する。第1の実施形態と異なる点は、図20に示す領域管理の画面1700において、各RAIDから構成される列において、ライト列に関して削除ボタン1801が追加されるとともに、割合1730においてユーザ領域及びライト領域がユーザ、もしくは、ライト領域となり、このユーザ、もしくは、ライト領域と代替領域とをスライドバー1804で分けるように構成されている。

10

【0128】

管理者が、削除ボタン1801を押した後、図18で示したライト領域の縮小処理が実施される。この際、新たなライト領域は、残りのFM-PKのうちライト領域を持つFM-PKに限られる。FM-PK内のライト領域が全削除された後、そのFM-PKはフラッシュメモリアダプタ220から抜くことができる。この際、図21のような、フラッシュメモリアダプタ220上のFM-PKスロット2111から外すことができる。各FM-PKに対応するLED2113, 2114は、で挿抜可能なFM-PKであるか否かを示している。LED2113はFM-PKが脱着可能であることを示し、LED2114は脱着不可であることを示している。この表示により管理者は容易に減設できるFM-PKを見つけることができる。

20

【0129】

本第2の実施形態の計算機システム1によると、ユーザ領域とライト領域が分離することによって、ライト性能に応じて、FM-PK単位で追加・削除できる。

【産業上の利用可能性】

【0130】

本発明は、ストレージ装置、その記憶領域管理方法及びフラッシュメモリパッケージに広く適用することができる。

【図面の簡単な説明】

【0131】

30

【図1】本発明に係わる第1の実施形態における計算機システムの構成の一例を示す図である。

【図2】同第1の実施形態における計算機システムの論理構成例を示す図である。

【図3】同第1の実施形態におけるFM-アドレステーブルの例を示す図である。

【図4】同第1の実施形態におけるFM-PKを用いたRAID構成テーブルの例を示す図である。

【図5】同第1の実施形態におけるFM-PKに対するユーザ・ライト・代替BLK領域テーブルの例を示す図である。

【図6A】同第1の実施形態におけるFM-PKを用いたRAID構成の中に構成された各領域のイメージ図である。

40

【図6B】同第1の実施形態における各LDEVに割り当てられたライト領域の例を示す図である。

【図7】同第1の実施形態における論理デバイスの構成と性能情報を示すテーブルの例である。

【図8】同第1の実施形態における論理デバイスに対するライト領域の割当て済みページ数と使用済みページ数、そして、リクラメーション頻度を示すテーブル例である。

【図9】同第1の実施形態におけるポート、LU、LDEV間のマッピングテーブル例である。

【図10】同第1の実施形態におけるLDEVに対するライト領域におけるページ管理表の一例である。

50

【図 1 1】同第 1 の実施形態における FM-PK に対して、RAID 構成、並びに、各種領域を割当てての手順である。

【図 1 2】同第 1 の実施形態におけるライト I/O のフローチャート例である。

【図 1 3】同第 1 の実施形態におけるリード I/O のフローチャート例である。

【図 1 4】同第 1 の実施形態におけるリクラメーション処理のフローチャート例である。

【図 1 5】同第 1 の実施形態における論理デバイスに対するライト領域の管理 GUI の例である。

【図 1 6】同第 1 の実施形態における管理コンソールが、リクラメーション回数からライト領域の増減設をアラートするフローチャート例である。

