

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7500991号
(P7500991)

(45)発行日 令和6年6月18日(2024.6.18)

(24)登録日 令和6年6月10日(2024.6.10)

(51)国際特許分類

F I

G 0 6 F	3/06 (2006.01)	G 0 6 F	3/06	3 0 1 Z
G 0 6 F	13/10 (2006.01)	G 0 6 F	3/06	3 0 1 W
G 0 6 F	16/174 (2019.01)	G 0 6 F	3/06	3 0 4 N
G 0 6 F	16/13 (2019.01)	G 0 6 F	13/10	3 4 0 A
		G 0 6 F	16/174	

請求項の数 11 (全22頁) 最終頁に続く

(21)出願番号 特願2020-28716(P2020-28716)
 (22)出願日 令和2年2月21日(2020.2.21)
 (65)公開番号 特開2021-135538(P2021-135538
 A)
 (43)公開日 令和3年9月13日(2021.9.13)
 審査請求日 令和4年11月17日(2022.11.17)

(73)特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1
 番1号
 (74)代理人 110002147
 弁理士法人酒井国際特許事務所
 (72)発明者 浦田 一宏
 神奈川県川崎市中原区上小田中4丁目1
 番1号 富士通株式会社内
 審査官 田名網 忠雄

最終頁に続く

(54)【発明の名称】 ストレージ制御装置及びストレージ制御プログラム

(57)【特許請求の範囲】

【請求項1】

ストレージ装置の物理的な格納領域に格納されたデータそれぞれに対応付けられたデータ番号と論理ボリュームにおける格納位置とを紐づける情報を保持するテーブルであって、前記論理ボリュームにおける格納位置に紐付けられた前記データ番号に対応する前記格納領域に格納されたデータが使用中データにあたる、管理テーブルと、

データの書込命令を受信し、前記格納領域に対する前記データの書込処理を実行し、前記書込処理を実行した場合、前記書込処理がデータを追加する書込命令又はデータを上書きする書込命令のいずれであるか、及び、重複データの有無を基に、前記管理テーブルに登録された論理ボリュームにおけるデータの格納位置と前記データ番号とを紐づける情報を更新し、前記管理テーブルにおいて前記論理ボリュームにおける格納位置に紐づけられた前記データ番号に対応する前記格納領域に格納されたデータの総量を前記使用中データの第1使用量として取得する書込処理部と、前記書込処理部により取得された前記第1使用量及び前記格納領域に格納された全データによる第2使用量を基に、空き容量確保処理の設定を決定し、決定した前記設定で前記空き容量確保処理を実行する容量確保実行部と

を備えたことを特徴とするストレージ制御装置。

【請求項2】

前記容量確保実行部は、前記第1使用量及び前記第2使用量に加えて、前記ストレージ装置の処理負荷を基に、空き容量確保処理の設定を決定することを特徴とする請求項1に

記載のストレージ制御装置。

【請求項 3】

前記容量確保実行部は、前記第 1 使用量と前記第 2 使用量との差を基に、前記設定を決定することを特徴とする請求項 1 又は 2 に記載のストレージ制御装置。

【請求項 4】

前記設定は、前記ストレージ装置が実行する処理における前記空き容量確保処理の割合であることを特徴とする請求項 1 ~ 3 のいずれか一つに記載のストレージ制御装置。

【請求項 5】

前記容量確保実行部は、前記第 1 使用量と前記第 2 使用量との差が閾値以上の場合に、前記空き容量確保処理の前記割合を上昇させることを特徴とする請求項 4 に記載のストレージ制御装置。

10

【請求項 6】

前記書込処理部は、前記データの書込処理において重複排除処理及び圧縮処理を実行し、既存データと重複しない場合はデータを書き込み且つ書き込んだデータに対応付けられたデータ番号と書込処理の対象とした論理ボリュームにおける格納位置とを紐づける情報を前記管理テーブルに格納し、前記既存データと重複する場合は前記既存データに対応付けられたデータ番号と書込処理の対象とした論理ボリュームにおける格納位置とを紐づける情報を前記管理テーブルに格納し、前記既存データを上書きする場合は前記既存データに対応付けられたデータ番号と論理ボリュームにおける格納位置とを紐づける情報を前記管理テーブルから削除することを特徴とする請求項 1 ~ 5 のいずれか一つに記載のストレージ制御装置。

20

【請求項 7】

前記容量確保実行部は、前記空き容量確保処理として前記格納領域における前記使用中データ以外の不要データの削除を実行することを特徴とする請求項 1 ~ 6 のいずれか一つに記載のストレージ制御装置。

【請求項 8】

前記容量確保実行部は、前記第 2 使用量が使用量閾値以上の場合、前記空き容量確保処理の前記割合を上昇させることを特徴とする請求項 4 又は 5 に記載のストレージ制御装置。

【請求項 9】

前記第 2 使用量が使用量閾値以上の場合、前記第 1 使用量と前記第 2 使用量との差を基にストレージ装置の状態を判定して報知する報知部をさらに備えたことを特徴とする請求項 1 ~ 8 のいずれか一つに記載のストレージ制御装置。

30

【請求項 10】

前記書込処理部は、複数の物理ディスクがまとめられて形成されたプールについて前記第 1 使用量を算出し、

前記容量確保実行部は、前記第 1 使用量及び前記プールに格納された全データによる前記第 2 使用量を基に、空き容量確保処理の設定を決定する

ことを特徴とする請求項 1 に記載のストレージ制御装置。

【請求項 11】

ストレージ装置の物理的な格納領域に格納されたデータそれぞれに対応付けられたデータ番号と論理ボリュームにおける格納位置とを紐づける情報を保持するテーブルであって、前記論理ボリュームにおける格納位置に紐付けられた前記データ番号に対応する前記格納領域に格納されたデータが使用中データにあたる、管理テーブルを有し、

40

データの書込命令を受信し、前記格納領域に対する前記データの書込処理を実行し、

前記書込処理を実行した場合、前記書込処理がデータを追加する書込命令又はデータを上書きする書込命令のいずれであるか、及び、重複データの有無を基に、前記管理テーブルに登録された論理ボリュームにおけるデータの格納位置と前記データ番号とを紐づける情報を更新し、前記管理テーブルにおいて前記論理ボリュームにおける格納位置に紐づけられた前記データ番号に対応する前記格納領域に格納されたデータの総量を前記使用中データの第 1 使用量として取得し、

50

取得した前記第 1 使用量及び前記格納領域に格納された全データによる第 2 使用量を基に、空き容量確保処理の設定を決定し、

決定した前記設定で前記空き容量確保処理を実行する

処理をコンピュータに実行させることを特徴とするストレージ制御プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ストレージ制御装置及びストレージ制御プログラムに関する。

【背景技術】

【0002】

ストレージ装置には、一度書き込んだデータの消去変更を禁止する追記型の記憶方式を採用した装置が存在する。さらに、追記型のストレージ装置の中には、重複排除機能及び圧縮機能を有する装置もある。

【0003】

重複解除機能及び圧縮機能を有する追記型の排除ストレージでは、重複しない新規の書込データは物理ディスクに上書きせずに追記される。また、追記型の削除や上書きが行われることで参照されなくなった物理ディスク上のデータは、ガベージコレクションと呼ばれる不要データ削除機能によってデータのインプット及びアウトプットとは非同期に、物理ディスクからの削除が行われる。このため、物理ディスクの使用量は、書き込み時に一時的に増加し、その後ガベージコレクションの動作によって減少するといった経過をたどる。

【0004】

ストレージ装置において物理ディスクの使用量は重要な性能指標であり、使用量が少ないほどデータの格納と言うストレージ装置本来の機能を十分に活用することができる。そのため、ストレージ装置では、物理ディスクの使用量はなるべく小さく抑えられることが好ましい。物理ディスクの使用量をなるべく抑えるためには、追記型のストレージ装置ではガベージコレクションを動作させることとなる。

【0005】

このようなガベージコレクションの技術として、書き込み領域が不足した時点でガベージコレクションを動作させる従来技術がある。また、新しい圧縮データを格納するための十分な大きさの空き領域が物理ディスクに存在しない場合に、ガベージコレクションを動作させる従来技術がある。さらに、物理ディスクにおける未使用領域が一定値以下になり、ホスト装置からのアクセスが一定時間到来しない場合に、ガベージコレクションを実行する従来技術がある。

【先行技術文献】

【特許文献】

【0006】

【文献】国際公開第 2015/097739 号

【文献】特開平 7 - 129470 号公報

【文献】特開平 9 - 330185 号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、ストレージ装置では、ガベージコレクションを動作させた場合の負荷による性能影響が大きい。そのため、ガベージコレクションを頻繁に実行することはなるべく避けることが好ましい。

【0008】

一方で、ガベージコレクションの実行頻度を低下させると、物理ディスクの使用量が実際に使用可能なデータに比べて大きくなってしまおうという問題がある。また、ガベージコレクションを実施しなければ、削除すべきデータが存在するかどうか判定困難であり、ガ

