

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6229347号
(P6229347)

(45) 発行日 平成29年11月15日(2017.11.15)

(24) 登録日 平成29年10月27日(2017.10.27)

(51) Int.Cl.

G06F 3/06 (2006.01)

F 1

G 06 F 3/06 3 O 1 F

請求項の数 7 (全 16 頁)

(21) 出願番号 特願2013-154585 (P2013-154585)
 (22) 出願日 平成25年7月25日 (2013.7.25)
 (65) 公開番号 特開2015-26183 (P2015-26183A)
 (43) 公開日 平成27年2月5日 (2015.2.5)
 審査請求日 平成28年4月5日 (2016.4.5)

(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 100092152
 弁理士 服部 毅巖
 (72) 発明者 鈴木 友憲
 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
 (72) 発明者 遠藤 篤
 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 桜井 茂行

最終頁に続く

(54) 【発明の名称】ストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法

(57) 【特許請求の範囲】

【請求項 1】

論理ボリュームにおけるストレージ装置の構成、および前記ストレージ装置の性能にもとづいて前記論理ボリュームに対して単位時間当たりに発行可能なI/O発行可能数を可変設定する設定部と、

前記論理ボリュームに対して受け付けたI/O要求を要求蓄積部に蓄積し、前記要求蓄積部が蓄積する前記I/O要求のうちから、単位時間当たりのI/O発行数を設定された前記I/O発行可能数の範囲に制限して前記論理ボリュームに前記I/O要求の発行をおこなう発行制御部と、

を備えることを特徴とするストレージ制御装置。

10

【請求項 2】

前記設定部は、前記ストレージ装置の構成、および前記ストレージ装置の性能ごとに要求される性能情報にもとづいて、前記論理ボリュームごとに前記性能情報から導出されるI/O発行可能数を前記I/O発行可能数として設定することを特徴とする請求項1記載のストレージ制御装置。

【請求項 3】

前記設定部は、前記発行制御部が所定時間の間に発行した前記I/O要求の転送単位に応じて前記I/O発行可能数を変更することを特徴とする請求項1または請求項2記載のストレージ制御装置。

【請求項 4】

20

前記発行制御部は、前記所定時間の間に発行した前記 I / O 要求の転送単位を統計情報として保持し、前記統計情報にもとづいて前記転送単位を選択することを特徴とする請求項 3 記載のストレージ制御装置。

【請求項 5】

前記設定部は、前記論理ボリュームの生成時に前記性能情報にもとづいて転送単位ごとに対応する I / O 発行可能数を算出することを特徴とする請求項 2 記載のストレージ制御装置。

【請求項 6】

コンピュータに、

論理ボリュームにおけるストレージ装置の構成、および前記ストレージ装置の性能にもとづいて前記論理ボリュームに対して単位時間当たりに発行可能な I / O 発行可能数を可変設定し、10

前記論理ボリュームに対して受け付けた I / O 要求を要求蓄積部に蓄積し、前記要求蓄積部が蓄積する前記 I / O 要求のうちから、単位時間当たりの I / O 発行数を設定された前記 I / O 発行可能数の範囲に制限して前記論理ボリュームに前記 I / O 要求の発行をおこなう、

処理を実行することを特徴とするストレージ制御プログラム。

【請求項 7】

コンピュータが、

論理ボリュームにおけるストレージ装置の構成、および前記ストレージ装置の性能にもとづいて前記論理ボリュームに対して単位時間当たりに発行可能な I / O 発行可能数を可変設定し、20

前記論理ボリュームに対して受け付けた I / O 要求を要求蓄積部に蓄積し、前記要求蓄積部が蓄積する前記 I / O 要求のうちから、単位時間当たりの I / O 発行数を設定された前記 I / O 発行可能数の範囲に制限して前記論理ボリュームに前記 I / O 要求の発行をおこなう、

処理を実行することを特徴とするストレージ制御方法。

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は、ストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法に関する。

【背景技術】

【0002】

複数の HDD (Hard Disk Drive) を 1 つの論理ボリュームとして扱うストレージ制御装置として RAID (Redundant Array of Inexpensive Disks) 装置が知られている。 RAID 装置は、複数の HDD を搭載したドライブエンクロージャ (以下、DE (Drive Enclosure)) を備え、DE 内の HDD から RAID を構成する。また、RAID 装置は、DE 内の HDD の交換や追加、あるいは DE の交換や追加により HDD 構成を更新可能にする。40

【0003】

RAID 装置は、HDD 構成に変更が生じても安定した性能値 (たとえば、スループット、IOPS (Input/Output Per Second)) を提供することが求められる。そのため、RAID 装置は、所定の性能を有することを検証された HDD を搭載している。HDD に求められる所定の性能は、その下限だけでなく上限も規定した基準の範囲内に適合することが求められる。