【図 1 7】同第 1 の実施形態における各種領域を割当ての際に使用する GUI の例である

10

【図 1 8】本発明に係わる第 2 の実施形態におけるライト領域を削減する処理のフローチャートの例である。

【図 1 9】同第 2 の実施形態における FM-PK を用いた RAID 構成の中に構成された各領域のイメージ図である。

【図 2 0】同第 2 の実施形態におけるライト領域の管理 GUI の例である。

【図 2 1】同第 2 の実施形態における FM-PK のスロットと脱着可否を示す LED 例を示した図である。

【符号の説明】

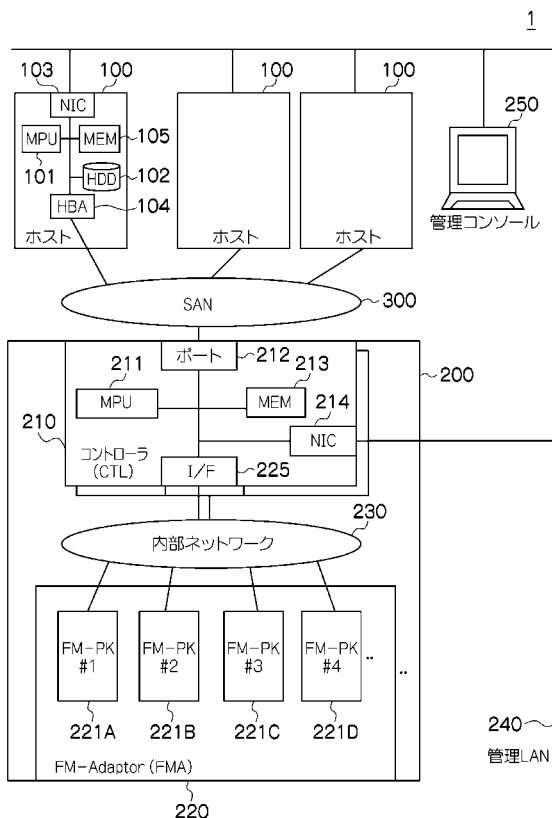
【0132】

20

1…計算機システム、100…ホスト計算機（ホスト）、200…ストレージ装置、コントローラ 210、211…MPU、221…フラッシュメモリパッケージ（FM-PK）、250…管理コンソール、251…FM-PK 管理部、252…ユーザ領域、253…ライト領域、254…代替領域、262…I/O 処理部

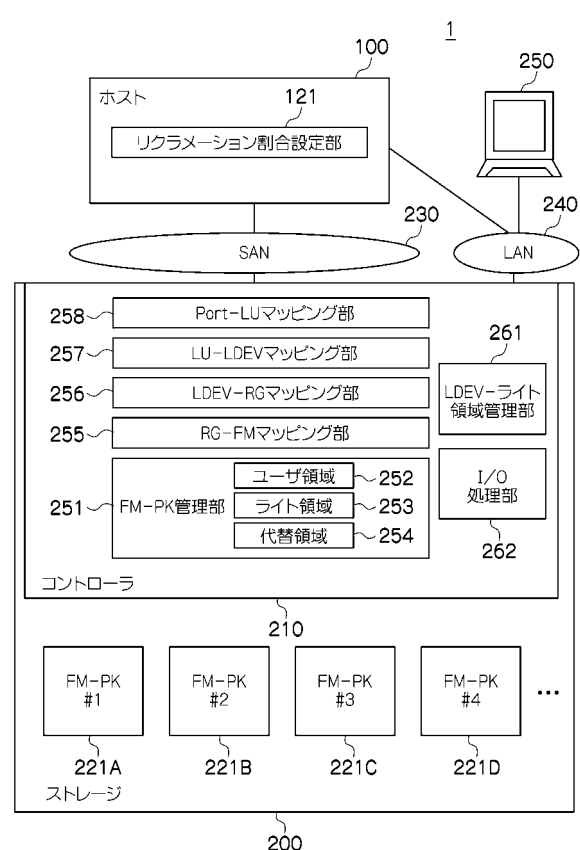
【図 1】

図1



【図 2】

図2



【図 3】

図3

301 FM-PK#	302 開始アドレス 位置	303 終了アドレス 位置	304 BLK サイズ	305 ページ サイズ
1	0x000...00000	0x0004...0000	256KB	2KB
2	0x000...00000	0x0008...0000	256KB	2KB
3	0x000...00000	0x0012...0000	256KB	2KB
...	...	...	...	...

【図 4】

図4

401 RG#	402 FM-PK	403 RAID構成
1	FM-PK #1	RAID5(3D+1P)
	FM-PK #2	
	FM-PK #3	
	FM-PK #4	
2	FM-PK #5, FM-PK #9	RAID5(3D+1P)
	FM-PK #6, FM-PK #10	
	FM-PK #7, FM-PK #11	
	FM-PK #8, FM-PK #12	
...	...	...

【図 5】

図5

501 FM-PK#	502 ユーザ領域 開始位置～ 終了位置	503 ライト領域 開始位置～ 終了位置	504 代替領域 開始位置～ 終了位置
1	0x00 0000 ～ 0x00 9999	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
2	0x00 0000 ～ 0x00 9999	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
3	0x00 0000 ～ 0x00 9999	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
...	...	...	...
9	—	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
10	—	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
11	—	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
12	—	0x0002 0000 ～ 0x0002 9999	0x0003 0000 ～ 0x0004 000
...	...	...	...