10

20

30

40

50

ページコレクションの頻度を下げた場合、不要なデータの増加に気づかず、物理ディスク上に無駄な領域が増えてしまう。さらに、参照されない不要なデータを除いたディスク実使用量もガベージコレクションを実行しないと不明であり、物理ディスクの容量不足などの事態の発生を迅速に把握することが困難となる。これらの事態が発生すると、ストレージ装置において記憶領域の確保が不十分となり、装置性能を向上させることが困難となる。

【0009】

この点、書き込み領域が不足した時点でガベージコレクションを動作させる従来技術では、書き込み領域が不足する以前に不要なデータの増加を検出することは困難であり、ガベージコレクションの実行が遅れる場合がある。その場合、ストレージ装置の装置性能を向上させることが困難となるおそれがある。これは、新しい圧縮データの格納領域の存否に応じてガベージコレクションを動作させる従来技術や、物理ディスクにおける未使用領域及びアクセス頻度に応じてガベージコレクションを実行する従来技術でも同様である。

10

【0010】

開示の技術は、上記に鑑みてなされたものであって、ストレージ装置の装置性能を向上させるストレージ制御装置及びストレージ制御プログラムを提供することを目的とする。

【課題を解決するための手段】

【0011】

本願の開示するストレージ制御装置及びストレージ制御プログラムの一つの態様において、管理テーブルは、ストレージ装置の物理的な格納領域に格納されたデータそれぞれに対応付けられたデータ番号と論理ボリュームにおける格納位置とを紐づける情報を保持するテーブルであって、前記論理ボリュームにおける格納位置に紐付けられた前記データ番号に対応する前記格納領域に格納されたデータが使用中データにあたる。書込処理部は、データの書込命令を受信し、前記格納領域に対する前記データの書込処理を実行し、前記書込処理を実行した場合、前記書込処理がデータを追加する書込命令又はデータを上書きする書込命令のいずれであるか、及び、重複データの有無を基に、前記管理テーブルに登録された論理ボリュームにおけるデータの格納位置と前記データ番号とを紐づける情報を更新し、前記管理テーブルにおいて前記論理ボリュームにおける格納位置に紐づけられた前記データ番号に対応する前記格納領域に格納されたデータの総量を前記使用中データの第1使用量として取得する。容量確保実行部は、前記書込処理部により取得された前記第1使用量及び前記格納領域に格納された全データによる第2使用量を基に、空き容量確保処理の設定を決定し、決定した前記設定で前記空き容量確保処理を実行する。

20

30

【発明の効果】

【0012】

1つの側面では、本発明は、ストレージ装置の記装置性能を向上させることができる。

【図面の簡単な説明】

【0013】

【図1】図1は、ストレージシステムのハードウェア構成図である。

【図2】図2は、実施例1に係るコントローラモジュールのブロック図である。

【図3】図3は、論理ボリューム側管理テーブルの一例を表す図である。

【図4】図4は、物理ボリューム側管理テーブルの一例を表す図である。

40

【図5】図5は、新規データ書き込み時の管理テーブルの遷移を表す図である。

【図6】図6は、重複データ書き込み時の管理テーブルの遷移を表す図である。

【図7】図7は、データ上書き時の管理テーブルの遷移を表す図である。

【図8】図8は、ガベージコレクション処理が割り当てられない場合の処理割り当てを表す図である。

【図9】図9は、ガベージコレクション処理の優先度が通常の場合の処理割り当てを表す図である。

【図10】図10は、ガベージコレクション処理の優先度が高い場合の処理割り当てを表す図である。

【図11】図11は、ガベージコレクション処理の全体のフローチャートである。

50

【図 1 2】図 1 2 は、プール使用量算出処理のフローチャートである。

【図 1 3】図 1 3 は、実施例 1 に係る優先度設定処理のフローチャートである。

【図 1 4】図 1 4 は、実施例 2 に係るコントローラモジュールのブロック図である。

【図 1 5】図 1 5 は、実施例 2 に係る優先度設定処理のフローチャートである。

【発明を実施するための形態】

【0014】

以下に、本願の開示するストレージ制御装置及びストレージ制御プログラムの実施例を図面に基づいて詳細に説明する。なお、以下の実施例により本願の開示するストレージ制御装置及びストレージ制御プログラムが限定されるものではない。

【実施例 1】

【0015】

図 1 は、ストレージシステムのハードウェア構成図である。図 1 に示すように、ストレージシステム 1 は、サーバなどのホスト 2 と接続される。そして、ストレージシステム 1 は、コントローラモジュール (Controller Module) 1 0 及びディスク (Disk) 2 0 を有する。

【0016】

ホスト 2 は、ストレージシステム 1 に対して命令を送信する。ストレージシステム 1 は、ホスト 2 から受信した命令を処理し、その命令に対する応答をホスト 2 へ返す。ホスト 2 からの命令は、データの書込命令や読出命令などがなどである。データの書込命令には、ストレージシステム 1 が保持しないデータを書き込む新規データの書込命令及び既にストレージシステム 1 が保持する既存データに重複するデータを書き込む重複データの書込命令が含まれる、さらに、書込命令には、既にストレージシステム 1 が保持する既存データを更新する上書命令が含まれる。

【0017】

コントローラモジュール 1 0 は、ディスク 2 0 の論理構成の生成やディスク 2 0 に対するデータの読み出し及び書き込みを実行するストレージ制御装置である。コントローラモジュール 1 0 は、チャンネルアダプタ (Channel Adapter) 1 1、CPU (Central Processing Unit) 1 2、DRAM (Dynamic Random Access Memory) 1 3、ディスクインタフェース 1 4 を有する。

【0018】

チャンネルアダプタ 1 1 は、ホスト 2 に接続されるホスト 2 との間の通信インタフェースである。チャンネルアダプタ 1 1 は、CPU 1 2 に接続され、ホスト 2 から受信した命令を CPU 1 2 へ出力する。また、チャンネルアダプタ 1 1 は、ホスト 2 から受信した命令に対する応答を CPU 1 2 から受信する。そして、チャンネルアダプタ 1 1 は、受信した応答をホスト 2 へ送信する。

【0019】

CPU 1 2 は、ホスト 2 から送信された命令の入力をチャンネルアダプタ 1 1 から受ける。そして、CPU 1 2 は、受信した命令を処理する。例えば、CPU 1 2 は、ディスクインタフェース 1 4 を介してディスク 2 0 にアクセスし、データの書き込みや読み出しの処理を実行する。そして、CPU 1 2 は、処理結果を命令に対する応答としてチャンネルアダプタ 1 1 を介してホスト 2 へ送信する。また、CPU 1 2 は、複数のディスク 2 0 をまとめてプール 2 0 0 を形成する。このプール 2 0 0 が、「格納領域」の一例にあたる。さらに、CPU 1 2 は、プール 2 0 0 の中にディスク 2 0 をまとめた論理構成を構築する。例えば、CPU 1 2 は、複数のディスク 2 0 を使用して RAID (Redundant Arrays of Inexpensive Disks) を構築して論理ボリュームを構成する。

【0020】

CPU 1 2 は、実際には物理的なディスクであるディスク 2 0 に対してデータの書き込み及び読み出しを行う。例えば、CPU 1 2 は、ホスト 2 からの命令では論理的なディスクであるボリュームに対しての書き込みや読み出しが指示される。そこで、CPU 1 2 は、ホスト 2 からの命令で指定されたボリュームにおけるアクセス先の情報をディスク 2 0

10

20

30

40

50

におけるアドレスに変換して、ディスク 20 に対して書込処理や読出処理を実行する。すなわち、書込処理や読出処理は、ホスト 2 からは論理ボリュームに対する処理として指定され、コントローラモジュール 10 により実際のデータは物理ボリュームであるディスク 20 に格納される。

【0021】

また、CPU 12 は、ストレージシステム 1 の制御プログラムを DRAM 13 上に展開して実行する。ストレージシステム 1 の制御プログラムとしては、例えば、ガベージコレクションなどを動作させるためのプログラムなどがある。

【0022】

DRAM 13 は、主記憶装置である。DRAM 13 は、ストレージシステム 1 におけるキャッシュとしても使用される。

10

【0023】

ディスクインタフェース 14 は、ディスク 20 との間の通信インタフェースである。ディスクインタフェース 14 は、CPU 12 とディスク 20 との間のデータの送受信を仲介する。

【0024】

ディスク 20 は、ハードディスクなどの物理ディスクであり補助記憶装置である。ディスク 20 は、複数まとめて 1 つのプール 200 を形成する。さらに、ディスク 20 はコントローラモジュール 10 により論理的構成が構築される。例えば、複数のディスク 20 を用いて 1 つの論理ボリュームが構築される。

20

【0025】

次に、図 2 を参照して、コントローラモジュール 10 の詳細について説明する。図 2 は、実施例 1 に係るコントローラモジュールのブロック図である。

【0026】

コントローラモジュール 10 は、CPU 12 により実現される、重複圧縮制御部 102、キャッシュメモリ制御部 104 及びバックエンド制御部 105 を有する。また、DRAM 13 には、メタデータテーブル 103 が格納される。