【0004】

RAID 装置ベンダと HDD ベンダは、HDD に求められる性能の上限が HDD 内部の処理に委ねられることから、ファームウェアの変更と検証を繰り返してファームウェアの50

最適化を図っている。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開平8-69359号公報

【特許文献2】特開平8-63298号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、市場や生産拠点のグローバル化に伴うサプライチェーンの最適化を図る上で、HDDのマルチベンダ化が進展している。このようなマルチベンダ化の流れは、自然災害や政情不安などの不安定要因に対抗するため一層進展する傾向にある。10

【0007】

また、ユーザのニーズにより細やかに対応するため、HDDの多機種化（サイズ、回転数など）に伴い様々な性能特性を有するHDDが存在する。

RAID装置ベンダとHDDベンダによるファームウェアの最適化を図るための作業負担は、ますます増大する傾向にある。

【0008】

1つの側面では、本発明は、ストレージ装置の性能最適化に伴う作業負担を軽減可能なストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法を提供することを目的とする。20

【課題を解決するための手段】

【0009】

上記目的を達成するために、以下に示すような、ストレージ制御装置が提供される。ストレージ制御装置は、設定部と、発行制御部を備える。設定部は、論理ボリュームにおけるストレージ装置の構成、およびストレージ装置の性能にもとづいて論理ボリュームに対して単位時間当たりに発行可能なI/O発行可能数を可変設定する。発行制御部は、論理ボリュームに対して受け付けたI/O要求を要求蓄積部に蓄積し、要求蓄積部が蓄積するI/O要求のうちから、単位時間当たりのI/O発行数を設定されたI/O発行可能数の範囲に制限して論理ボリュームにI/O要求の発行をおこなう。30

【0010】

また、上記目的を達成するために、以下に示すような、ストレージ制御プログラム、およびストレージ制御方法が提供される。ストレージ制御プログラムは、以下に示すような処理をコンピュータに実行させ、ストレージ制御方法は、以下に示すような処理をコンピュータが実行する。コンピュータは、論理ボリュームにおけるストレージ装置の構成、およびストレージ装置の性能にもとづいて論理ボリュームに対して単位時間当たりに発行可能なI/O発行可能数を可変設定し、論理ボリュームに対して受け付けたI/O要求を要求蓄積部に蓄積し要求蓄積部が蓄積するI/O要求のうちから、単位時間当たりのI/O発行数を設定されたI/O発行可能数の範囲に制限して論理ボリュームにI/O要求の発行をおこなう処理を実行する。40

【発明の効果】

【0011】

1態様によれば、ストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法において、ストレージ装置の性能最適化に伴う作業負担を軽減する。

【図面の簡単な説明】

【0012】

【図1】第1の実施形態のストレージ制御装置の構成例を示す図である。

【図2】第2の実施形態のストレージシステムの構成例を示す図である。

【図3】異なるHDD種別におけるRAID性能の測定例を示す図である。

【図4】第2の実施形態のコントローラモジュールの構成例を示す図である。50

【図5】第2の実施形態のRAID装置のハードウェア構成例を示す図である。

【図6】第2の実施形態のRAIDグループ追加処理のフローチャートを示す図である。

【図7】第2の実施形態の性能情報テーブルの一例を示す図である。

【図8】第2の実施形態のI/O発行数管理テーブルの一例を示す図である。

【図9】第2の実施形態のRAIDグループI/O受付処理のフローチャートを示す図である。

【図10】第2の実施形態の統計情報管理テーブルの一例を示す図である。

【図11】第2の実施形態のRAIDグループI/O発行数管理処理のフローチャートを示す図である。

【発明を実施するための形態】

10

【0013】

以下、図面を参照して実施の形態を詳細に説明する。

[第1の実施形態]

まず、第1の実施形態のストレージ制御装置について図1を用いて説明する。図1は、第1の実施形態のストレージ制御装置の構成例を示す図である。

【0014】

ストレージ制御装置1は、ホストコンピュータ9から論理ボリュームへのI/O要求7を受け付ける。ストレージ制御装置1は、複数のストレージ装置5(5a, 5b, ..., 5n)の組合せから論理ボリュームを生成する。論理ボリュームは、たとえば、RAIDがある。

20

【0015】

ストレージ装置5は、たとえば、HDDであるが、フラッシュメモリドライブなどのSSD(Solid State Drive)などであってもよい。また、ストレージ装置5は、ストレージ制御装置1が内蔵するものであってもよいし、ストレージ制御装置1に外付けされるものであってもよい。