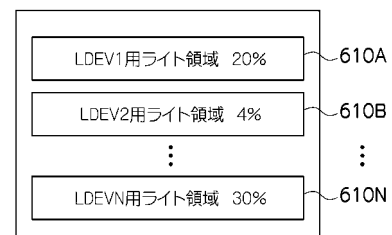
【図 6 A】

図6A



【図 6 B】

図6B



【図 7】

図7

700				
701	702	703	704	705
LDEV#	RG#	開始アドレス	容量	ライト性能 (iops)
1	1	0x00...00	32GB	12131
2	1	0x01...00	16GB	10
3	1	0x02...00	2GB	3232
...	...	...	...	...
10	2	0x02...00	100GB	2312
11	2	0x04...00	20GB	0

【図 8】

図8

800					
801	802	803	804	805	806
LDEV#	RG#	開始 アドレス	割り当て ページ数	使用済み ページ数	リクレーション 頻度 [回/日]
1	1	0x0002... ... 1000	XXXXXX	xxxxxx	P
2	1	0x0002... ... 2000	ZZZZZZ	zzzzzz	Q
3	1	0x0002... ... 3000	YYYYYY	yyyyyy	R
...	...	...	...	...	...
10	2	0x0011... ... 1000	...	...	...
11	2	0x0012... ... 1000	...	...	...
...	...	...	...	...	...

【図 9】

図9

900			
901	902	903	904
WWN	ポート#	LUN	LDEV#
50060E80C93BC394	1	1	1
50060E80C93BC394	1	2	2
50060E80C93BC394	1	3	3
...	...	...	...
50060E80C93BC395	2	1	10
50060E80C93BC395	2	2	11
...	...	...	...

【図 10】

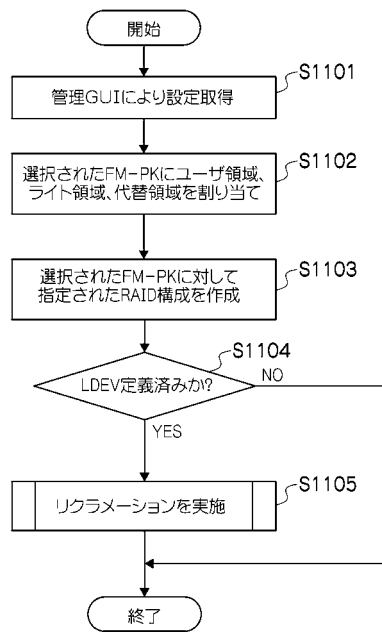
図10

1000			
1010	1011	1012	1013
BLK#	Page#	対応BLK番号	サイズ(B)
11...	1	23231	1024
11...	2	124121	2048
11...	3	235512	512
11...	4	100	1024
...	...	...	...
...	...	...	...
...	...	...	...
...	...	...	...