【0027】

メタデータテーブル 103 は、図 3 に示す論理ボリューム側の管理テーブル 131 及び図 4 に示す物理ボリューム側の管理テーブル 132 を含む。図 3 は、論理ボリューム側管理テーブルの一例を表す図である。また、図 4 は、物理ボリューム側管理テーブルの一例を表す図である。

30

【0028】

管理テーブル 131 は、論理的なディスクである論理ボリュームにおけるデータの格納位置を表すテーブルである。管理テーブル 131 には、図 3 に示すように、論理ボリューム LBA (Logical Block Addressing) とその論理ボリューム LBA で示される領域に格納されたデータの識別情報であるデータ番号とが対応付けられて格納される。管理テーブル 131 に登録されたデータ番号により、物理ボリューム側の管理テーブル 132 を介して、ディスク 20 の集まりである物理ボリューム上のどのデータを参照しているかが確認可能となる。

40

【0029】

管理テーブル 132 は、物理ディスクであるディスク 20 におけるデータの格納位置を表すテーブルである。管理テーブル 132 には、図 4 に示すように、データ番号、参照カウンタ、物理ディスクアドレス及びデータサイズが対応付けられて格納される。データ番号は、論理ボリューム側の管理テーブル 131 に格納されたデータ番号が用いられる。参照カウンタは、そのデータが参照される参照数を表す。本実施例に係るストレージシステム 1 は重複排除機能を有するため、1 つのデータが複数の異なる情報として参照される場合がある。物理ディスクアドレスは、そのデータが格納されたディスク 20 におけるアドレスを表す。そして、管理テーブル 132 の物理ディスクアドレスで指定されたディスク 20 上の領域にデータ番号に対応するデータが格納される。格納データ 210 は、管理テ

50

プール 1 3 2 に登録された各情報に対応するディスク 2 0 に格納された実際のデータを表す。

【 0 0 3 0 】

重複圧縮制御部 1 0 2 は、入出力制御部 1 2 1 及びガベージコレクション制御部 1 2 2 を有する。入出力制御部 1 2 1 は、論理ボリューム L B A を用いて指定されることで参照されているデータ、すなわち使用中のデータによるプール 2 0 0 の使用量を表す使用中データ使用量を保持する。言い換えれば、プール使用量は、プール 2 0 0 の記憶領域に格納された全データから参照されないデータである不要データを除いたデータの使用量である。このプール使用量が、「第 1 使用量」の一例にあたる。ここで、本実施例では、プール 2 0 0 を基準として記憶領域の使用量を算出しているが、データを格納する記憶領域を対象とするものであれば他の記憶領域を基準としてもよく、例えば、論理ボリュームを基準としてもよい。入出力制御部 1 2 1 は、プール 2 0 0 の作成時にプール使用量を初期化して 0 にする。

10

【 0 0 3 1 】

入出力制御部 1 2 1 は、チャンネルアダプタ 1 1 を介してホスト 2 から送信された命令の入力を受ける。そして、入出力制御部 1 2 1 は、取得した命令を処理する。以下に入出力制御部 1 2 1 の命令処理の動作について説明する。

【 0 0 3 2 】

読出命令の場合、入出力制御部 1 2 1 は、メタデータテーブル 1 0 3 を参照して、読み出す対象のデータの格納先を特定する。そして、入出力制御部 1 2 1 は、キャッシュメモリ制御部 1 0 4 に特定した格納先のデータの読み出しを依頼する。その後、入出力制御部 1 2 1 は、読み出し対象のデータの入力をキャッシュメモリ制御部 1 0 4 から受ける。そして、入出力制御部 1 2 1 は、取得したデータをホスト 2 へチャンネルアダプタ 1 1 を介して送信する。

20

【 0 0 3 3 】

書込命令の場合、入出力制御部 1 2 1 は、既存データに対する上書き命令であるか、又は、データを追加する書込み命令であるかを判定する。さらに、データを追加する書込み命令である場合、入出力制御部 1 2 1 は、書込み対象のデータが既存データと重複しない新規データであるか又は重複する重複データであるかを判定する。

【 0 0 3 4 】

データを追加する書込命令であって書き込み対象のデータが新規データである場合、入出力制御部 1 2 1 は、書き込み対象の新規データのディスク 2 0 における格納先を決定する。次に、入出力制御部 1 2 1 は、書き込み対象の新規データを圧縮しキャッシュメモリ制御部 1 0 4 へ出力して決定した格納先への格納を依頼する。さらに、入出力制御部 1 2 1 は、メタデータテーブル 1 0 3 を更新する。この場合のメタデータテーブル 1 0 3 の詳細を以下に説明する。

30

【 0 0 3 5 】

図 5 は、新規データ書き込み時の管理テーブルの遷移を表す図である。ここでは、新規データを書き込む以前の管理テーブル 1 3 1 及び 1 3 2 の状態が、図 3 及び図 4 で示される状態であった場合で説明する。

40

【 0 0 3 6 】

入出力制御部 1 2 1 は、論理ボリューム側の管理テーブル 1 3 1 の行 3 0 1 に、新規データのデータ番号とともにその新規データの論理ボリューム L B A を登録する。また、入出力制御部 1 2 1 は、物理ボリューム側の管理テーブル 1 3 2 に新規データのための新しい行 3 0 2 を作成し、データ番号を登録するとともに、物理ディスクアドレス及びデータサイズを格納する。さらに、入出力制御部 1 2 1 は、新たに追加した論理ボリューム L B A により格納した新規データが参照されるので、管理テーブル 1 3 2 における新規データの行 3 0 2 の参照カウンタを 1 に設定する。この場合、行 3 0 2 の新規データの情報に対応するデータ 2 1 1 が格納データ 2 1 0 として物理ボリュームに格納される。

【 0 0 3 7 】

50

この場合、参照されているデータである新規データがプール 200 に追加格納されるので、入出力制御部 121 は、プール使用量に新規データによる使用量を加算する。

【0038】

一方、データを追加する書込命令であって書き込み対象のデータが重複データである場合、入出力制御部 121 は、書き込み対象の重複データのディスク 20 における格納先を決定する。次に、入出力制御部 121 は、決定した格納先に対する重複した既存データを示す情報の格納をキャッシュメモリ制御部 104 へ出力する。その後、入出力制御部 121 は、書き込み完了の応答をキャッシュメモリ制御部 104 から受ける。そして、入出力制御部 121 は、チャンネルアダプタ 11 を介して Host 2 へ書き込み完了の応答を送信する。また、入出力制御部 121 は、メタデータテーブル 103 を更新する。メタデータテーブル 103 を更新する。この場合のメタデータテーブル 103 の詳細を以下に説明する。

10

【0039】

図 6 は、重複データ書き込み時の管理テーブルの遷移を表す図である。ここでは、重複データを書き込む以前の管理テーブル 131 及び 132 の状態が図 5 で示される状態であった場合で説明する。

【0040】

入出力制御部 121 は、論理ボリューム側の管理テーブル 131 を参照して、重複する既存データである元データを表す行を特定する。そして、入出力制御部 121 は、特定した行のデータ番号を表す欄 304 から元データのデータ番号を取得する。次に、入出力制御部 121 は、論理ボリューム側の管理テーブル 131 の新たな行 303 に、重複データのデータ番号として元データのデータ番号を登録し、且つ、重複データの論理ボリューム LBA を登録する。また、入出力制御部 121 は、物理ボリューム側の管理テーブル 132 における元データを表す行を特定する。そして、入出力制御部 121 は、元データに対して今回格納したアドレスからの参照が 1 つ増えたので、特定した行の参照カウンタの欄 305 の値を 1 つインクリメントする。この場合、格納データ 210 には重複データの新たな格納は行われない。

20

【0041】

この場合、重複データによるプール 200 の使用量の増加は発生しないため、入出力制御部 121 は、プール使用量をそのままの値で維持する。

【0042】

これに対して、データを上書きする書込命令である場合、入出力制御部 121 は、更新データについては、新規データか重複データかを判定し、それぞれの場合について上述した方法でデータの格納、並びに、管理テーブル 131 及び 132 の更新を実行する。一方、上書きされる元データについては、入出力制御部 121 は、メタデータテーブル 103 を参照して上書き対象の元データの情報を管理テーブル 132 の中から特定する。そして、入出力制御部 121 は、管理テーブル 132 における上書き対象の元データの参照カウンタを 1 つデクリメントする。この場合のメタデータテーブル 103 の詳細を以下に説明する。

30

【0043】

図 7 は、データ上書き時の管理テーブルの遷移を表す図である。ここでは、データの上書きを行う以前の管理テーブル 131 及び 132 の状態が図 6 で示される状態であった場合で説明する。図 7 は、更新データが重複データの場合の上書きを表す。

40

【0044】

入出力制御部 121 は、論理ボリューム側の管理テーブル 131 を参照して、上書き対象の元データを表す行を特定する。以降の処理は、更新データが重複データか新規データかで処理が異なる。

【0045】

更新データが重複データの場合、入出力制御部 121 は、特定した行のデータ番号を表す欄 306 のデータ番号に、更新データが重複する元データのデータ番号を登録する。また、入出力制御部 121 は、物理ボリューム側の管理テーブル 132 における更新データ