【0016】

ストレージ制御装置1は、設定部2と、発行制御部3と、要求蓄積部4を含む。設定部2は、論理ボリュームに対して単位時間当たりに発行可能なI/O発行可能数6を設定する。設定部2は、論理ボリュームにおけるストレージ装置5の構成、およびストレージ装置5の性能にもとづいて、I/O発行可能数を設定する。

30

【0017】

これにより、ストレージ制御装置1は、論理ボリュームを構成するストレージ装置5の性能を反映したI/O発行可能数を設定することができる。

発行制御部3は、論理ボリュームに対して受け付けたI/O要求7を要求蓄積部4に蓄積する。発行制御部3は、要求蓄積部4が蓄積するI/O要求7から論理ボリュームにI/O要求の発行をおこなう。このとき、発行制御部3は、単位時間当たりのI/O発行数をI/O発行可能数6の範囲に制限する。

【0018】

これにより、ストレージ制御装置1は、ストレージ装置5が過大な応答性能を有する場合があっても、論理ボリュームが過大な応答性能を有することを抑制する。このようなストレージ制御装置1は、論理ボリュームを構成するストレージ装置5の組合せに変更が生じても安定した性能値を発揮することができる。そのため、ストレージ制御装置1は、ストレージ装置5の性能最適化に伴う作業負担を軽減することができる。

40

【0019】

[第2の実施形態]

次に、第2の実施形態のストレージ装置の接続例について図2を用いて説明する。図2は、第2の実施形態のストレージシステムの構成例を示す図である。

【0020】

ストレージシステム10は、1以上のホストコンピュータ11と、ホストコンピュータ11と通信可能に接続する1以上のRAID装置12とを含んで構成される。RAID装

50

置12は、複数のHDD19を1つの論理ボリュームとして扱うストレージ装置の1つである。RAID装置12は、RAIDへのI/O要求をホストコンピュータ11から受け付ける。RAID装置12は、コントローラモジュール(以下、CM(Controller Module)13と、ドライブエンクロージャ(DE)14とを含む。

【0021】

DE14は、複数のHDD19を収容する筐体である。DE14は、複数のHDD19への電源供給を担うとともに、HDD19とCM13とを接続するインターフェースを有する。RAID装置12は、DE14が収容する複数のHDD19の組合せによりRAIDを構成する。なお、DE14は、RAID装置12が内蔵するものであってもよいし、RAID装置12に外付けされるものであってもよい。また、図示するRAID装置12は、1つのDE14を備えるが2以上のDE14を備えてよい。
10

【0022】

CM13は、ストレージ制御装置の一形態であり、ホストコンピュータ11からのI/O要求(たとえば、Write要求、Read要求など)を受け付けて、DE14が備える複数のHDD19へのアクセスを制御する。CM13は、ホストI/O制御部15と、RAID制御部16と、ディスクI/O制御部17を含む。なお、図示するRAID装置12は、1つのCM13を備えるが2以上のCM13を備える冗長構成としてもよい。

【0023】

ホストI/O制御部15は、ホストコンピュータ11とCM13とのインターフェース制御をおこなう。CM13は、ホストI/O制御部15を介してホストコンピュータ11との間で所要のデータの送受信をおこなう。RAID制御部16は、RAIDグループごとのI/O発行数を管理する。ディスクI/O制御部17は、DE14とCM13とのインターフェース制御をおこなう。CM13は、ディスクI/O制御部17を介してDE14との間で所要のデータの送受信をおこなう。
20

【0024】

ここで、異なるHDD種別におけるRAID性能の測定例について図3を用いて説明する。図3は、異なるHDD種別におけるRAID性能の測定例を示す図である。

HDD_AとHDD_Bとは、RAID装置12において相互に交換可能なHDDである。所定条件下でアクセス性能を測定したとき、HDD_AはP1(MB/s)、HDD_BはP2(MB/s)であり、HDD_AとHDD_Bの間にd1(MB/s)の性能差がある。
30

【0025】

HDD_Aは、要求性能下限PLと要求性能上限PHを満足するが、HDD_Bは、要求性能下限PLを満足するものの要求性能上限PHを満足していない。たとえば、HDD_Bを含むHDDの組合せで、ユーザが要求する性能を満足するRAIDを構成していたときに、HDD_BをHDD_Aに交換するとユーザが要求する性能を満たさなくなってしまう場合がある。

【0026】

このように、HDDごとのI/O発行数に上限値を設けたとしてもI/O要求の発行後の処理は、HDD側に任せられるため、RAID性能は、HDD内部の処理方法に委ねられる。そのため、従来であれば、HDD_AとHDD_Bが互換性を有して運用できるよう、設計段階でHDDのファームウェア性能の最適化を図っている。
40