FM-PK#:1 → 1004
 LDEV#:1 → 1001
 ポインタBLK:1020 → 1002
 Page:20 → 1003

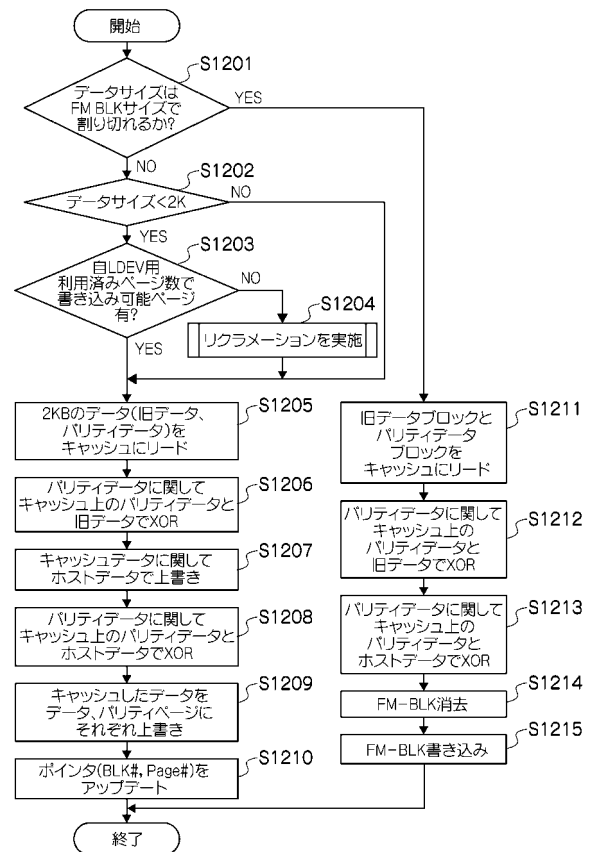
【図 1 1】

図11



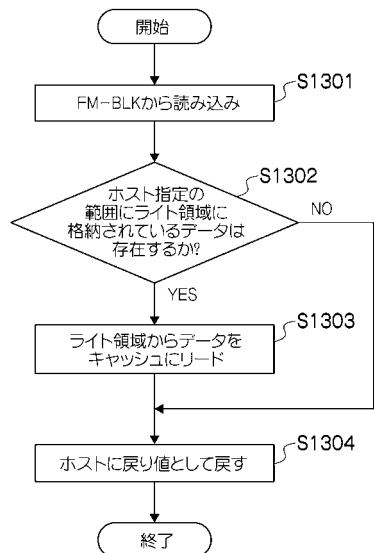
【図 1 2】

図12



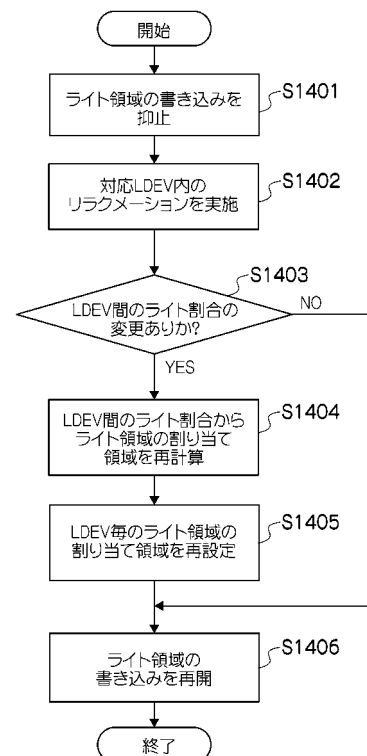
【図 1 3】

図13



【図 1 4】

図14



【図 15】

図15

1500

✕

ライト領域管理

RG#	LDEV #	割り当て 割合%	推奨割合 %	アラート	リクレー ション 回数 (回/日)	リクレー ション 割合%	AP設定 割合 (下限/ 上限)	アラート
1	1	20	30	!!	P	82	10/80	増設要
	2	4	10		Q	10	—	削減要
	5	5	5		R	40	—	
	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

1501 1502 1503 1504 1511 1510

リクレーション閾値(上限値): %

リクレーション閾値(下限値): %

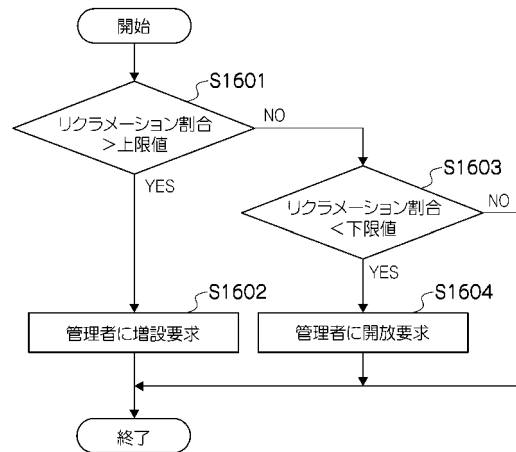
1509

Cancel Apply OK

1511 1512 1513

【図 16】

図16



【図 17】

図17

1700

✕

領域管理

RG#	FM-PK	ユーザ領域	ライト領域	代替領域
1751 ☐ RG作成	FM-PK #1	...	...	...
	FM-PK #2	...	...	...
	FM-PK #3	...	...	...
	FM-PK #4	...	...	...
1752 ☐ RG削除	FM-PK #5	...	...	...
	FM-PK #6	...	...	...
	FM-PK #7	...	...	...
	FM-PK #8	...	...	...
1706 ☐ RG削除	FM-PK #9	...	...	...
	FM-PK #10	...	...	...
	FM-PK #11	...	...	...
	FM-PK #12	...	...	...
...	...	...	...	...

1710 RAID Type 1730 割合 1731 ユーザ領域 30%(1733) 1732 ライト領域 50%(1734) 代替領域 20%(1735)

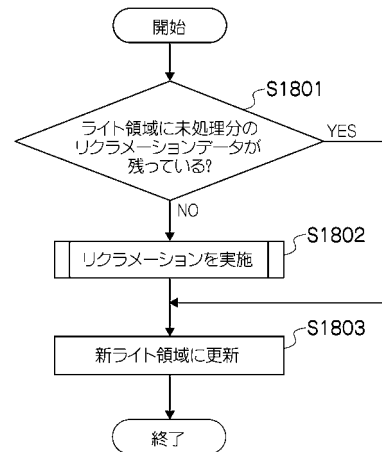
空き PK 1721 1722 1723 1724 1720

選択	FM-PK	開始位置	終了位置
☐	13	0x000000000000	0x000400000000
☐	14	0x000000000000	0x000800000000
☐	15	0x000000000000	0x001200000000
...	...	...	...