50

が重複する元データを表す行を特定する。そして、入出力制御部 1 2 1 は、更新データが重複する元データに対して今回の更新データのアドレスからの参照が 1 つ増えるので、特定した行の参照カウンタの欄 3 0 8 における値を 1 つインクリメントする。この場合、格納データ 2 1 0 には更新データの新たな格納は行われない。

【 0 0 4 6 】

この場合、更新データによるプール 2 0 0 の使用量の増加は発生しないため、入出力制御部 1 2 1 は、プール使用量をそのままの値で維持する。

【 0 0 4 7 】

これに対して、更新データが新規データの場合、入出力制御部 1 2 1 は、新たにデータ番号を割り当てて論理ボリューム側の管理テーブル 1 3 1 に新たに更新データの情報を登録する。また、入出力制御部 1 2 1 は、更新データの情報を物理ボリューム側の管理テーブル 1 3 2 にも登録する。

10

【 0 0 4 8 】

この場合、参照されているデータである新規データがプール 2 0 0 に追加格納されるので、入出力制御部 1 2 1 は、プール使用量に新規データによる使用量を加算する。

【 0 0 4 9 】

さらに、更新データが新規データ又は重複データのいずれの場合においても、入出力制御部 1 2 1 は、以下の処理を実行する。入出力制御部 1 2 1 は、上書き対象の元データを表す行を特定する。そして、入出力制御部 1 2 1 は、上書き対象の元データに対して今回の更新データのアドレスからの参照が 1 つ減るので、特定した行の参照カウンタの欄 3 0 7 における値を 1 つデクリメントする。その後、入出力制御部 1 2 1 は、上書き対象の元データの参照カウンタが 0 か否かを判定する。

20

【 0 0 5 0 】

参照カウンタが 0 でなければそのデータはいずれかの論理ボリューム L B A を用いて参照されているので、入出力制御部 1 2 1 は、上書き対象の元データを使用中のデータであると判定する。この場合、入出力制御部 1 2 1 は、プール使用量をそのまま維持する。これに対して、上書き対象の元データの参照カウンタが 0 である場合、入出力制御部 1 2 1 は、上書き対象の元データは参照されておらず不要なデータであると判定する。この場合、上書き対象の元データが不要なデータとなったため、入出力制御部 1 2 1 は、プール使用量から上書き対象の元データによる使用量を減算する。

30

【 0 0 5 1 】

また、入出力制御部 1 2 1 は、例えば、ホスト 2 からプール使用量の通知要求を受けた場合、保持するプール使用量の情報をホスト 2 へチャネルアダプタ 1 1 を介して送信する。これにより、管理者は、プール使用量を確認することができ、ある時点での圧縮重複排除後の使用中のデータ量を判断することができる。

【 0 0 5 2 】

図 2 に戻って説明を続ける。ガベージコレクション制御部 1 2 2 は、ガベージコレクションの定期実行を判定するためのタイマを有する。そして、ガベージコレクション制御部 1 2 2 は、タイマを用いてガベージコレクションの定期実行のタイミングの到来を検知して、ガベージコレクションの実行を開始する。ここで、本実施例では、ガベージコレクションは定期実行されるが、不定期の実行でもよい。例えば、プール 2 0 0 の使用量に基づいてガベージコレクションを実行してもよいし、管理者からの指示によりガベージコレクションを実行してもよい。

40

【 0 0 5 3 】

ガベージコレクション制御部 1 2 2 は、ガベージコレクションを実行する場合に、ガベージコレクションの設定を決定して、決定した設定に基づいてガベージコレクションを実行する。本実施例では、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの設定として、ストレージシステム 1 で実行される処理全体におけるガベージコレクション処理が実行される割合を表す優先度を用いる。すなわち、本実施例に係る特定の処理の優先度とは、優先度が高くなるほど、ストレージシステム 1 で実行される処理全体における

50

その特定の処理が実行される割合が上昇することを表す指標である。以下に、ガベージコレクション制御部 1 2 2 によるガベージコレクション処理の詳細について説明する。

【 0 0 5 4 】

ガベージコレクション制御部 1 2 2 は、ストレージシステム 1 のシステム負荷が閾値以下か否かを判定する。システム負荷が閾値より大きい場合、ガベージコレクションを優先的に処理するための処理能力や資源の余裕をストレージシステム 1 が有さないと考えられるため、ガベージコレクション制御部 1 2 2 は、ガベージコレクション処理の優先度を通常に設定する。

【 0 0 5 5 】

これに対して、システム負荷が閾値以下の場合、ガベージコレクション制御部 1 2 2 は、入出力制御部 1 2 1 からプール使用量を取得する。また、ガベージコレクション制御部 1 2 2 は、プール 2 0 0 に格納された全てのデータによる使用量であるディスク実使用量をバックエンド制御部 1 0 5 から取得する。このディスク実使用量が、「第 2 使用量」の一例にあたる。

【 0 0 5 6 】

次に、ガベージコレクション制御部 1 2 2 は、ディスク実使用量からプール使用量を減算する。そして、ガベージコレクション制御部 1 2 2 は、ディスク実使用量とプール使用量との差を表す減算結果が閾値以上か否かを判定する。

【 0 0 5 7 】

ディスク実使用量とプール使用量との差が閾値未満の場合、不要データが少ないと考えられ、ガベージコレクションを実行しても未使用領域の増加はそれほど見込めない。そのため、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの優先度を通常に設定する。

【 0 0 5 8 】

これに対して、ディスク実使用量とプール使用量との差が閾値以上の場合、不要データが多いと考えられ、ガベージコレクションを実行することで未使用領域のある程度の増加が見込まれる。そのため、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの優先度を高くする。本実施例では、ガベージコレクションの優先度として、通常の優先度及び高い優先度の 2 種類が存在する場合で説明する。

【 0 0 5 9 】

その後、ガベージコレクション制御部 1 2 2 は、設定した優先度でガベージコレクションを CPU コアに割り当てるとともにバックエンド制御部 1 0 5 に実行させる。ここで、本実施例に係る処理の優先度について説明する。

【 0 0 6 0 】

本実施例に係る優先度設定は、ストレージシステム 1 におけるタスクスケジューラによる CPU 1 2 のコア割り当ての優先度やバックエンド制御部 1 0 5 によるディスクへのコマンド発行優先度に反映される。

【 0 0 6 1 】

ストレージシステム 1 に搭載される CPU 1 2 は、複数のコアを有する。そして、タスクスケジューラと呼ばれる制御が、ストレージシステム 1 が実行する処理を、各コアに割り当てて実行させる。この割り当ての際に、タスクスケジューラは、特定のタスクについて割り当てるコアを固定したり、あるいは優先度の高い処理を優先度が低い処理より先に実行させたりする。

【 0 0 6 2 】

例えば、CPU 1 2 がコア # 1 ~ # 9 を有し、コア # 1 ~ # 9 が、ホスト 2 からの読み出し命令及び書き込み命令を処理する I/O 処理及びガベージコレクション処理を実行する場合で説明する。例えば、ガベージコレクションが実行されていない場合、コア # 1 ~ # 9 には、図 8 のように I/O 処理が割り当てられる。図 8 は、ガベージコレクション処理が割り当てられない場合の処理割り当てを表す図である。図 8 における実行中処理にあたる処理が、各コア # 1 ~ # 9 が実行している処理である。そして、実行待ち処理行列にあ

10

20

30

40

50

る処理が、各コア # 1 ~ # 9 のそれぞれに既に割り当てられて実行中処理が終了すると順次紙面に向かって上から処理されていく処理である。

【 0 0 6 3 】

図 9 は、ガベージコレクション処理の優先度が通常の場合の処理割り当てを表す図である。図 9 において C G 処理と記載したものがガベージコレクション処理にあたる。ガベージコレクションの優先度が通常の場合、例えば、コア # 9 にガベージコレクション処理を割り当て、残りのコア # 1 ~ # 8 は、I O 処理を割り当てる。他にも、ガベージコレクション処理が通常の場合の優先度に設定された場合に、I O 処理がガベージコレクション処理よりも優先して実行され、ガベージコレクション処理は I O 処理が空いたタイミングで実施されるように設定してもよい。このように、通常の場合の優先度にガベージコレクションが設定された場合、I O 処理を阻害することなくガベージコレクション処理が実行される。

10

【 0 0 6 4 】

図 1 0 は、ガベージコレクション処理の優先度が高い場合の処理割り当てを表す図である。ガベージコレクション処理の優先度が高い場合、ガベージコレクション処理は I O 処理と均等に各コア # 1 ~ # 9 に割り当てられる。すなわち、平均すると、5 つのコアで I O 処理が実施され、残りの 5 つのコアでガベージコレクション処理が実施されることになる。この場合、I O 処理又はガベージコレクション処理のうち、先に登録されたものが先に実行される。これによって、ガベージコレクション処理の処理速度は通常時と比べて大きく向上する。逆に、I O 処理は実行がある程度阻害される。ただし、コア # 1 ~ # 9 に対する処理割り当ては固定されないため、I O 処理が無い場合には、全てのコア # 1 ~ # 9 でガベージコレクション処理が動作することが可能である。逆に、ガベージコレクション処理が無い場合には、全てのコア # 1 ~ # 9 で I O 処理が動作することが可能である。

