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

特許第5075761号
(P5075761)

(45)発行日　平成24年11月21日(2012.11.21)　(24)登録日　平成24年8月31日(2012.8.31)

(51)Int.Cl.		F I		
G06F 3/06	(2006.01)	G06F	3/06	301Z
G06F 3/08	(2006.01)	G06F	3/08	H
G06F 13/10	(2006.01)	G06F	13/10	340A
G06F 12/16	(2006.01)	G06F	12/16	310A
		G06F	12/16	310J

請求項の数 6　(全 22 頁)　最終頁に続く

(21)出願番号　特願2008-213464 (P2008-213464)
(22)出願日　平成20年8月22日 (2008.8.22)
(65)公開番号　特開2009-301525 (P2009-301525A)
(43)公開日　平成21年12月24日 (2009.12.24)
審査請求日　平成23年2月16日 (2011.2.16)
(31)優先権主張番号　特願2008-126608 (P2008-126608)
(32)優先日　平成20年5月14日 (2008.5.14)
(33)優先権主張国　日本国(JP)

(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　100098660
弁理士　戸田　裕二
(72)発明者　山本　政行
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　山本　彰
東京都千代田区丸の内一丁目６番１号　株式会社日立製作所研究開発本部内

最終頁に続く

(54)【発明の名称】フラッシュメモリを用いたストレージ装置

(57)【特許請求の範囲】
【請求項１】
一つ以上のフラッシュメモリモジュールとキャッシュメモリを有する第一のストレージ装置及び前記第一のストレージ装置に接続されたホスト計算機から構成される計算機システムにおける前記第一のストレージ装置の記憶領域作成方法において、
前記第一のストレージ装置は、
前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを用いた記憶領域の作成要求を受信すると、受信した前記作成要求において、前記記憶領域とコピーペアの関係になる記憶領域が指定されているか否かを判断し、前記記憶領域とコピーペアの関係になる記憶領域が指定されている場合、前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを用いた記憶領域とコピーペアの関係となる記憶領域へのライトアクセス頻度から代替領域容量を定義し、前記ホスト計算機に提供する記憶領域の容量と前記代替領域容量とを合算した容量で前記第一のストレージ装置が有する前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを作成する
ことを特徴とするストレージ装置の記憶領域作成方法。
【請求項２】
請求項１記載のストレージ装置の記憶領域作成方法において、
前記第一のストレージ装置は、
前記ライトアクセス頻度及び前記記憶領域の耐用年数から前記代替領域容量を定義する
ことを特徴とするストレージ装置の記憶領域作成方法。

【請求項３】
請求項１記載のストレージ装置の記憶領域作成方法において、
前記計算機システムは、
前記第一のストレージ装置に接続されて一つ以上のハードディスクドライブを有する第二のストレージ装置を備え、
前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを用いた記憶領域とコピーペアの関係となる記憶領域は、前記第二のストレージ装置の前記一つ以上のハードディスクドライブから構成される記憶領域である
ことを特徴とするストレージ装置の記憶領域作成方法。
【請求項４】
一つ以上のフラッシュメモリモジュールと、キャッシュメモリと、ストレージコントローラとを有し、ホスト計算機に接続される第一のストレージ装置において、
前記第一のストレージ装置の前記ストレージコントローラは、
前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを用いた記憶領域の作成要求を受信すると、受信した前記作成要求において、前記記憶領域とコピーペアの関係になる記憶領域が指定されているか否かを判断し、前記記憶領域とコピーペアの関係になる記憶領域が指定されている場合、前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを用いた記憶領域とコピーペアの関係になる記憶領域へのライトアクセス頻度から代替領域容量を定義し、前記ホスト計算機に提供する記憶領域の容量と、前記代替領域容量とを合算して論理デバイス容量を算出し、前記第一のストレージ装置が有する前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを作成する
ことを特徴とするストレージ装置。
【請求項５】
請求項４記載のストレージ装置において、
前記第一のストレージ装置の前記ストレージコントローラは、
前記ライトアクセス頻度及び前記記憶領域の耐用年数から前記代替領域容量を定義する
ことを特徴とするストレージ装置。
【請求項６】
請求項４記載のストレージ装置において、
一つ以上のハードディスクドライブを有する第二のストレージ装置に接続され、
前記一つ以上のフラッシュメモリモジュールから構成される論理デバイスを用いた記憶領域とコピーペアの関係となる記憶領域は、前記第二のストレージ装置の前記一つ以上のハードディスクドライブから構成される記憶領域である
ことを特徴とするストレージ装置。
【発明の詳細な説明】
【技術分野】
【０００１】
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する。特に本発明は、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に関する。
【背景技術】
【０００２】
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。
【０００３】
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、SCSI(Small Computer System Interface)等の従来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。

【０００４】

【特許文献１】特開2004-5370号公報

【特許文献２】米国特許第6529416号公報

【発明の開示】

【発明が解決しようとする課題】

【０００５】

先に述べたフラッシュメモリおよびハードディスク双方の記憶媒体の特徴を活用する形態として、ホスト計算機に対して記憶領域を提供する論理デバイスとして、例えばデータウェアハウスのような大量データのランダムリードアクセスの高性能化が求められる場合はフラッシュメモリから構成される論理デバイスを提供し、例えば長期データのアーカイブといったデータ保持コストを抑えたい場合はハードディスクから構成される論理デバイスを提供することが考えられる。このように、二つの記憶媒体の論理デバイスを一つのストレージ装置で提供できることが望ましい。

【０００６】

上記のようなストレージ装置を実現する形態として、特許文献２のフラッシュメモリディスクを特許文献１に示すストレージ装置に搭載し、フラッシュメモリディスクとハードディスクを混在させることで、SCSI等のアクセス手段を変更すること無く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て用いることはできる。しかし、ストレージ装置に搭載できるフラッシュメモリは、フラッシュメモリディスクが準じるハードディスクのフォームファクタの制約を受けるため、フラッシュメモリをそのまま接続したストレージ装置よりも、フラッシュメモリモジュールの高密度な実装が困難であるという課題がある。また、フラッシュメモリディスクの性能は磁気ディスクのアクセス手段のピーク性能に準じるため、将来フラッシュメモリのアクセス性能が向上してもその性能を十分発揮できない可能性があるという課題がある。

【０００７】

上記の課題を解決するには、フラッシュメモリを、従来ハードディスクのアクセス手段ではなく、直接入出力するアクセス手段を備えつつ、ハードディスクのアクセス手段も具備したストレージ装置を実現する必要がある。

【課題を解決するための手段】

【０００８】

本発明は前記課題を鑑みたものである。

【０００９】

本発明のストレージ装置は、少なくともデータを格納するフラッシュメモリと、ストレージ装置の制御を行うストレージコントローラとを備え、I/0要求を発行するホスト計算機と、ストレージシステムを管理する管理計算機と、磁気ディスクを備える第二のストレージ装置が接続される。フラッシュメモリはフラッシュメモリパッケージ(基板)上に実装され、メモリモジュールやパッケージなどの単位でストレージ装置に増設できる。前記ストレージコントローラは、ホスト計算機上で稼働するオペレーティングシステム（以下OSと略記）等のデータを格納する記憶領域である論理ボリュームを作成するとき、当該ストレージ装置のフラッシュメモリパッケージ上のフラッシュメモリから記憶領域を構成することもできるし、特許文献１の技術により、前記第二のストレージ装置の磁気ディスクから構成される記憶領域を用いて記憶領域を構成することもできる。ホスト計算機からのI/0要求が発行されたとき、前記ストレージコントローラは、もし当該ストレージ装置のフラッシュメモリから記憶領域を構成したときは、フラッシュメモリに直接アクセスしてI/0要求を処理する。もし前記第二のストレージ装置の磁気ディスクから構成される記憶領域を用いて記憶領域を構成したときは、SCSIコマンド等の磁気ディスクのアクセス手段を

用いてI/0要求を処理する。当該ストレージ装置でフラッシュメモリから構成される記憶領域を定義するときは、ホスト計算機に提供する記憶領域の容量と、フラッシュメモリの消去回数の制約を考慮した代替領域容量を合算して記憶領域を定義する。そのとき、当該記憶領域のをどのようなアプリケーションで用いるかを示す「用途」を指定し、代替領域容量のストレージ管理者からの入力を省略してもよい。また、フラッシュメモリから構成される記憶領域を定義するとき、当該記憶領域とコピーペアの関係になるハードディスクドライブから構成される記憶領域へのライトアクセス頻度から、代替領域容量を算出してもよい。

【発明の効果】

【００１０】

本発明によれば、ストレージ装置にフラッシュメモリパッケージでフラッシュメモリを搭載することで、高密度な実装が可能となる。

【００１１】

さらに、フラッシュメモリの消去回数の制約を考慮した代替領域容量の設定を設け、フラッシュメモリから構成される記憶領域を耐用年数を向上させる。さらに、代替領域容量の定義を簡素化し、フラッシュメモリに関する知識の有無にかかわらず、フラッシュメモリから構成される記憶領域を利用することができる。

【発明を実施するための最良の形態】

【００１２】

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。

【実施例１】

【００１３】

本実施形態は、フラッシュメモリパッケージでフラッシュメモリを搭載し、かつ磁気ディスクを搭載した従来のストレージ装置に接続したストレージ装置により、フラッシュメモリへの直接アクセスと従来の磁気ディスクのアクセスの両方のI/0要求処理を実現できることを示す。