1753 追加

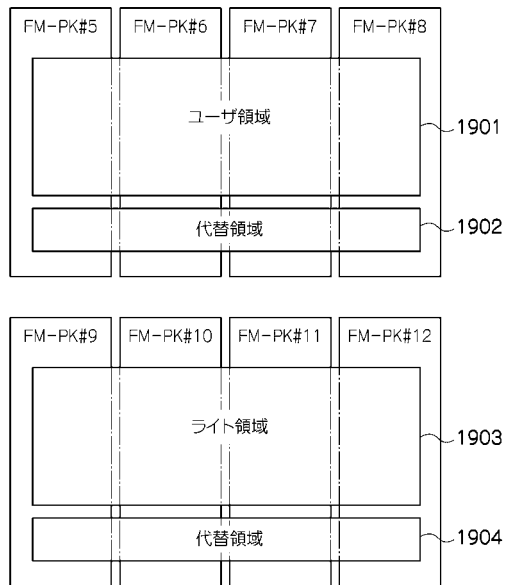
【図 18】

図18



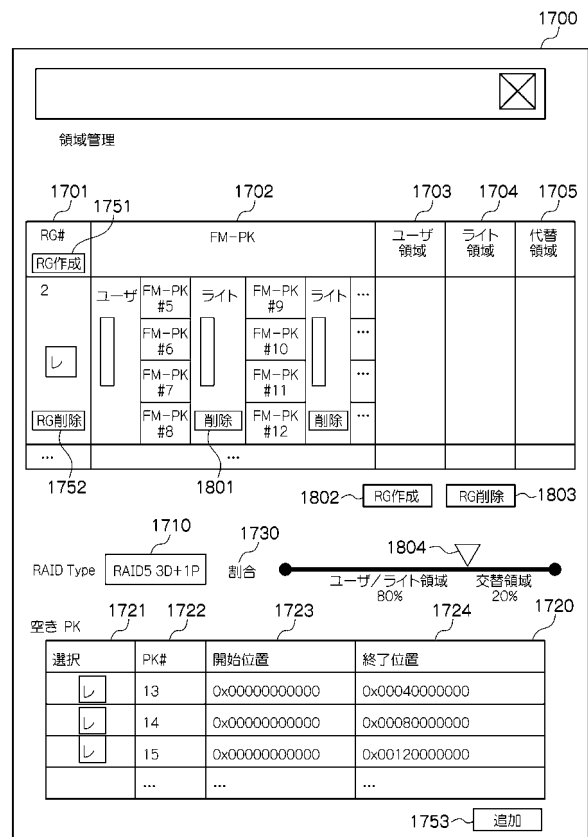
【図 19】

図19



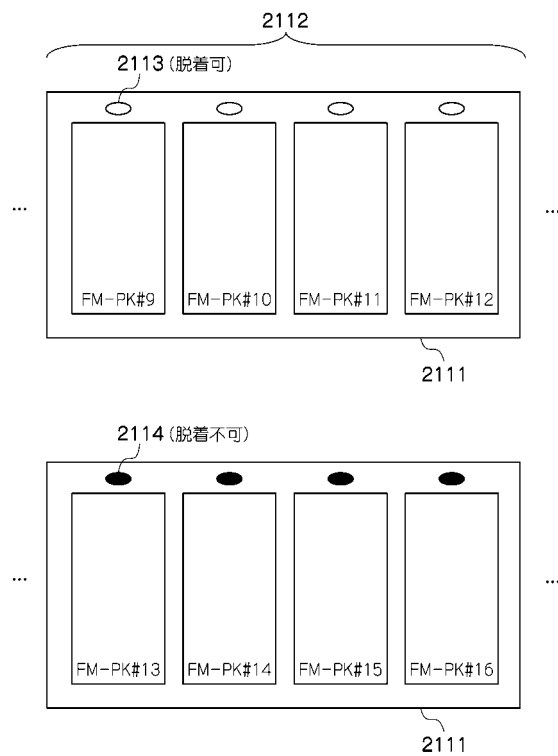
【図 20】

図20



【図 21】

図21



フロントページの続き

- (56)参考文献 米国特許出願公開第2008/0104309 (US, A1)
特開2001-184176 (JP, A)
米国特許出願公開第2002/0097594 (US, A1)
特開平08-235076 (JP, A)
特開2008-003932 (JP, A)
特開2005-242897 (JP, A)
特開2006-318246 (JP, A)
特開2006-252031 (JP, A)
特開2003-162439 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06- 3/08
G06F 12/00-12/06
G06F 12/16
G06F 13/10-13/14