20

【 0 0 6 5 】

ここで、C P U 1 2 のそれぞれのコアは、入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4、バックエンド制御部 1 0 5 及びディスクインタフェース 1 4 の機能をそれぞれが実現する。すなわち、ガベージコレクション制御部 1 2 2 は、各コアで動作する入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4、バックエンド制御部 1 0 5 及びディスクインタフェース 1 4 に対して設定した優先度を通知して処理を実行させるといえる。このように、本実施例において優先度を上げるということは、具体的には、ガベージコレクションの実行の設定を変更して、コントローラモジュール 1 0 が実行する処理全体のうちのガベージコレクション処理の割合を上昇させることにあたる。

30

【 0 0 6 6 】

さらにタスクスケジューリングに加えて、本実施例に係るストレージシステム 1 では、バックエンド制御部 1 0 5 が実行するディスク 2 0 に対するデータの流量制御時の割合にも優先度が反映される。バックエンド制御部 1 0 5 は、R A I D グループを構成する複数のディスク 2 0 へのコマンド発行の際に、ガベージコレクション処理に対する延長コマンドをどの程度発行するかを優先度に応じて決定する。ガベージコレクション処理の優先度が通常の場合、バックエンド制御部 1 0 5 は、I O 処理のコマンドの発行を優先して行い、I O 処理のコマンド発行後にガベージコレクション処理の延長コマンドを発行する。これに対して、ガベージコレクション処理の優先度が高い場合、バックエンド制御部 1 0 5 は、ガベージコレクション処理の延長コマンドを I O 処理のコマンドと均等に発行する。

40

【 0 0 6 7 】

このガベージコレクション制御部 1 2 2 が、「容量確保実行部」の一例にあたる。また、ガベージコレクション制御部 1 2 2 により実行されるガベージコレクションが、「空き容量確保処理」の一例にあたる。ただし、プール使用量とディスク実使用量との差を基に実行させる空き容量確保処理は、ディスク 2 0 の空き容量を増やすことができれば他の処理を対象としてもよい。

【 0 0 6 8 】

以上の説明では、ガベージコレクションの優先度として、通常の場合の優先度及び高い優先度の 2 種類が存在する場合で説明したが、優先度は通常の場合の優先度から最高の優先度まで複数

50

段階あってもよい。ガベージコレクション制御部 1 2 2 は、ガベージコレクションの優先度を高くする場合、複数段階の優先度のうち通常の優先度よりも高い優先度を選択して設定する。ガベージコレクション制御部 1 2 2 は、この優先度の選択を、ディスク実容量の大きさやディスク実容量とプール使用量との差の大きさに応じて選択してもよい。

【 0 0 6 9 】

図 2 に戻って説明を続ける。キャッシュメモリ制御部 1 0 4 は、データの書込命令の入力を入出力制御部 1 2 1 から受ける。そして、キャッシュメモリ制御部 1 0 4 は、書き込み対象のデータを D R A M 1 3 のキャッシュ領域に書き込み、書き込み完了の応答を入出力制御部 1 2 1 へ出力する。その後、キャッシュメモリ制御部 1 0 4 は、非同期で D R A M 1 3 のキャッシュ領域から書き込み対象のデータを読み出して、バックエンド制御部 1 0 5 に書込命令を出力する。

10

【 0 0 7 0 】

また、キャッシュメモリ制御部 1 0 4 は、データの読出命令の入力を入出力制御部 1 2 1 から受ける。そして、キャッシュメモリ制御部 1 0 4 は、D R A M 1 3 のキャッシュ領域に読み出し対象のデータが存在するか否かを確認する。キャッシュヒットした場合、キャッシュメモリ制御部 1 0 4 は、D R A M 1 3 のキャッシュ領域から読み出し対象のデータを読み出して入出力制御部 1 2 1 へ出力する。

【 0 0 7 1 】

これに対して、キャッシュミスヒットの場合、キャッシュメモリ制御部 1 0 4 は、データの読出命令をバックエンド制御部 1 0 5 へ出力する。その後、キャッシュメモリ制御部 1 0 4 は、読み出し対象のデータの入力をバックエンド制御部 1 0 5 から受ける。そして、キャッシュメモリ制御部 1 0 4 は、取得した読み出し対象のデータを D R A M 1 3 のキャッシュ領域に格納するとともに、キャッシュがフルであれば不要なデータを削除する。さらに、キャッシュメモリ制御部 1 0 4 は、読み出し対象のデータを入出力制御部 1 2 1 へ出力する。

20

【 0 0 7 2 】

バックエンド制御部 1 0 5 は、ホスト 2 から送信されたディスク 2 0 の構成情報に基づいてプール 2 0 0 及び論理ボリュームを生成する。この際、バックエンド制御部 1 0 5 は、プール 2 0 0 における全てのデータによる使用量であるディスク実使用量を初期化して 0 とする。

30

【 0 0 7 3 】

バックエンド制御部 1 0 5 は、データの書込命令をキャッシュメモリ制御部 1 0 4 から受ける。そして、バックエンド制御部 1 0 5 は、データの書込コマンドをディスクインタフェース 1 4 を介してディスク 2 0 に発行してデータを格納させる。

【 0 0 7 4 】

また、バックエンド制御部 1 0 5 は、データの読出命令をキャッシュメモリ制御部 1 0 4 から受ける。そして、バックエンド制御部 1 0 5 は、データの読出コマンドをディスクインタフェース 1 4 を介してディスク 2 0 に発行してデータを取得する。その後、バックエンド制御部 1 0 5 は、読み出したデータをキャッシュメモリ制御部 1 0 4 へ出力する。

【 0 0 7 5 】

40

また、バックエンド制御部 1 0 5 は、ガベージコレクションの実行の指示をガベージコレクション制御部 1 2 2 から受ける。そして、バックエンド制御部 1 0 5 は、メタデータテーブル 1 0 3 を参照して参照が無いデータである不要データを特定する。そして、バックエンド制御部 1 0 5 は、不要データを削除する。この時、バックエンド制御部 1 0 5 は、ガベージコレクション実行の指示で指定されたガベージコレクションの優先度に応じてガベージコレクション処理の延長コマンドの発行を行う。

【 0 0 7 6 】

次に、図 1 1 を参照して本実施例に係るコントローラモジュール 1 0 によるガベージコレクション処理の全体の流れを説明する。図 1 1 は、ガベージコレクション処理の全体のフローチャートである。

50

【 0 0 7 7 】

入出力制御部 1 2 1 は、ホスト 2 から送信された書込命令をチャンネルアダプタ 1 1 を介して受信する（ステップ S 1）。

【 0 0 7 8 】

次に、入出力制御部 1 2 1 は、保持するプール使用量を更新する（ステップ S 2）。

【 0 0 7 9 】

ガベージコレクション制御部 1 2 2 は、タイマを使用してガベージコレクションの動作タイミングが到来したか否かを判定する（ステップ S 3）。ガベージコレクションの動作タイミングが到来しない場合（ステップ S 3：否定）、重複圧縮制御部 1 0 2 の処理はステップ S 1 へ戻る。

10

【 0 0 8 0 】

これに対して、ガベージコレクションの動作タイミングが到来したい場合（ステップ S 3：肯定）、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの定期動作を開始する（ステップ S 4）。

【 0 0 8 1 】

次に、ガベージコレクション制御部 1 2 2 は、ストレージシステム 1 のシステム負荷を確認する。また、ガベージコレクション制御部 1 2 2 は、ディスク実使用量をバックエンド制御部 1 0 5 から取得して確認する（ステップ S 5）。

【 0 0 8 2 】

次に、ガベージコレクション制御部 1 2 2 は、入出力制御部 1 2 1 からプール使用量を取得する（ステップ S 6）。

20

【 0 0 8 3 】

次に、ガベージコレクション制御部 1 2 2 は、プール使用量及びディスク実使用量を用いてガベージコレクションの優先度を設定する（ステップ S 7）。

【 0 0 8 4 】

その後、ガベージコレクション制御部 1 2 2 は、設定した優先度でのガベージコレクションの実行を入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4 及びバックエンド制御部 1 0 5 に指示する。入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4 及びバックエンド制御部 1 0 5 は、I/O 処理を実行しつつ、設定された優先度でガベージコレクションを実行する（ステップ S 8）。

30

【 0 0 8 5 】

ここで、図 1 1 のステップ S 4 ~ S 8 においてガベージコレクション処理の実行中にも、コントローラモジュール 1 0 は、I/O 処理を並行して実施する。

【 0 0 8 6 】

次に、図 1 2 を参照して、プール使用量算出処理の流れについて説明する。図 1 2 は、プール使用量算出処理のフローチャートである。図 1 2 に示した処理は、図 1 1 におけるステップ S 1 及び S 2 で実行される処理の一例にあたる。