(1-1)実施例1における計算機システムの構成

実施例1における計算機システム構成について説明する。図1から図2は計算機システムの構成および計算機システムに接続される装置の構成を示し、図3から図7は各装置に具備される管理情報を示す。

【００１４】

図1に計算機システムの構成を示す。1台以上のホスト計算機10000と、後述するフラッシュメモリを搭載するストレージ装置30000と、後述するハードディスクを搭載する1台以上のストレージ装置40000とが、ストレージネットワーク20000で互いに接続される。また、ストレージ装置20000とストレージ装置30000は、管理ネットワーク60000を介して、管理計算機50000に接続される。

【００１５】

ホスト計算機10000の詳細な構成について説明する。ホスト計算機10000は、CPU(Central Processing Unit)11000と、メモリ12000と、ストレージネットワーク20000に接続するための一つ以上のI/0ポート13000と、処理結果を出力するためのディスプレイ装置等の出力部16000と、キーボードやマウス等の入力部15000とを有し、これらは内部バス14000で互いに接続される。メモリ12000には、OS(図示せず)と、ストレージ装置へのデータアクセスを伴う処理を行うアプリケーション(図示せず)が存在する。これらのプログラムは、ハードディスク等の記憶媒体(図示せず)からロードされ、プロセッサ11000がこれらのプログラムを参照するものとする。

【００１６】

ストレージ装置30000は、I/0ポート31000と、ストレージコントローラ32000と、フラッシュメモリを搭載するフラッシュメモリパッケージ33000と、キャッシュメモリ34000と、管理ネットワーク60000に接続するための管理ポート35000とを有する。詳細な構成例は後

述する。

【００１７】

ストレージ装置40000は、I/Oポート41000と、ストレージコントローラ42000と、ハードディスク43000と、キャッシュメモリ44000と、管理ポート45000とを有する。以降の説明を簡単にするため、本発明におけるストレージ装置40000は、特許文献１の技術により、ストレージ装置30000の記憶領域として、ストレージ装置40000のハードディスク43000の記憶領域を提供する装置である。よって、従来のストレージ装置と同様のI/O処理を実施するため、詳細な説明を省略する。

【００１８】

管理計算機50000の詳細な構成について説明する。管理計算機50000は、CPU51000と、メモリ52000と、管理ポート53000と、処理結果を出力するためのディスプレイ装置等の出力部56000と、キーボードやマウス等の入力部55000とを有し、これらは内部バス54000で互いに接続される。管理計算機50000は、ストレージ装置30000または40000の運用管理を実行するため、ハードディスク等の記憶媒体(図示せず)またはストレージ装置30000または40000との通信によりプログラムをメモリ52000にロードして実行する。また、図示していないが、メモリ52000には、OS(オペレーティングシステム)が記憶媒体からロードされ、プロセッサ51000がこれらのプログラムを実行している。

【００１９】

ホスト計算機とストレージ装置の間、および、ストレージ装置とストレージ装置の間のストレージネットワーク50000は、スイッチにより構成されてもよいし、装置間で直接接続されていてもよい。

【００２０】

以降の説明の都合上、実施例1では、ホスト計算機がストレージネットワークを介してストレージ装置ST1,ST2に接続されるものとする。また、ストレージネットワーク20000はFC(Fibre Channel)プロトコルを用いたネットワーク、管理ネットワーク60000はIPプロトコルを用いたネットワークであるとする。

【００２１】

図2にストレージ装置30000の詳細な構成例を示す。ストレージ装置30000は、一つ以上のI/Oポート31000と、ストレージコントローラ32000と、フラッシュメモリモジュール33100を搭載する一つ以上のフラッシュメモリパッケージ33000と、キャッシュメモリ(CM)34000と、管理ネットワーク60000に接続するための管理ポート35000とを有しこれらはストレージコントローラ32000を介して互いに接続される。

【００２２】

ストレージコントローラ32000には、ストレージ装置内の制御を行うマイクロプログラム(図示せず)が存在し、特許文献１で開示された技術を用いて、ストレージ装置40000の記憶領域を外部デバイス(EXDEV)として認識し、ホスト計算機10000に対する論理デバイス(LDEV)として提供できるものとする。さらに、ストレージコントローラ32000には、論理デバイスを管理するための論理デバイステーブル(LDEV TBL)32100と、後述するフラッシュメモリパッケージ上のフラッシュメモリモジュールから構成される記憶領域である内部デバイス(INDEV)を管理するための内部デバイステーブル(INDEV TBL)32200と、外部デバイスを管理するための外部デバイステーブル(EXDEV TBL)32300と、LDEVの一部領域のCM格納状況を管理するためのキャッシュ割り当てテーブル(CM TBL)32400と、フラッシュメモリパッケージを管理するためのフラッシュメモリパッケージテーブル(FPK TBL)32500と、デバイスの割り当てや状態を管理するデバイス管理プログラム(DEV管理PG)32600と、LDEVへのI/Oを制御するI/O制御プログラム(I/O制御PG)32700と、後述するINDEVの用途に対応した代替領域容量を定義する用途別代替領域容量テーブル(USAGE TBL)32800と、DEV管理PG32600においてINDEVを設定するときに呼び出される内部デバイス設定プログラム(INDEV PG)32900を備える。これらは、図示しないが、ストレージコントローラ内に存在するハードディスク等の記憶媒体からメモリにロードされてプロセッサにより実行されるものとする。

【００２３】
フラッシュメモリパッケージ(FPK)33000は、一つ以上のフラッシュメモリモジュール（FM）33100から構成される基板である。ここで、フラッシュメモリモジュール（FM）33100とは、例えばDIMM(Dual Inline Memory Module)のような形状の、プリント基板にフラッシュメモリチップを複数搭載した基板である。後述するDEV管理PG32600において、ストレージ管理者から指定された容量を満たすようFPKおよびFMを組み合わせてINDEVは作成され、RAID技術によりデータを分散して格納する。
【００２４】
なお、以降の説明の都合上、I/Oポート31000はP11とP12の2個であり、P11はホスト計算機との接続に用いられ、P12はストレージ装置ST2との接続に用いられているものとする。また、FPK33000は4枚搭載され、各FPKは容量256GBのFMをn枚(nは2以上の整数)搭載されているものとする。
【００２５】
図3に、ストレージ装置ST1が具備するLDEV TBL32100の例を示す。
【００２６】
図3は、ストレージ装置ST1がLDEVを管理するためのテーブルである。LDEV TBL32100は、ストレージ装置ST1内でLDEVの一意の識別子となるLDEV IDを登録するフィールド32110と、LDEVのホスト計算機に提供するホスト認識容量を登録するフィールド32120と、LDEVに対応する記憶領域がINDEVかEXDEVのどちらかを示す対応DEV IDフィールド32130と、当該LDEVがホスト計算機に割り当て済みか否かを示す状態フィールド32140と、当該LDEVがホスト計算機に割り当て済みの場合、ホスト接続先I/Oポート番号・SCSI ターゲットID・SCSI LUNを登録するフィールド32150とから構成される。
【００２７】
状態フィールド32140には、当該LDEVがホスト計算機に割り当て済みの場合には「Allocated」を、ホスト計算機に未割り当ての場合には「Unallocated」が登録される。また、当該LDEVがホスト計算機に未割り当ての場合、ポート番号/ターゲットID/LUNフィールド32150には、「N/A」(Not Applicable)が登録される。
【００２８】
図4に、ストレージ装置ST1が具備するINDEV TBL32200の例を示す。
【００２９】
図4は、ストレージ装置ST1がINDEVを管理するためのテーブルである。INDEV TBL32200は、ストレージ装置ST1内でINDEVの一意の識別子となるINDEV IDを登録するフィールド32210と、後述するINDEVの基本容量を登録するフィールド32220と、後述する代替領域をINDEVが確保するための代替領域容量を登録するフィールド32230と、INDEVがLDEVに割り当て済みの場合対応するLDEV IDを登録するフィールド32240と、INDEVのRAID構成を示すフィールド32250と、INDEVのストライプサイズを示すフィールド32260と、INDEVを構成するFPKを登録するフィールド32270と、INDEVを構成するFPK内のFMを登録するフィールド32280と、後述するINDEVのアドレス空間対応リストを登録するフィールド32290から構成される。
【００３０】
ここで、基本容量と代替領域容量について説明する。フラッシュメモリの各ビットの更新は１から０（または０から１）の一方向に限定される。逆の変更が必要な場合は、フラッシュメモリチップのブロック（以下、「メモリブロック」と呼ぶ）の消去を行って一旦メモリブロック全体を１（または０）にする必要がある。また、この消去回数に制約(上限回数)があり、例えばＮＡＮＤ型フラッシュメモリの場合は一万から十万回程度が限度である。そのため、フラッシュメモリをハードディスクの代替として計算機に接続する場合、メモリブロック毎の書き込み頻度の偏りによって、一部のメモリブロックのみが消去回数の上限に達して使えなくなってしまう可能性がある。例えば一般的なファイルシステムでは、ディレクトリやiノードに割り当てられたメモリブロックは他のメモリブロックに比べて書き換え頻度が高いので、これらのメモリブロックのみ消去回数の上限に達する

可能性が高い。そこで、使えなくなったメモリブロック（不良メモリブロック）に対して代替となるメモリブロック（代替メモリブロック）を割り当てることで、INDEVの寿命を延ばすものとする。基本容量は、ホスト認識容量を提供するために必要な容量である。なお、基本容量には、後述するRAID構成によるパリティデータ格納容量を含むため、基本容量とホスト認識容量とは一致しない場合がある。例えば図4のINDEV IN01の場合、基本容量は768GBであるが、後述するRAID5(3D+1P)構成をとるため、ホスト認識容量は768GBの75%の576GBとなり、LDEV L01のホスト認識容量に等しくなる。代替領域容量は、代替メモリブロックを確保するために確保する記憶領域の容量である。なお、代替領域容量も、基本容量と同じRAID構成をとるものとし、パリティデータ格納容量を含む容量を定義する。