【 0 0 8 7 】

入出力制御部 1 2 1 は、ホスト 2 から送信された書込命令をチャンネルアダプタ 1 0 1 を介して受信する（ステップ S 1 0 1）。

40

【 0 0 8 8 】

次に、入出力制御部 1 2 1 は、重複排除圧縮処理を実行し指定されたデータの書込みを実行する（ステップ S 1 0 2）。

【 0 0 8 9 】

次に、入出力制御部 1 2 1 は、書き込み対象のデータが既存データと重複しないか否かを判定する（ステップ S 1 0 3）。書き込み対象のデータが既存データと重複する場合（ステップ S 1 0 3：否定）、入出力制御部 1 2 1 は、ステップ S 1 0 5 へ進む。

【 0 0 9 0 】

これに対して、書き込み対象のデータが既存データと重複しない場合（ステップ S 1 0 3：肯定）、入出力制御部 1 2 1 は、プール使用量に書き込み対象のデータによる使用量

50

を加算する（ステップ S 1 0 4）。

【 0 0 9 1 】

その後、入出力制御部 1 2 1 は、書き込みが上書きの場合の元データの参照カウンタが 0 か否かを判定する（ステップ S 1 0 5）。元データの参照カウンタが 0 でない場合（ステップ S 1 0 5：否定）、入出力制御部 1 2 1 は、ステップ 1 0 7 へ進む。

【 0 0 9 2 】

これに対して、元データの参照カウンタが 0 の場合（ステップ S 1 0 5：肯定）、入出力制御部 1 2 1 は、プール使用量から元データによる使用量を減算する（ステップ S 1 0 6）。

【 0 0 9 3 】

その後、入出力制御部 1 2 1 は、重複する既存データが無い場合は書き込むデータをキャッシュメモリ制御部 1 0 4 へ出力する。そして、キャッシュメモリ制御部 1 0 4 は、キャッシュにデータを書き込む（ステップ S 1 0 7）。

【 0 0 9 4 】

その後、キャッシュメモリ制御部 1 0 4 は、非同期で書き込むデータをキャッシュから読み出し、バックエンド制御部 1 0 5 へ出力する。バックエンド制御部 1 0 5 は、キャッシュメモリ制御部 1 0 4 から入力されたデータを書き込む書込コマンドをディスクインタフェース 1 4 を介してディスク 2 0 に発行してディスク 2 0 にデータを書き込む（ステップ S 1 0 8）。

【 0 0 9 5 】

次に、図 1 3 を参照して、実施例 1 に係るコントローラモジュール 1 0 による優先度設定処理の流れについて説明する。図 1 3 は、実施例 1 に係る優先度設定処理のフローチャートである。図 1 3 に示した処理は、図 1 2 におけるステップ S 4 ~ S 8 で実行される処理の一例にあたる。

【 0 0 9 6 】

ガベージコレクション制御部 1 2 2 は、ガベージコレクションの定期動作を開始する（ステップ S 2 0 1）。

【 0 0 9 7 】

次に、ガベージコレクション制御部 1 2 2 は、ストレージシステム 1 のシステム負荷を取得し、システム負荷が負荷閾値以下か否かを判定する（ステップ S 2 0 2）。システム負荷が負荷閾値以下の場合（ステップ S 2 0 2：肯定）、ガベージコレクション制御部 1 2 2 は、ディスク使用量からプール使用量を減算し、プール使用量とディスク使用量の差が閾値以上か否かを判定する（ステップ S 2 0 3）。

【 0 0 9 8 】

プール使用量とディスク使用量の差が閾値以上の場合（ステップ S 2 0 3：肯定）、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの優先度を高に設定する（ステップ S 2 0 4）。

【 0 0 9 9 】

一方、システム負荷が負荷閾値より大きい場合（ステップ S 2 0 2：否定）、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの優先度を通常に設定する（ステップ S 2 0 5）。同様に、プール使用量とディスク使用量の差が閾値未満の場合（ステップ S 2 0 3：否定）、ガベージコレクション制御部 1 2 2 は、ガベージコレクションの優先度を通常に設定する（ステップ S 2 0 5）。

【 0 1 0 0 】

その後、ガベージコレクション制御部 1 2 2 は、設定した優先度でのガベージコレクションの実行を入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4 及びバックエンド制御部 1 0 5 に指示する。入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4 及びバックエンド制御部 1 0 5 は、I/O 処理を実行しつつ、設定された優先度でガベージコレクションを実行する（ステップ S 2 0 6）。

【 0 1 0 1 】

10

20

30

40

50

以上に説明したように、本実施例に係るコントローラモジュールは、使用中のデータによる使用量であるプール使用量と、全データによる使用量であるディスク実使用量との差を算出する。そして、コントローラモジュールは、ディスク実使用量とプール使用量との差が閾値以上であれば、ガベージコレクションの実行の優先度を上昇させる。すなわち、コントローラモジュールは、ガベージコレクションの実行の設定を変更して、コントローラモジュールが実行する処理全体のうちのガベージコレクションの割合を上昇させる。

【0102】

これにより、ガベージコレクションの実行によるディスクの空き容量の確保の効果が高い場合に、ガベージコレクションの割合を上げることができ、迅速に空き容量を確保することができる。言い換えれば、ガベージコレクションを事項しても十分な効果を得られない場合には、ガベージコレクションの処理の割合を維持することで、より多くのCPU性能をホストコンピュータからのIO処理に割くことが可能となる。結果的に、限りあるディスクのリソースを効率的に使用することが可能となり、ストレージシステムの装置性能を有効に発揮させることができる。

10

【0103】

また、システム負荷を用いてガベージコレクションの設定を変更することで、例えば、システム負荷が低く余裕があるタイミングで、ガベージコレクションを優先的に動作させてディスクの空き容量を確保することができ、IO処理への影響を抑えることができる。

【0104】

以上のように、本実施例に係るコントローラモジュールは、ストレージ装置における空き容量の確保と処理負荷とのバランスを適切に保つことができ、ストレージ装置の装置性能を向上させることができる。

20

【実施例2】

【0105】

図14は、実施例2に係るコントローラモジュールのブロック図である。本実施例に係るコントローラモジュール10は、ディスク実使用量が閾値以上の場合にガベージコレクションの優先度を高に設定した上で管理者に通知を行うことが実施例1と異なる。本実施例に係るコントローラモジュール10は、実施例1の各部に加えて報知部106を有する。以下の説明では、実施例1で説明した各部の動作については説明を省略する。

【0106】

ガベージコレクション制御部122は、ガベージコレクションの定期動作を開始すると、ディスク実使用量をバックエンド制御部105から取得する。そして、ガベージコレクション制御部122は、ディスク実使用量が予め決められた使用量閾値以上か否かを判定する。

30

【0107】

ディスク実使用量が使用量閾値以上の場合、ディスク20の空き容量が僅少であり危険であると判定できることから、ガベージコレクション制御部122は、ガベージコレクションの優先度を高に設定する。次に、ガベージコレクション制御部122は、プール使用量を入出力制御部121から取得する。そして、ガベージコレクション制御部122は、ディスク実使用量からプール使用量を減算して、減算結果であるプール使用量とディスク実使用量との差が閾値以上か否かを判定する。そして、ガベージコレクション制御部122は、判定結果を報知部106に通知する。

40

【0108】

一方、ディスク実使用量が使用量閾値未満の場合はディスク20の空き容量に余裕があるので、ガベージコレクション制御部122は、実施例1と同様にシステム負荷及びプール使用量とディスク実使用量との差を用いてガベージコレクションの優先度を決定する。

【0109】

報知部106は、プール使用量とディスク実使用量との差が閾値以上か否かの判定結果の通知をガベージコレクション制御部122から受ける。

【0110】

50

プール使用量とディスク実使用量との差が閾値以上の場合、ガベージコレクションを実行することでディスクの空き容量がある程度確保できることが想定できるので、報知部 106 は、ストレージシステム 1 の性能低下の通知を管理者に通知する。

【0111】

これに対して、プール使用量とディスク実使用量との差が閾値未満の場合、ガベージコレクションを実行してもディスクの空き容量の確保が困難であると想定できるので、報知部 106 は、ディスク 20 の追加推奨を管理者に通知する。この報知部 106 の機能も、CPU 12 により実現される。

【0112】

次に、図 15 を参照して、本実施例に係るコントローラモジュール 10 による優先度設定処理の流れについて説明する。図 15 は、実施例 2 に係る優先度設定処理のフローチャートである。

【0113】

ガベージコレクション制御部 122 は、ガベージコレクションの定期動作を開始する（ステップ S301）。

【0114】

次に、ガベージコレクション制御部 122 は、ディスク実使用量をバックエンド制御部 105 から取得する。そして、ガベージコレクション制御部 122 は、ディスク実使用量が使用量閾値以上か否かを判定する（ステップ S302）。

【0115】

ディスク実使用量が使用量閾値以上の場合（ステップ S302：否定）、ガベージコレクション制御部 122 は、ストレージシステム 1 のシステム負荷を取得し、システム負荷が負荷閾値以下か否かを判定する（ステップ S303）。システム負荷が負荷閾値以下の場合（ステップ S303：肯定）、ガベージコレクション制御部 122 は、ディスク使用量からプール使用量を減算し、プール使用量とディスク使用量の差が閾値以上か否かを判定する（ステップ S304）。

【0116】

プール使用量とディスク使用量の差が閾値以上の場合（ステップ S304：肯定）、ガベージコレクション制御部 122 は、ガベージコレクションの優先度を高に設定する（ステップ S305）。

【0117】

一方、システム負荷が負荷閾値より大きい場合（ステップ S303：否定）、ガベージコレクション制御部 122 は、ガベージコレクションの優先度を通常に設定する（ステップ S306）。同様に、プール使用量とディスク使用量の差が閾値未満の場合（ステップ S304：否定）、ガベージコレクション制御部 122 は、ガベージコレクションの優先度を通常に設定する（ステップ S306）。

【0118】

一方、ディスク実使用量が使用量閾値以上の場合（ステップ S302：肯定）、ガベージコレクション制御部 122 は、ガベージコレクションの優先度を高に設定する（ステップ S307）。

【0119】

次に、ガベージコレクション制御部 122 は、プール使用量を入出力制御部 121 から取得する。そして、ガベージコレクション制御部 122 は、ディスク実使用量からプール使用量を減算して、プール使用量とディスク実使用量との差が閾値以上か否かを判定する（ステップ S308）。そして、ガベージコレクション制御部 122 は、判定結果を報知部 106 へ通知する。

【0120】

プール使用量とディスク実使用量との差が閾値以上の場合（ステップ S308：肯定）、報知部 106 は、ストレージシステム 1 の性能低下を管理者に通知する（ステップ S309）。

10

20

30

40

50

【 0 1 2 1 】

これに対して、プール使用量とディスク実使用量との差が閾値未満の場合（ステップ S 3 0 8：否定）、報知部 1 0 6 は、ディスク 2 0 の追加推奨を管理者に通知する（ステップ S 3 1 0）。

【 0 1 2 2 】

その後、ガベージコレクション制御部 1 2 2 は、設定した優先度でのガベージコレクションの実行を入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4 及びバックエンド制御部 1 0 5 に指示する。入出力制御部 1 2 1、キャッシュメモリ制御部 1 0 4 及びバックエンド制御部 1 0 5 は、I/O 処理を実行しつつ、設定された優先度でガベージコレクションを実行する（ステップ S 3 1 1）。

10

【 0 1 2 3 】

以上に説明したように、本実施例に係るコントローラモジュールは、ディスク実使用量が使用量閾値以上の場合、ガベージコレクションの優先度を高に設定する。さらに、コントローラモジュールは、プール使用量とディスク実使用量の差から判定される現在のストレージシステムの状態を管理者に通知する。

【 0 1 2 4 】

これにより、ディスクの空き容量が僅少な場合に、ガベージコレクション処理を優先させることができ、迅速にディスクの空き容量を確保することができる。また、ディスクの空き容量が僅少で危険な状態と考えられる場合に、管理者にストレージシステムの状態を通知することで障害発生前の対応を促すことができ、ストレージシステムの運転の継続性を維持して信頼性を確保することができる。

20

【符号の説明】

【 0 1 2 5 】

- 1 ストレージシステム
- 2 ホスト
- 1 0 コントローラモジュール
- 1 1 チャンネルアダプタ
- 1 2 C P U
- 1 3 D R A M
- 1 4 ディスクインタフェース
- 2 0 ディスク
- 1 0 2 重複圧縮制御部
- 1 0 3 メタデータテーブル
- 1 0 4 キャッシュメモリ制御部
- 1 0 5 バックエンド制御部
- 1 0 6 報知部
- 2 0 0 プール

30

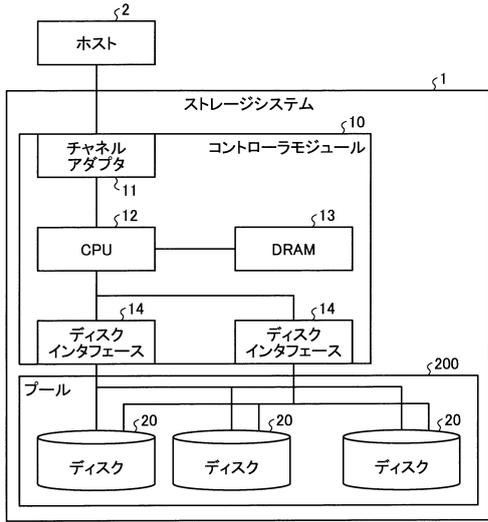
40

50

【 図面 】

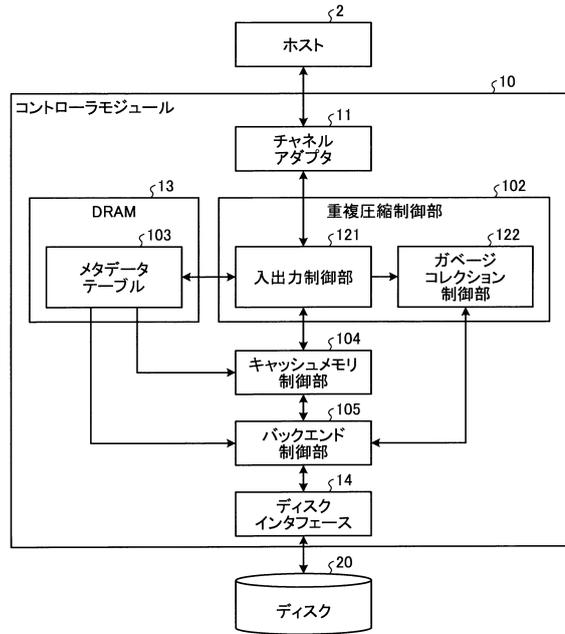
【 図 1 】

ストレージシステムのハードウェア構成図



【 図 2 】

実施例1に係るコントローラモジュールのブロック図



10

20

【 図 3 】

論理ボリューム側管理テーブルの一例を表す図

論理ボリュームLBA	データ番号
0x0000	0
0x0008	1
0x0010	2
0x0018	1
0x0020	1
0x0028	3
0x0030	4
0x0038	2

【 図 4 】

物理ボリューム側管理テーブルの一例を表す図

データ番号	参照カウンタ	物理ディスクアドレス	データサイズ
0	1	0x0000	3
1	3	0x0003	5
2	2	0x0008	8
3	1	0x0010	1
4	1	0x0011	8

↔

AAA
BBBBB
CCCCCCC
D
EEEEEEEE

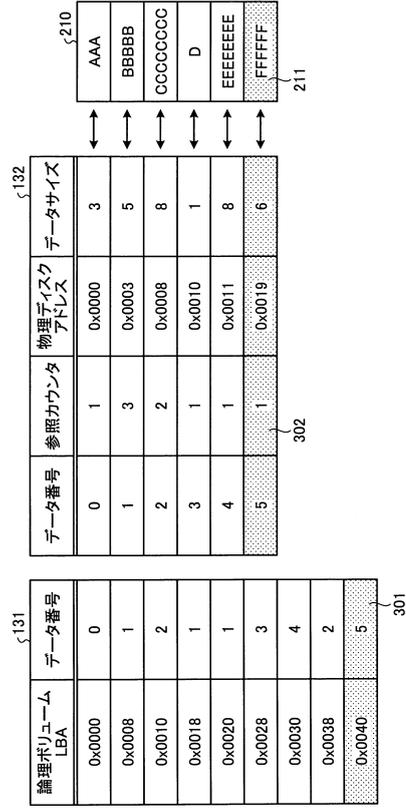
30

40

50

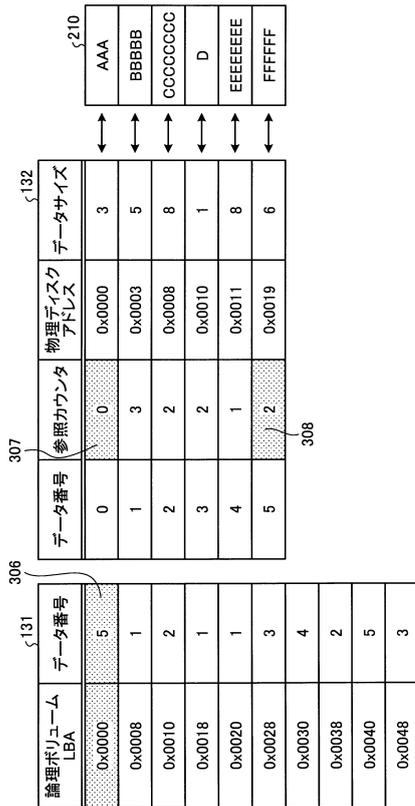
【図 5】

新規データ書き込み時の管理テーブルの遷移を表す図



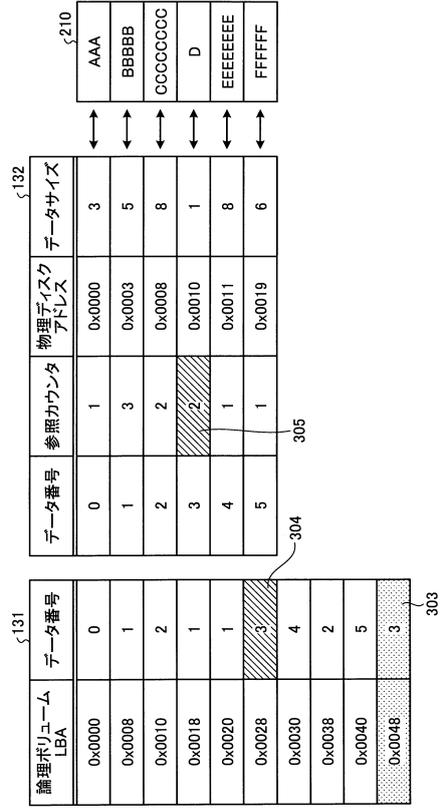
【図 7】

データ上書き時の管理テーブルの遷移を表す図



【図 6】

重複データ書き込み時の管理テーブルの遷移を表す図



【図 8】

ガベージコレクション処理が割り当てられない場合の処理割り当てを表す図



【図 9】

ガベージコレクション処理の優先度が通常の場合の処理割り当てを表す図

コア番号	#0	#1	#2	#3	#4	#5	#6	#7	#8	#9
実行中処理	IO処理	GC処理								
実行待ち処理	IO処理									
実行待ち処理	IO処理	GC処理								
実行待ち処理	IO処理	GC処理								
実行待ち処理	IO処理	GC処理								