【００３１】

また、ストライプサイズフィールド32260には、メモリブロック長を定義してもよい。以降の説明の都合上、本発明では、すべてのFPKのFMのメモリブロック長は256KBとする。

【００３２】

対応FPKフィールド32270において、「FPKn-m」(n,mはn=<mを満たす1以上の整数)のような表記は、「FPKnからmまでを用いる」ことを示す。本発明ではストレージ装置ST1にはFPK1からFPK4の4枚のFPKが搭載されており、図4のINDEV IN01の場合、「FPK1-4」であることから、FPK1からFPK4の4枚すべてのFPKを用いることを示す。また、対応FM番号フィールド32280では、1以上の整数kが登録される。この整数は、対応FPKフィールドに示されたFPKのk番目のFMを用いることを示す。この二つのフィールドから、例えば図4のINDEV IN01の場合、FPK1のFM1-1，FPK2のFM2-1，FPK3のFM3-1，FPK4のFM4-1を用いることがわかる。

RAID構成フィールド32250には、本ストレージ装置ST1で対応可能なRAID構成を登録する。たとえばRAID5(3D+1P)とは、4枚のFPKのうち3枚でストライプされたデータを格納し(3Dと記載)、1枚でパリティを格納する(+1Pと記載)ことを示す。

【００３３】

アドレス空間対応リスト32290では、INDEVのホスト認識容量のどのアドレスのデータを、FMのどのメモリブロックで格納するかの対応関係を示す。具体的には、図4のINDEV IN01の場合、リストの第一要素「(0 to 768kB，FM1-#1)」の意味は、INDEVアドレス0から768KBまでのデータを各FPKの1番目のFM(つまりFM1-1,FM2-1,FM3-1,FM4-1)の1番のメモリブロックで格納することである。リストの第二要素「(768kB to 1536KB，FM1-#850)」の意味は、INDEVアドレス768KBから1536KBまでのデータを各FPKの1番目のFMの850番のメモリブロックで格納することである。

【００３４】

図5に、ストレージ装置ST1が具備するEXDEV TBL32300の例を示す。

【００３５】

図5は、ストレージ装置ST1がEXDEVを管理するためのテーブルである。EXDEV TBL32300は、ストレージ装置ST1内でEXDEVの一意の識別子となるEXDEV IDを登録するフィールド32310と、EXDEVのホスト計算機に提供するホスト認識容量を登録するフィールド32320と、EXDEVに対応するLDEV IDを示すフィールド32330と、当該EXDEVへストレージ装置ST1がアクセスするためのI/Oポートを示すフィールド32340と、当該EXDEVの識別情報として、装置IDを示すフィールド32350と、LDEV IDを示すフィールド32360と、ストレージ装置ST1への接続先I/Oポート番号・SCSI ターゲットID・SCSI LUNを登録するフィールド32370とから構成される。

【００３６】

ストレージ識別情報に関連するフィールド32350,32360,32370の値は、当該EXDEVに対してストレージ装置ST1からSCSI Inquiryコマンドを発行することにより取得できる。

【００３７】

図6に、ストレージ装置ST1が具備するCM TBL32400の例を示す。

【００３８】

図6は、ストレージ装置ST1がCMの割り当て状態を管理するためのテーブルである。CMは全体容量32460を、ブロックの単位に区切って利用する。そのブロック長は32450に登録さ

れる。ブロック長32450は、たとえばストレージ装置に搭載されたフラッシュメモリのメモリブロック長や対応するRAID構成をもとにストレージ装置の固定値としてもよいし、ストレージ管理者により定義できてもよい。本発明では、キャッシュ容量は512GB、ストレージ装置ST1がメモリブロック長256KBのFMを4個使ってRAID構成をとるので、キャッシュブロック長を1024KBと定めてある。よって、キャッシュブロック数は512個となり、各ブロックのレコードが存在する。各レコードには、キャッシュブロックを一意に識別するブロックIDフィールド32410と、キャッシュ状態を示すフィールド32420と、キャッシュ割り当てされたLDEVを示すフィールド32430と、キャッシュ割り当てされたアドレスを示すフィールド32440から構成される。

【００３９】

状態フィールド32420は、CMのデータとINDEVまたはEXDEVに格納されたデータが一致する「Clean」状態、CMのデータが更新されているがまだINDEVまたはEXDEVに反映されていない「Dirty」状態、まだ割り当てされていない「Not Used」状態が存在する。「Not Used」状態のとき、対応LDEV IDフィールド32430とLDEVアドレスフィールド32440は「N/A」(Not Applicable)を登録する。

【００４０】

図7に、ストレージ装置ST1が具備するFPK TBL32500の例を示す。

【００４１】

図7は、ストレージ装置ST1がFPKの利用状態を管理するためのテーブルである。FPK TBL32500は、FPKを一意に識別できるIDを登録するフィールド32510と、各FPK内のFMを一意に識別できるIDを登録するフィールド32520と、各FMの容量を示すフィールド32530と、各FMのアドレスを示すフィールド32540と、アドレスフィールド32540のフラッシュメモリブロックの利用状態を示すフィールド32550と、アドレスフィールド32540のフラッシュメモリブロックに割り当てられたINDEVを示すフィールド32560と、当該アドレスの書き換え回数を示すフィールド32570から構成される。

【００４２】

状態フィールド32420は、利用済みを示す「Used」と、未使用を示す「Not Used」のいずれかが登録される。「Not Used」状態のとき、割当INDEV IDフィールド32560と書換回数フィールド32570には「N/A」(Not Applicable)を登録する。書換回数フィールドは、各FMの書換許容回数の仕様との割合により、どれだけ多くのフラッシュメモリブロックが不良ブロック化している可能性があるかの目安として用いる。

【００４３】

以上が実施例1における計算機システムの構成である。

(1-2)実施例1におけるストレージ装置ST1のLDEVに関連する処理

次に、本実施例における、ストレージ装置ST1が行う処理について説明する。本処理は、ストレージ装置ST1 30000内のDEV管理PG32600と、I/O制御PG32700と、INDEV PG32900とによって実現する。

【００４４】

DEV管理PG32600は、EXDEVを認識してLDEVに登録する処理、INDEVを作成してLDEVに登録する処理を行うプログラムである。I/O制御PG32700は、LDEVへのホスト計算機からのI/Oを制御するプログラムである。INDEV PG32900は、INDEVを作成し、LDEVに登録するプログラムである。

【００４５】

以下順にプログラムのフローチャートを示す。なお、特に断りが無ければ、各プログラムのステップは、ストレージ装置ST1のストレージコントローラ32000が実行するものとする。

【００４６】

図8に、DEV管理PG32600のフローチャートを示す。

【００４７】

まず、管理計算機は、EXDEV認識をストレージ装置ST1に指示する(ステップS32610)。具

体的には、ストレージ装置ST2のLDEVを作成し、当該LDEVをストレージ装置ST1のI/0ポートから認識できるように接続を確立し、どのI/0ポートからEXDEVを認識すればよいかをストレージ装置ST1に送信する。

【００４８】

ステップS32610でI/0ポートを指定されたストレージ装置ST1はEXDEV認識を実行する(ステップS32620)。具体的には、指定されたI/0ポートからSCSI Inquiryコマンドを発行し、その結果をもとにEXDEV TBL32300を作成し、その一覧を管理計算機に送信する。

【００４９】

次に管理計算機は、認識できたEXDEVのうちLDEVに登録するEXDEVを指定する(ステップS32630)。具体的には、ステップS32620で作成されたEXDEVの中から、ホスト計算機にLDEVとして提供するEXDEVを指定し、その一覧をストレージ装置ST1に送信する。ここで、ホスト計算機へのLDEVの割り当ても指定してもよい。

【００５０】

ステップS32630でLDEVに登録するEXDEVを指定されたストレージ装置ST1は、LDEV登録を実行する(ステップS32640)。具体的には、指定されたEXDEVのエントリからLDEV TBL32100を作成する。もしホスト計算機へのLDEVの割り当ても指定されていれば、状態フィールド32140やポート番号/ターゲットID/LUNフィールド32150も登録してもよい。ストレージ装置ST1は、LDEV登録結果およびホスト計算機への割り当て結果を管理計算機に送信する。

【００５１】

管理計算機は、送信されたLDEV一覧を表示する(ステップS32650)。表示内容については後述する。

【００５２】

以上がEXDEVの認識およびLDEVへの登録に関する処理である。

【００５３】

次に、INDEVの作成およびLDEVへの登録に関する処理について説明する。

【００５４】

まず、管理計算機は、INDEV作成をストレージ装置ST1に指示する(ステップS32660)。具体的には、ホスト認識容量・代替領域容量・RAID構成等を指定してLDEV作成を指示すればよい。詳細は表示内容にて後述する。

【００５５】

ステップS32660でホスト認識容量・代替領域容量・RAID構成等を指定されたストレージ装置ST1はINDEV作成を実行する(ステップS32670)。具体的には、後述するINDEV PG32800を呼び出す。