【図 10】

ガベージコレクション処理の優先度が高い場合の処理割り当てを表す図

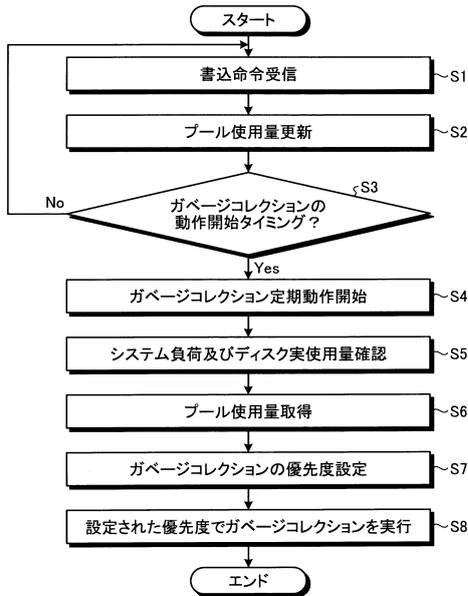
コア番号	#0	#1	#2	#3	#4	#5	#6	#7	#8	#9
実行中処理	IO処理	GC処理	IO処理	IO処理	GC処理	GC処理	IO処理	IO処理	IO処理	GC処理
実行待ち処理	GC処理	IO処理	IO処理	IO処理	GC処理	IO処理	IO処理	GC処理	GC処理	IO処理
実行待ち処理	IO処理	IO処理	GC処理	GC処理	GC処理	GC処理	IO処理	GC処理	GC処理	GC処理
実行待ち処理	GC処理	IO処理	GC処理	GC処理	GC処理	IO処理	IO処理	GC処理	GC処理	GC処理
実行待ち処理	GC処理	IO処理	GC処理	GC処理	GC処理	IO処理	IO処理	GC処理	GC処理	GC処理

10

20

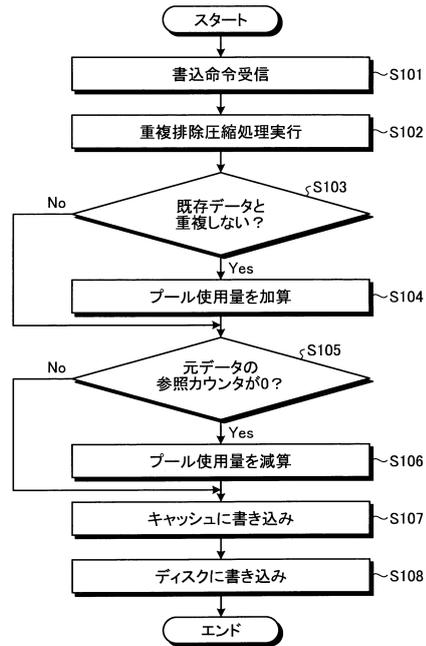
【図 11】

ガベージコレクション処理の全体のフローチャート



【図 12】

プール使用量算出処理のフローチャート



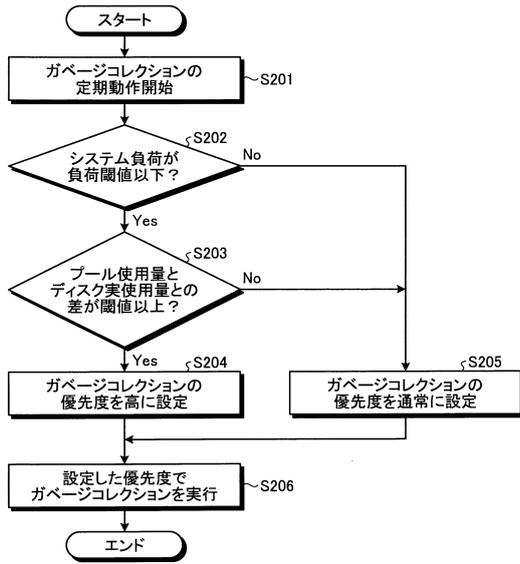
30

40

50

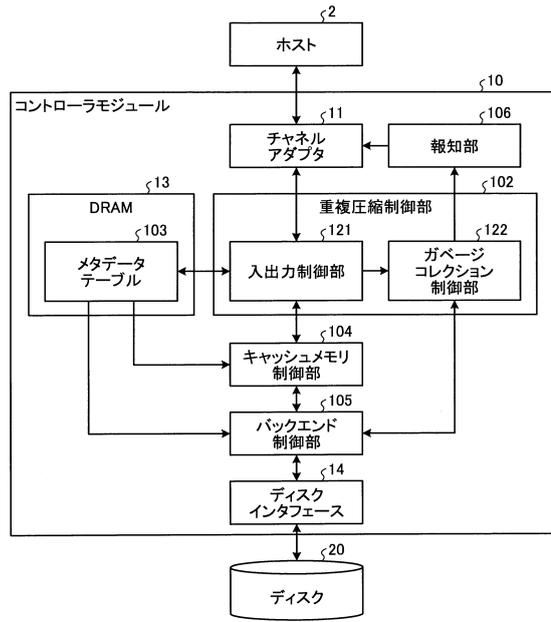
【 図 1 3 】

実施例1に係る優先度設定処理のフローチャート



【 図 1 4 】

実施例2に係るコントローラモジュールのブロック図

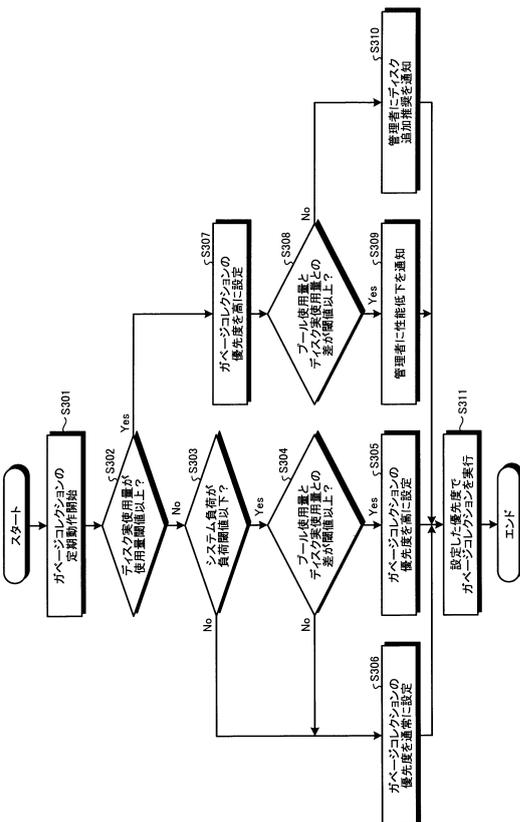


10

20

【 図 1 5 】

実施例2に係る優先度設定処理のフローチャート



30

40

50

フロントページの続き

(51)国際特許分類

F I
G 0 6 F 16/13 1 2 0

(56)参考文献

米国特許出願公開第 2 0 1 5 / 0 2 6 8 8 6 1 (U S , A 1)
米国特許出願公開第 2 0 1 5 / 0 0 1 9 7 9 7 (U S , A 1)
国際公開第 2 0 1 5 / 0 5 2 8 3 3 (W O , A 1)
特開 2 0 1 9 - 1 2 8 9 0 6 (J P , A)
特開 2 0 1 7 - 1 1 7 0 5 5 (J P , A)
特開 2 0 1 1 - 1 9 2 2 6 0 (J P , A)
特開 2 0 1 6 - 1 7 0 5 8 3 (J P , A)
米国特許出願公開第 2 0 1 6 / 0 1 2 4 8 4 8 (U S , A 1)

(58)調査した分野 (Int.Cl. , D B 名)

G 0 6 F 3 / 0 6 - 3 / 0 8
G 0 6 F 1 3 / 1 0 - 1 3 / 1 4
G 0 6 F 1 6 / 1 0 - 1 6 / 1 8 8