【００５６】

管理計算機は、送信されたLDEV一覧を表示する(ステップS32650)。表示内容については後述する。

【００５７】

以上がDEV管理PG32600のフローチャートである。

【００５８】

次に、図13に示す内部デバイス設定プログラム(INDEV PG)32900について説明する。説明にあたり、図9に示す内部デバイス作成画面70000と図12に示す用途別代替容量テーブル(USAGE TBL)32800を用いる。

【００５９】

前記DEV管理PG32600のステップS32670により、図13に示すINDEV PG32900が実行されると、ストレージ装置ST1は、管理計算機からのINDEV作成要求を受信する(ステップS91000)。

【００６０】

以下、図13のステップS91000について詳しく説明するため、図9のストレージ装置ST1がデバイス管理プログラムを実行中に管理計算機が表示する、内部デバイス作成画面について説明する。内部デバイス作成画面は、ストレージ管理者からINDEV設定に必要なパラメ

ータを受信する画面である。内部デバイス作成画面70000は、ホスト認識容量を指定するフィールド70010と、代替領域容量を指定するフィールド70020と、RAID構成を指定するフィールド70030と、INDEV作成を送信するために押し下げするボタン70050から構成される。なお、図示しないが、ホスト認識容量を指定するのではなく、基本容量を直接指定してもよい。

【００６１】

図13のステップS91000に戻って、図9のようなパラメータが指定されたとき、ストレージ装置ST1は、まず、指定されたホスト認識容量とRAID構成から基本容量を算出する。たとえば、ホスト認識容量576GBでRAID構成が「RAID5(3D+1P)」の場合、ホスト認識容量の3分の4倍を基本容量とすればよいので、基本容量は768GBとなる。

【００６２】

以上がステップS91000の説明である。

【００６３】

次に、ストレージ装置ST1は、代替領域容量を決定するため、図9の内部デバイス作成画面において、代替領域容量を指定されているか、用途を指定されているか判断する(ステップS91010)。代替領域容量を指定されているときは、指定された容量をそのまま用いるものとして、ステップS91030にジャンプする。用途を指定されているときは、ステップS91020にジャンプする。

【００６４】

ステップS91010において、用途を指定されているとき、ストレージ装置ST1は、用途から代替領域容量を決定する(ステップS91020)。

【００６５】

以下図13のステップS91030について詳しく説明するため、図12に示す用途別代替容量テーブル(USAGE TBL)32800について説明する。USAGE TBLは、INDEVを用いるアプリケーション等を示す用途32320と、当該用途において必要な年あたりの代替領域容量32330をストレージ管理者等が予め定義しておくテーブルである。

【００６６】

図13のステップS91030に戻って、USAGE TBLにより、図9の内部デバイス作成画面において、LDEV用途フィールド70040の値と、作成するINDEVを何年利用することを想定しているかを示す耐用年数フィールド70060の値を受信し、LDEV用途フィールド70040の値に合致する用途における、図12の年あたりの代替領域容量32330の値と耐用年数フィールド70060の値の積により、代替領域容量を計算できる。たとえば、LDEV用途が「データベース」で耐用年数が「3年」の場合、用途「データベース」の年あたりの代替領域容量は「50GB」であるので、代替領域容量は150GBとなる。

【００６７】

以上がステップS91020の説明である。

【００６８】

ステップS91010またはステップS91020において代替領域容量が決定すると、ストレージ装置ST1は、基本容量と代替領域容量の和を求め、割り当て可能なFMを探す(ステップS91030)。たとえば、基本容量が768GBで代替領域容量を256GBの場合、和は1024GBとなる。求めた和に1024GBに一致するFMのメモリブロック構成を決定する場合、FPK TBL32500の各FPKの先頭FMから順に、状態フィールド32550が「Not Used」である領域を累積加算する。図7の例でいえば、各FPKの一番目のFMの全領域と二番目のFMの先頭128GBは「Used」である。よって、1024GBの領域を確保するためには、二番目のFMの残り128GBと三番目のFMの先頭128GBを利用すればよい。ここでは先頭FMから詰めて領域を確保する例を述べたが、領域の確保の方法は任意の方法であってもよい。

【００６９】

割り当て可能なFMを探すことができたら、ストレージ装置ST1は、INDEV TBLとLDEV TBLを更新する(ステップS91040)。具体的には、FPK TBL32500で、ステップS91030において確保したメモリブロックの状態を「Used」とし、割当INDEV IDに新規IDを割り当てる。さら

に、作成した新規INDEV IDから、INDEV TBL32200の新規エントリを作成する。作成したINDEV TBL32200の新規エントリに対し、ステップS91010で決定した基本容量と、ステップS91010またはステップS91020で決定した代替領域容量と、ステップS91010で決定したRAID構成と、本ステップで決定した対応FPKと対応FM番号を登録する。ストライプサイズは規定の値であってもよい。さらに、LDEV TBL32100の新規エントリを作成し、LDEVに登録する。さらに、LDEV TBL32100の新規エントリのLDEV IDを先に作成したINDEV TBLのエントリに登録する。

【００７０】

以上がステップS91030の説明である。

【００７１】

最後にストレージ装置ST1は、管理計算機に、作成した新規INDEVを登録したLDEVの一覧を送信し、完了報告とする(ステップS91050)。

【００７２】

以上が図13に示す内部デバイス設定プログラム(INDEV PG)32900の説明である。

【００７３】

図10にストレージ装置ST1がデバイス管理プログラムを実行中に管理計算機が表示する、論理デバイス一覧表示画面の一例を示す。

【００７４】

論理デバイス一覧表示画面80000は、LDEV IDを表示するフィールド80010と、LDEVのホスト計算機に提供するホスト認識容量を表示するフィールド80020と、LDEVに対応する記憶領域がINDEVかEXDEVのどちらかを示す対応DEV IDフィールド80030と、当該LDEVがホスト計算機に割り当て済みか否かを示す状態フィールド80040と、当該LDEVがホスト計算機に割り当て済みの場合、ホスト接続先I/Oポート番号・SCSI ターゲットID・SCSI LUNを登録するフィールド80050とから構成される。これらはLDEV TBL32100の値を表示すればよい。

【００７５】

さらに、LDEVがINDEVから構成されている場合、以下のフィールドを表示してもよい。それは、INDEVの基本容量を表示するフィールド80060と、INDEVの代替領域容量を表示するフィールド80070と、INDEVのRAID構成を示すフィールド80080である。これらは、INDEV TBL32200の値を表示すればよい。なお、これらのフィールドは、EXDEVの場合は「N/A」(Not Applicable)と表示すればよい。

【００７６】

さらに、LDEVがINDEVから構成されている場合、INDEVを構成するFPKまたはFMがあらかじめ定義された書換回数を超過する場合、不良ブロックが多くなっている可能性があることを示す警告を示すフィールド80090を設けてもよい。これは、FPK TBL32500の書換回数32570フィールドを定期的にモニタリングし、所定の回数を超過したFMを用いるINDEVに警告を示せばよい。なお、この警告は、画面表示だけでなく、メールやSNMP(Simple Network Management Protocol)による通知、Syslog等のログ蓄積を行ってもよい。

【００７７】

図11に、ストレージ装置ST1が実行するI/O制御プログラムの処理内容を示すフローチャートを示す。

【００７８】

まず、ストレージ装置ST1はホストからの入出力要求を受信する(ステップS90000)。具体的には、あるLDEVに対するSCSIのReadコマンドやWriteコマンドなどである。

【００７９】

入出力要求を受信したストレージ装置ST1は、その要求がデータのReadかWriteかを判断する(ステップS90010)。Readの場合はステップS90020に、Writeの場合はステップS90100にジャンプする。

【００８０】

ステップS90010において要求がReadの場合、ストレージ装置ST1は、Read先のアドレス

を解釈し、当該アドレスのデータがキャッシュヒットするか否かを判断する(ステップS90020)。具体的には、CM TBLの全エントリのうち、対応LDEV IDフィールド32430とLDEVアドレスフィールド32440が合致するものがあればキャッシュヒット、合致しなければキャッシュミスと判断する。キャッシュヒットの場合ステップS90030に、キャッシュミスの場合ステップS90040にジャンプする。

【００８１】

ステップS90020でキャッシュヒットしたとき、キャッシュデータをリードし(ステップS90030)、当該データをホストへ転送し(ステップS90080)、ホストへデータ入出力要求が完了したことを報告し(ステップS90090)、I/O制御PGを終了する。

【００８２】

ステップS90020でキャッシュミスしたとき、データアクセス先のLDEVがINDEVに対応するか、EXDEVに対応するか判断する(ステップS90040)。具体的には、LDEV TBL32100の対応DEV IDフィールド32130で判断すればよい。INDEVの場合はステップS90050に、EXDEVの場合はステップS90060にジャンプする。

【００８３】

ステップS90040でINDEVと判断された場合、INDEVからデータをリードする(ステップS90050)。具体的には、INDEV TBL32200の対応FPKフィール32270、対応FM番号フィールド32280、アドレス空間対応リストフィールド32290からどのFMのどのメモリブロックからデータを読み出せばよいか調べる。たとえば、INDEV IN01のINDEVアドレス0を先頭とした512KBの長さのデータをリードしたい場合、FPK1-4の一番目のFM(つまりFM1-1,FM2-1,FM3-1,FM4-1)のFMメモリブロック#1番から、先頭から512KB分のデータをリードすればよい。リードした後ステップS90070に進む。

【００８４】

テップS90040でEXDEVと判断された場合、EXDEVからデータをリードする(ステップS90060)。具体的には、EXDEV TBL32300のイニシエータポートID32340から、ストレージ識別情報に指定されたLDEVに対して、たとえばSCSIのリードコマンドを発行すればよい。リードした後ステップS90070に進む。

【００８５】

ステップS90050またはステップS90060でデータをリードした後、当該リードしたデータのためにCMを割り当て、CMデータを更新する(ステップS90070)。具体的には、CM TBL32400から状態が「Not Used」のブロックがないかを探し、見つかればそのブロックを割り当てるものとして、データを更新する。その際、フラッシュメモリやハードディスクのブロックサイズとCMのブロックサイズの違いにより、複数のブロックが必要であれば、複数のブロックを割り当てればよい。また、CMに「Not Used」のブロックが確保できなかった場合は、任意の方法で状態がCleanのブロックを置き換えて割当してもよい。CMの割り当ておよびデータ更新完了後、当該データをホストへ転送し(ステップS90080)、ホストへデータ入出力要求が完了したことを報告し(ステップS90090)、I/O制御PGを終了する。

【００８６】

ステップS90010において要求がWriteの場合、ストレージ装置ST1は、Write先のアドレスを解釈し、当該アドレスのデータがキャッシュヒットするか否かを判断する(ステップS90100)。具体的な処理はステップS90020と同じであるので説明を省略する。キャッシュヒットの場合ステップS90110に、キャッシュミスの場合ステップS90160にジャンプする。

【００８７】

ステップS90100でキャッシュヒットしたとき、ストレージ装置ST1はWriteデータでキャッシュを更新する(ステップS90110)。具体的には、CMブロックのデータ更新後、CM TBL32400で当該ブロックの状態を「Dirty」に変更する。ステップS90110終了後、ホストへデータ入出力要求が完了したことを報告する(ステップS90120)。さらに、CMブロックのDirty状態を解消するため、INDEVまたはEXDEVにデータをライトするため、データアクセス先のLDEVがINDEVに対応するか、EXDEVに対応するか判断する(ステップS90130)。具体的には、LDEV TBL32100の対応DEV IDフィールド32130で判断すればよい。INDEVの場合はステップS

90140に、EXDEVの場合はステップS90150にジャンプする。

【００８８】

ステップS90100でキャッシュミスしたとき、ストレージ装置ST1はWriteデータのためのキャッシュを割り当てる(ステップS90170)。具体的な処理は、ステップS90070におけるキャッシュ割り当てと同じであるので説明を省略する。ステップS90170の後、ホストへデータ入出力要求が完了したことを報告する(ステップS90180)。さらに、確保したCMブロックへデータを格納するため、INDEVまたはEXDEVのアクセス先のデータを先にCMブロックに読み込んでホスト計算機のWriteデータで修正する「Read-Modify-Write」を実施するため、データアクセス先のLDEVがINDEVに対応するか、EXDEVに対応するか判断する(ステップS90180)。具体的には、LDEV TBL32100の対応DEV IDフィールド32130で判断すればよい。

【００８９】

ステップS90180においてINDEVと判断された場合は、INDEVからデータを読み込む(ステップS90190)。具体的な処理は、ステップS90050に同じであるので説明を省略する。さらに、ステップS90190で読み込んだデータを用いてCMブロックのデータを更新する(ステップS90200)。さらに、ステップS90140にジャンプする。

【００９０】

ステップS90180においてEXDEVと判断された場合は、EXDEVからデータを読み込む(ステップS90210)。具体的な処理はステップS90060に同じであるので説明を省略する。さらに、ステップS90210で読み込んだデータを用いてCMブロックのデータを更新する(ステップS90220)。さらに、ステップS90150にジャンプする。

【００９１】

ステップS90140において、ストレージ装置ST1は、CMブロックにRead-Modify-WriteしたWriteデータをINDEVへ反映するため、INDEVへのデータの書き込みを行う。具体的には、INDEV TBL32200の対応FPKフィール32270、対応FM番号フィールド32280、アドレス空間対応リストフィールド32290からどのFMのどのメモリブロックへデータを書き込めばよいか調べる。たとえば、INDEV IN01のINDEVアドレス0を先頭とした512KBの長さのデータをライトしたい場合、FPK1-4の一番目のFM(つまりFM1-1,FM2-1,FM3-1,FM4-1)のFMメモリブロック#1番から、先頭から512KB分のデータを書き込めばよい。また、別の方法として、代替領域容量で確保されたFMメモリブロックへデータを書き込み、アドレス空間対応リストの対応関係の書き換えを行ってもよい。この場合、アドレス空間対応リストにおけるFMメモリブロックの使用順序の順序性の保証はなくなるが、FMメモリブロックの書換回数を平滑化でき、早期に不良化の対象となることを防ぐ、言いかえれば、FPKやFMの寿命を延ばす効果が期待できる。ステップS90140終了後、ストレージ装置ST1はI/O制御PGを終了する。

【００９２】

ステップS90150において、ストレージ装置ST1は、CMブロックにRead-Modify-WriteしたWriteデータをEXDEVへ反映するため、EXDEVへのデータの書き込みを行う。具体的には、EXDEV TBL32300のイニシエータポートID32340から、ストレージ識別情報に指定されたLDEVに対して、たとえばSCSIのライトコマンドを発行すればよい。ステップS90150終了後、ストレージ装置ST1はI/O制御PGを終了する。

【００９３】

以上がI/O制御PG32700のフローチャートである。

【００９４】

以上が、実施例1におけるストレージ装置ST1のLDEVに関連する処理である。

【００９５】

本実施例によれば、ストレージ装置にフラッシュメモリパッケージでフラッシュメモリを搭載し、磁気ディスクのアクセス手段を用いること無くフラッシュメモリモジュールに直接アクセスすることができる。また、従来の磁気ディスクのアクセス手段を用いたI/O要求処理を実現することで、既存の磁気ディスクストレージ装置をそのまま利用することができる。

【００９６】

以上が、実施例1の説明である。

【実施例２】

【００９７】

本実施形態は、フラッシュメモリから構成される記憶領域であるINDEVを定義するとき、当該記憶領域とコピーペアの関係になるハードディスクドライブから構成される記憶領域へのライトアクセス頻度から、INDEVの代替領域容量算出することで、INDEVの定義を簡素化することができることを示す。

なお、以降実施例2の説明において、特に断りがなければ、実施例1の構成およびテーブルおよびプログラムが実行されるものとし、実施例1との違いのみ説明する。

(2-1)実施例2における計算機システムの構成

実施例2における計算機システム構成について説明する。

【００９８】

図14に、実施例2における計算機システムの構成を示す。実施例1との違いは、論理デバイステーブル36000のエントリが増えること、内部デバイス設定プログラム36100が異なることである。具体的には、実施例2におけるストレージ装置ST1のLDEVに関する処理で詳しく説明する。

【００９９】

以上が実施例2における計算機システムの構成である。

(2-2)実施例2におけるストレージ装置ST1のLDEVに関連する処理

次に、本実施例における、ストレージ装置ST1が行う処理について説明する。本処理は、ストレージ装置ST1 30000内のDEV管理PG32600と、I/O制御PG32700と、INDEV PG36100によって実現する。　DEV管理PG32600とI/O制御PG32700は、実施例1と同じであるので説明を省略する。

【０１００】

以下順に、実施例1との違いがある論理デバイステーブル36000と、INDEV PG36100について説明する。

【０１０１】

図15に、実施例2においてストレージ装置ST1が具備するLDEV TBL36000の例を示す。

【０１０２】

実施例1との違いは、各LDEVのライトアクセス頻度を保持するエントリ32160を有する点である。ライトアクセス頻度とは、ストレージ装置ST1のストレージコントローラ32000がLDEVに対するデータの書き込み命令(例えばSCSIのライトコマンド等)を一定時間に受領した回数を示し、たとえば毎秒当たりのライトコマンド受領回数(IOPS(Input Output Per Second))の値を保持してもよいし、1分間といったモニタ時間のライトコマンド受領数をIOPSに変換してもよい。さらに、過去数回のIOPSの平均値を保持してもよいし、最大値や最小値であってもよい。さらに、過去のIOPSの履歴を蓄積し、後述の代替容量計算にあたり、ストレージ管理者に選択させてもよい。

【０１０３】

次に、図17に示す内部デバイス設定プログラム(INDEV PG)326100について説明する。説明にあたり、図16に示す内部デバイス作成画面71000と図12に示す用途別代替容量テーブル(USAGE TBL)32800を用いる。

【０１０４】

ステップS91000は実施例1と同じである。

【０１０５】

ステップS91010の実施例1との違いは、代替領域容量を指定されていない場合、ステップS92000にジャンプする点である。

【０１０６】

ステップS91010において、代替領域容量を指定されていない場合、図16に示す内部デバイス作成画面71000のペア対象LDEVフィールド70070と耐用年数フィールド70060により、耐用年数の指定があるか否かを判断する(ステップS92000)。指定があればステップS92010

にジャンプし、指定がなければステップS92020にジャンプする(ステップS92000)。

【０１０７】

ここで、ペア対象LDEVについて詳しく説明する。ペア対象LDEVとは、たとえば以下の二つのケースを想定した、ストレージ装置ST1とST2から構成されるストレージシステムが提供する記憶領域のデータ移行やデータコピーを実現するための対となるLDEVである。第一のケースは、現在EXDEVで構成されるLDEVに格納されたデータへのアクセス頻度が向上したため、高速なアクセスが可能となるINDEVで構成されるLDEVに移行し、LDEVの性能を向上させるケースである。第二のケースは、EXDEVまたはINDEVから構成されるLDEVに格納されたデータを、ストレージ装置ST1内部で、INDEVから構成される別のLDEVにコピーすることで、LDEVの可用性を向上させるケースである。このように、ペア対象となるLDEVの情報を取得することで、後述するステップS92010において代替領域容量を算出するにあたり、移行元またはコピー元のLDEVのライトアクセス頻度を考慮して代替領域容量を算出することができる。

【０１０８】

図17の説明に戻り、ステップS92010において、ストレージ装置ST1は、指定されたペア対象LDEVのライトアクセス頻度と耐用年数から、代替領域容量を算出する(ステップS92010)。具体的には、たとえば図18に示すような計算式を用いればよい。

【０１０９】

ここで、図18の計算式の示す内容を説明する。

【０１１０】

まず、ペア対象LDEVのライトアクセス頻度を一括書換数で割り、新規作成INDEVに対するフラッシュメモリブロック書換頻度(以下Wとする)を算出する。これは、ペア対象LDEVのライトアクセス頻度と同等のライトアクセス頻度があると仮定し、さらに、ある一定のライト回数まではキャッシュメモリ上のみで更新を行ってフラッシュメモリを更新する一括書換を考慮し、毎秒何回フラッシュメモリブロックを書き換えるかを示す値である。

【０１１１】

次に、Wと、書換ブロック長と、耐用年数の積をとる。前記三つの値の積(以下W1とする)は、耐用年数の間に発生する、新規作成LDEVの、のべライト容量である。耐用年数の間の書換頻度として、Wと耐用年数の積を利用し、先に述べたとおり、フラッシュメモリはフラッシュメモリブロック単位に消去と上書きを行うため、さらに書換ブロック長の積をとることで、W1が求められる。

【０１１２】

最後に、W1を書換回数上限で除算して代替領域容量を求める。これは、同一フラッシュメモリブロックの再利用によるW1の削減を考慮するものである。結果の単位はKBからGBに変換するための変換率(10のマイナス6乗)を乗算していることに注意する。

【０１１３】

例えば、図16で示すとおりペア対象LDEVをL02とし、耐用年数を3年としたときの代替領域容量は、図18によれば、以下の通り求められる。なお、一括書き換え数は10(つまり、10回分のライト結果をまとめてフラッシュメモリブロックに反映する)、書き換えブロック長は256KB、書換回数上限は一万回とする。L02のライトアクセス頻度がLDEVTBL36000から100IOPSであるので、Wは10となる。その結果のべライト容量W1は、10×256×(3×3600×24×365)で、242716GBとなる(先にKBからGBに変換するための変換率(10のマイナス6乗)を乗算していることに注意)。書換回数上限の除算の結果、約24.3GBとなる。

【０１１４】

以上が図18の計算式の説明である。

【０１１５】

図17の説明に戻り、上記のような計算式により、ステップS92010では、代替領域容量が計算できる。

【０１１６】

ステップS92000において、ペア対象LDEVと耐用年数の指定がない場合は、予め定めたラ

イトアクセス頻度と耐用年数を用いて、図18で述べた計算式から代替領域容量を計算する(ステップS92020)。ここで、予め定めたライトアクセス頻度として、ストレージ管理者が指定した値であってもよいし、LDEV TBLに登録されたLDEVのうち、同じ用途を指定して作成された任意のLDEVのライトアクセス頻度の値を用いてもよい。

【０１１７】

ステップS91030、ステップS91040、ステップS91050は、実施例1と同じである。

【０１１８】

以上が図17に示す内部デバイス設定プログラム(INDEV PG)32900の説明である。

【０１１９】

以上が、実施例2におけるストレージ装置ST1のLDEVに関連する処理である。

【０１２０】

本実施例によれば、代替領域容量の定義を簡素化し、フラッシュメモリに関する知識の有無にかかわらず、フラッシュメモリから構成される記憶領域を利用することができる。

【０１２１】

以上が、実施例2の説明である。

【０１２２】

なお、本実施例および前記実施例1の変形例として、計算機システムの構成は、図19に示すように、一つのストレージ装置ST1であってもよい。このとき、ストレージ装置ST1は、FPKが接続されたストレージコントローラでINDEVの管理を行い、ディスクドライブが接続されたストレージコントローラでEXDEVの管理を行い、両ストレージコントローラが互いに接続されている点で、実施例1及び実施例2と異なり、INDEVの作成方法及びLDEVの処理方法は、実施例1及び実施例2に同じである。

【図面の簡単な説明】

【０１２３】

【図１】本発明における計算機システムの構成例を示す図である。

【図２】本発明の実施例１におけるストレージ装置ST1の詳細な構成例を示す図である。

【図３】本発明の実施例１におけるストレージ装置ST1が具備する論理デバイステーブルの例を示す図である。

【図４】本発明の実施例１と実施例２におけるストレージ装置ST1が具備する内部デバイステーブルの例を示す図である。

【図５】本発明の実施例１と実施例２におけるストレージ装置ST1が具備する外部デバイステーブルの例を示す図である。

【図６】本発明の実施例１と実施例２におけるストレージ装置ST1が具備するキャッシュ割り当てテーブルの例を示す図である。

【図７】本発明の実施例１と実施例２におけるストレージ装置ST1が具備するフラッシュメモリパッケージテーブルの例を示す図である。

【図８】本発明の実施例１と実施例２におけるストレージ装置ST1が実行するデバイス管理プログラムの処理内容を示すフローチャートである。

【図９】本発明の実施例１におけるストレージ装置ST1がデバイス管理プログラムを実行中に管理計算機が表示する、内部デバイス作成画面の一例を示す図である。

【図１０】本発明の実施例１と実施例２におけるストレージ装置ST1がデバイス管理プログラムを実行中に管理計算機が表示する、論理デバイス一覧表示画面の一例を示す図である。

【図１１Ａ】本発明の実施例１と実施例２におけるストレージ装置ST1が実行するI/O制御プログラムの処理内容を示すフローチャートである。

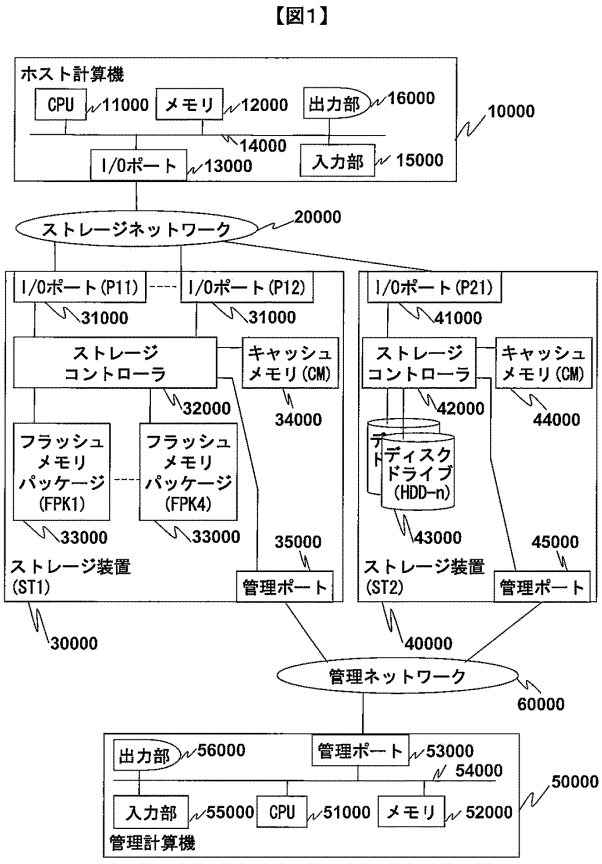
【図１１Ｂ】本発明の実施例１と実施例２におけるストレージ装置ST1が実行するI/O制御プログラムの処理内容を示すフローチャートである。

【図１２】本発明の実施例１と実施例２におけるストレージ装置ST1が具備する用途別代替領域容量テーブルの例を示す図である。

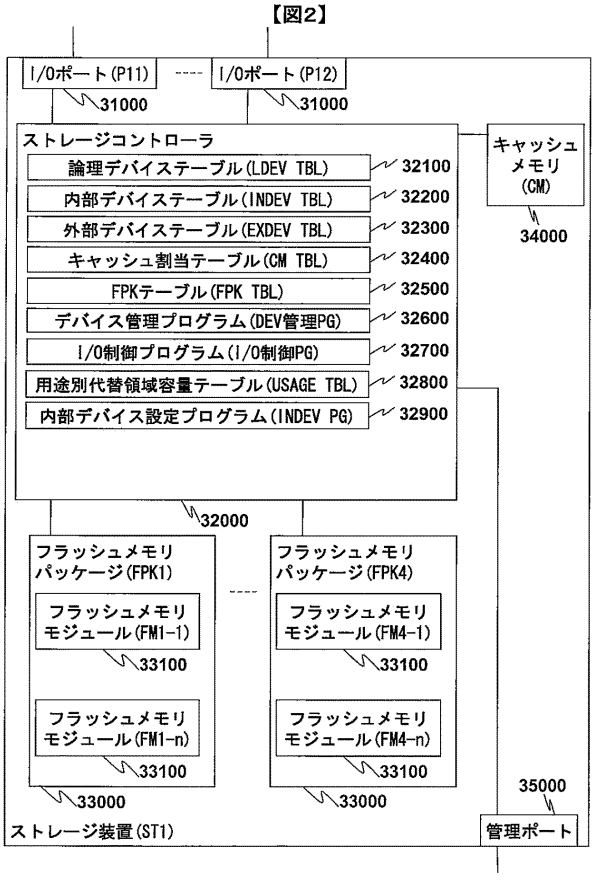
【図１３】本発明の実施例１におけるストレージ装置ST1が実行する内部デバイス設定プ

ログラムの処理内容を示すフローチャートである。
【図１４】本発明の実施例２におけるストレージ装置ST1の詳細な構成例を示す図である。
【図１５】本発明の実施例２におけるストレージ装置ST1が具備する論理デバイステーブルの例を示す図である。
【図１６】本発明の実施例２におけるストレージ装置ST1がデバイス管理プログラムを実行中に管理計算機が表示する、内部デバイス作成画面の一例を示す図である。
【図１７】本発明の実施例２におけるストレージ装置ST1が実行する内部デバイス設定プログラムの処理内容を示すフローチャートである。
【図１８】本発明の実施例２におけるストレージ装置ST1が実行する内部デバイス設定プログラムにおいて用いる計算式の一例を示す図である。
【図１９】本発明の実施例１または実施例２の変形例の計算機システムの構成例を示す図である。
【符号の説明】
【０１２４】
10000…ホスト計算機
20000…ストレージネットワーク
30000…ストレージ装置ST1
32000…ストレージコントローラ
32600…デバイス管理プログラム
32700…I/O制御プログラム
40000…ストレージ装置ST2
50000…管理計算機
60000…管理ネットワーク

【図１】



【図２】



【図３】

【図3】

32110 LDEV ID	32120 ホスト認識容量	32130 対応DEV ID	32140 状態	32150 ポート番号/ターゲットID/LUN
L01	576GB	IN01	Allocated	P11/2/0
L02	768GB	EX01	Allocated	P11/2/1
L03	360GB	IN02	Unallocated	N/A
L04	768GB	EX02	Unallocated	N/A

LDEV TBL 32100

【図４】

【図4】

32210 INDEV ID	32220 基本容量	32230 代替領域容量	32240 対応LDEV ID	32250 RAID構成
IN01	768GB	256GB	L01	RAID5(3D+1P)
IN02	480GB	32GB	L02	RAID5(3D+1P)

32260 ストライプサイズ	32270 対応FPK	32280 対応FM番号	32290 アドレス空間対応リスト (INDEVアドレス, FMメモリブロック)
256KB	FPK1-4	1	(0 to 768kB, FM1-#1), (768KB to 1536KB, FM1-#850), ... , (768GB-768KB to 768GB, FM1-#768)
256KB	FPK1-4	2	....

INDEV TBL 32200

【図５】

【図5】

32310 EXDEV ID	32320 ホスト認識容量	32330 対応LDEV ID	32340 イニシエータポートID
EX01	768GB	L02	P12
EX02	768GB	L04	P12

ストレージ識別情報		
装置ID	LDEV ID	ポート番号/ターゲットID/LUN
ST2	L11	P21/3/0
ST2	L12	P21/3/1

32350 32360 32370 EXDEV TBL 32300

【図６】

【図6】

CM TBL 32400

32410 Block ID	32420 状態	32430 対応LDEV ID	32440 LDEVアドレス
B0001	Clean	L01	0 to 768KB
B0002	Dirty	L01	768KB to 1536KB
B0003	Clean	L02	0 to 512KB
B0004	Not Used	N/A	N/A
....	....	....	....
B0512	Not Used	N/A	N/A

キャッシュBlock長 = 1024KB	32450
キャッシュ容量 = 512GB	32460

【図７】

【図7】

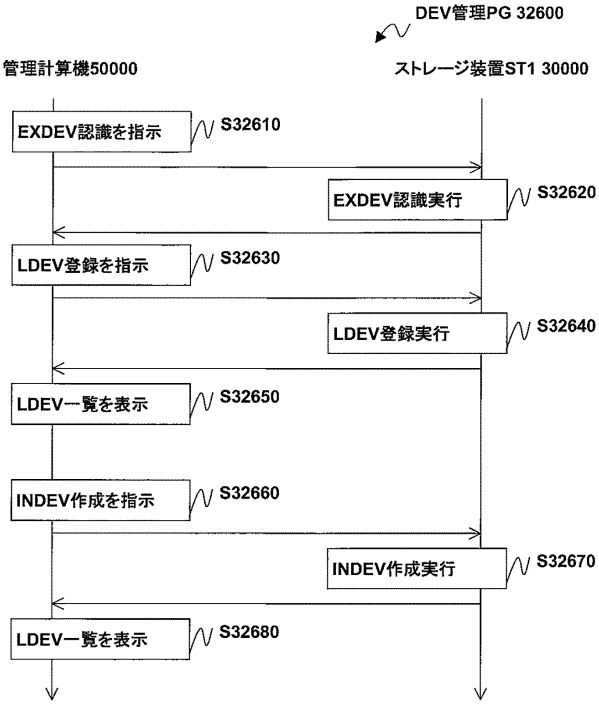
32510 FPK ID	32520 FM ID	32530 容量	32540 アドレス空間
FPK1	FM1-1	256GB	0-256GB
	FM1-2	256GB	0-128GB
			128GB-256GB
	.....	.....	.....
FPK2	FM2-1	256GB	0-256GB
	.....	.....	.....
FPK3	FM3-1	256GB	0-256GB
	.....	.....	.....
FPK4	FM4-1	256GB	0-256GB
	.....	.....	.....

32550 状態	32560 割当INDEV ID	32570 書換回数
Used	IN01	250
Used	IN02	128
Not Used	N/A	N/A
.....	.....	.....
Used	IN01	140
.....	.....	.....
Used	IN01	580
.....	.....	.....
Used	IN01	280
.....	.....	.....

FPK TBL 32500

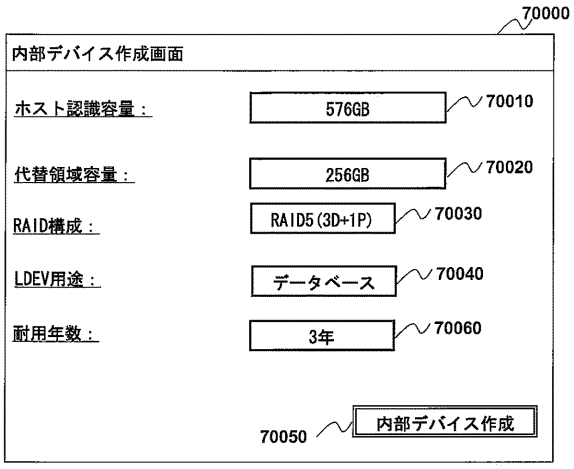
【図８】

【図8】



【図９】

【図9】



【図１０】

【図10】

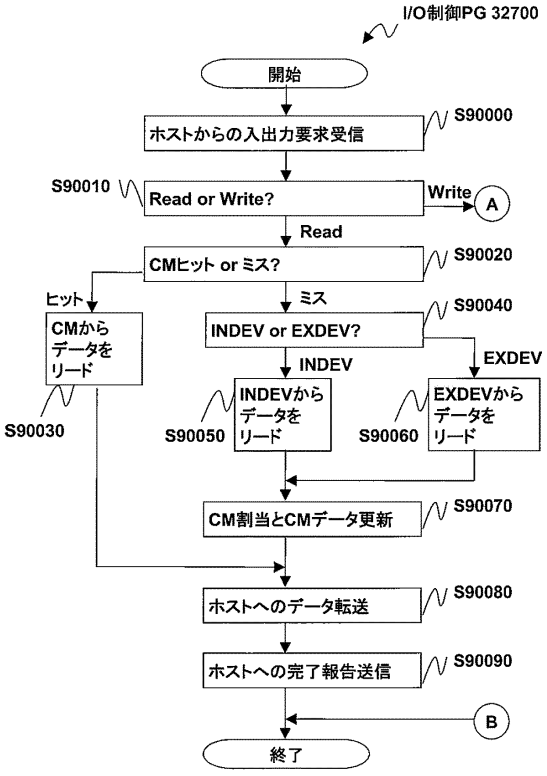
80000 論理デバイス一覧表示画面

80010 LDEV ID	80020 ホスト認識容量	80030 対応DEV ID	80040 状態	80050 ポート番号/ターゲットID/LUN
L01	576GB	IN01	Allocated	P11/2/0
L02	768GB	EX01	Allocated	P11/2/1
L03	360GB	IN02	Unallocated	N/A
L04	768GB	EX02	Unallocated	N/A

80060 基本容量	80070 代替領域容量	80080 RAID構成	80090 不良化警告
768GB	256GB	RAID5(3D+1P)	X
N/A	N/A	N/A	N/A
480GB	32GB	RAID5(3D+1P)	
N/A	N/A	N/A	N/A

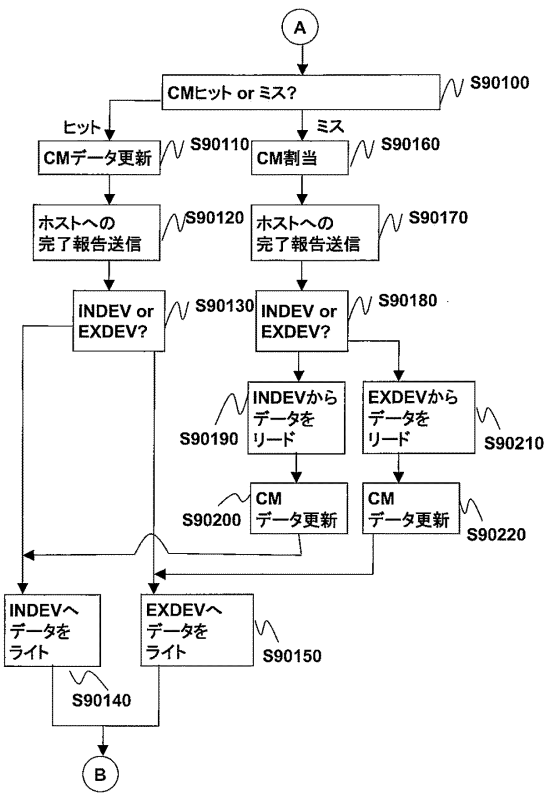
【図１１Ａ】

【図11A】



【図１１Ｂ】

【図11B】



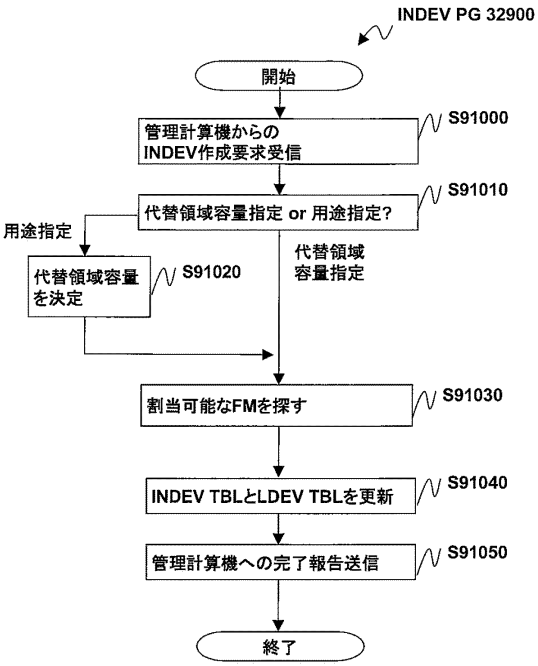
【図１２】

【図12】

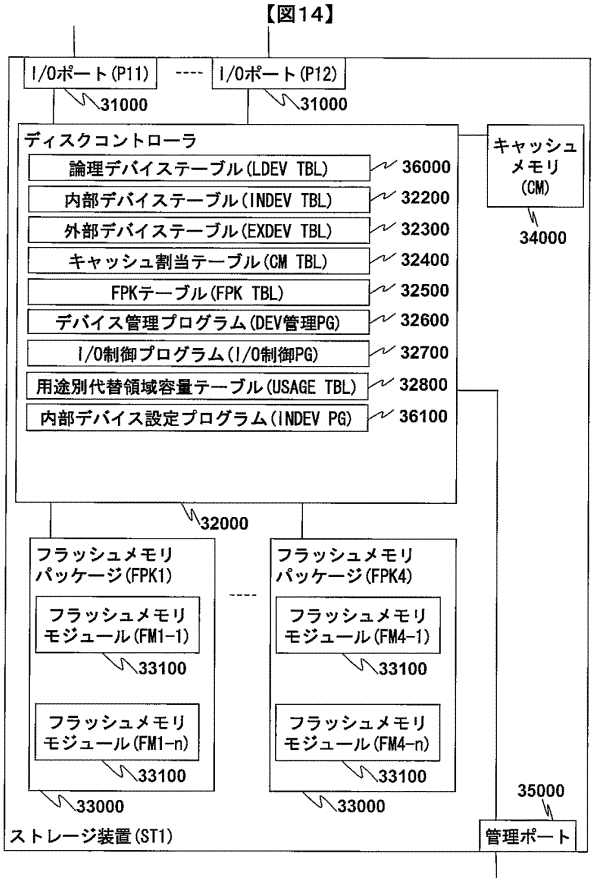
32310 USAGE ID	32320 用途	32330 年あたり代替領域容量
USAGE01	データベース	50GB
USAGE02	メール	20GB

USAGE TBL 32800

【図１３】



【図１４】



【図１５】

【図15】

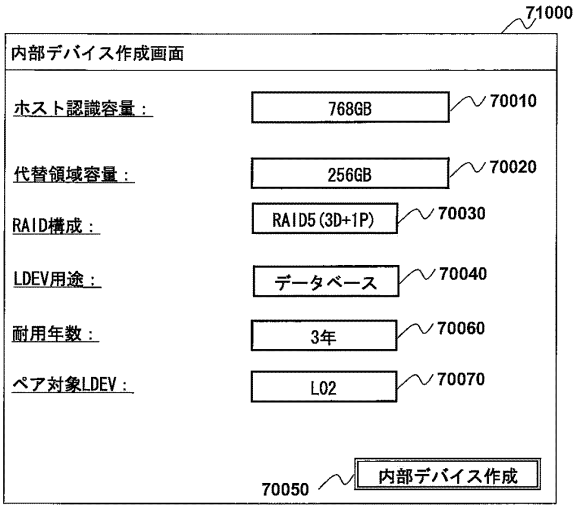
32110 LDEV ID	32120 ホスト認識容量	32130 対応DEV ID	32140 状態
L01	576GB	IN01	Allocated
L02	768GB	EX01	Allocated
L03	360GB	IN02	Unallocated
L04	768GB	EX02	Unallocated

32150 ポート番号/ターゲットID/LUN	32160 ライトアクセス頻度
P11/2/0	80 IOPS
P11/2/1	100 IOPS
N/A	90 IOPS
N/A	110 IOPS

LDEV TBL 36000

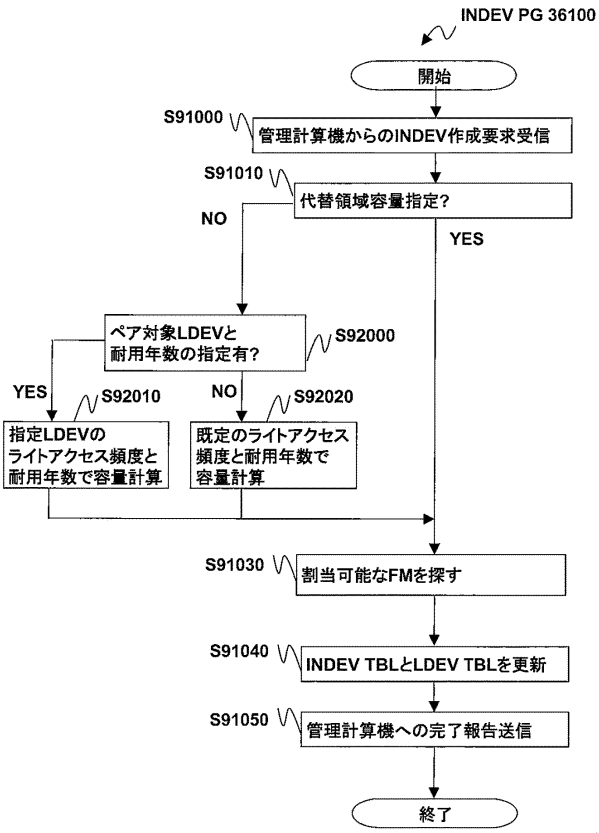
【図１６】

【図16】



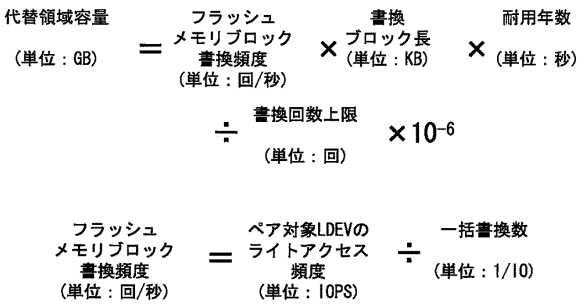
【図１７】

【図17】



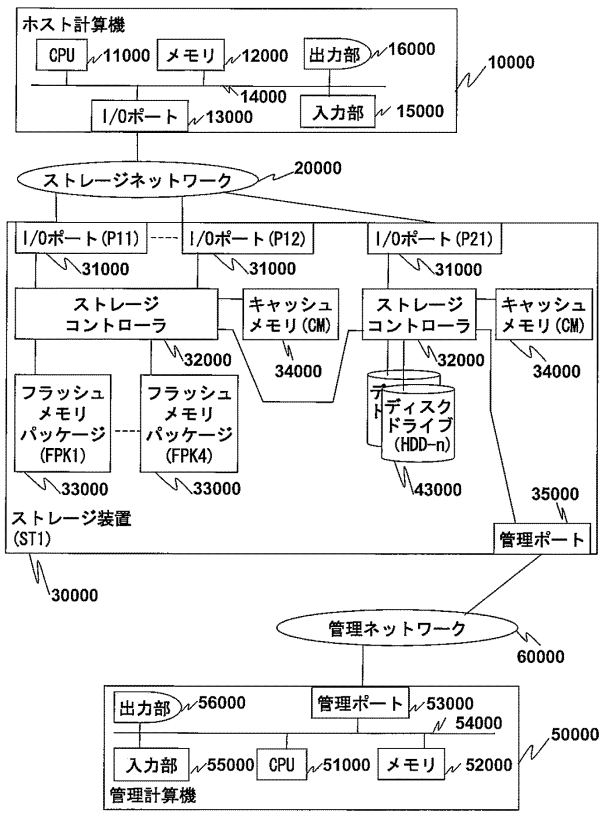
【図１８】

【図18】



【図１９】

【図19】



フロントページの続き

(51)Int.Cl. FI
G06F 3/06 306H

(72)発明者 藤林 昭
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 北原 潤
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 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

審査官 木村 貴俊

(56)参考文献 特開平０６－３３２８０６（ＪＰ，Ａ）
特開２００８－１０２９００（ＪＰ，Ａ）

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