【0027】

たとえば、RAID装置ベンダは、新規のHDDの性能測定を実施し、測定結果にもとづいて要求性能値を算出してHDDベンダに通知する。HDDベンダは、通知された要求性能値にもとづいてHDDのファームウェアを更新してRAID装置ベンダに提供する。RAID装置ベンダとHDDベンダは、要求性能値が満たされるまで上記作業を反復する。

【0028】

このようなHDDの性能最適化に伴う作業負担を軽減するための処理を実行するRAI
50

D制御部について、図4を用いて説明する。図4は、第2の実施形態のコントローラモジュールの構成例を示す図である。

【0029】

C M (コントローラモジュール) 13は、前述した通り、ホストI/O制御部15と、RAID制御部16と、ディスクI/O制御部17を含む。ホストI/O制御部15は、I/O受付部20と、I/O完了受付部21を含む。I/O受付部20は、ホストコンピュータ11からI/O要求を受け付けてRAID制御部16にI/O要求を通知する。I/O完了受付部21は、RAID制御部16からI/O完了応答を受け付けてホストコンピュータ11にI/O完了応答を通知する。

【0030】

RAID制御部16は、I/O受付部22と、I/O発行可能数算出部25と、I/O完了受付部28を含む。I/O受付部22は、ホストI/O制御部15からI/O要求を受け付けて、RAIDグループごとに時間当たりのI/O発行数を制限してディスクI/O制御部17にI/O要求を通知する。I/O受付部22は、I/O発行数管理部23と、管理時間監視部24を含む。

10

【0031】

I/O発行数管理部23は、I/O受付部22が受け付けたI/O要求をI/O発行待ちキューに蓄積し、時間当たりのI/O発行数をI/O発行可能数27の範囲に制限しながらディスクI/O制御部17にI/O要求の発行(I/O発行)をおこなう。管理時間監視部24は、I/O発行数を制限する際の単位時間を管理時間として監視する。

20

【0032】

I/O発行可能数算出部25は、あらかじめ設定される性能情報テーブル26を保持し、性能情報テーブル26にもとづいてRAIDグループごとにI/O発行可能数27を算出し、I/O発行可能数27を保持する。性能情報テーブル26は、RAIDの構成条件別の要求性能を示す。性能情報テーブル26の詳細は、後で図7を用いて説明する。I/O完了受付部28は、ディスクI/O制御部17からI/O完了応答を受け付けてホストI/O制御部15にI/O完了応答を通知する。

【0033】

ディスクI/O制御部17は、I/O受付部29と、I/O完了受付部30を含む。I/O受付部29は、RAID制御部16からI/O要求を受け付けてRAIDを構成するHDD19にI/O要求を通知する。HDD19は、I/O要求を受け付け、I/O要求に対応するコマンド処理をおこなう。HDD19は、コマンド処理の実行後にI/O完了受付部30にI/O完了応答を通知する。I/O完了受付部30は、HDD19からI/O完了応答を受け付けてRAID制御部16にI/O完了応答を通知する。

30

【0034】

次に、第2の実施形態のRAID装置のハードウェア構成について図5を用いて説明する。図5は、第2の実施形態のRAID装置のハードウェア構成例を示す図である。

RAID装置12は、チャネルアダプタ31と、コントローラモジュール13と、DE14を備える。RAID装置12は、チャネルアダプタ31を介してホストコンピュータ11と接続する。コントローラモジュール13は、チャネルアダプタ31を複数(たとえば、2つ)備え、複数のチャネルアダプタ31による冗長構成のもと複数系統でホストコンピュータ11と接続する。コントローラモジュール13は、ディスクアダプタ34を介して、DE14が収容するHDD19と接続する。

40

【0035】

コントローラモジュール13は、プロセッサ32、メモリ33、ディスクアダプタ34を備え、図示しないバスを介して接続されている。プロセッサ32は、コントローラモジュール13全体を制御し、RAID制御を含むストレージ制御をおこなう。プロセッサ32は、マルチプロセッサであってもよい。プロセッサ32は、たとえばCPU(Central Processing Unit)、MPU(Micro Processing Unit)、DSP(Digital Signal Processor)、ASIC(Application Specific Integrated Circuit)、またはPLD(Programmable Logic Device)である。

50

ammable Logic Device) である。またプロセッサ 32 は、CPU、MPU、DSP、ASIC、PLD のうちの 2 以上の要素の組合せであってもよい。

【0036】

メモリ 33 は、たとえば、RAM (Random Access Memory) や不揮発性メモリを含む。メモリ 33 は、HDD 19 からデータを読み出したときにデータを保持するほか、I/O 要求を一時的に蓄積する I/O 発行待ちキューや HDD 19 にデータを書き込むときのバッファとなる。また、メモリ 33 は、ユーザデータや制御情報を格納する。ディスクアダプタ 34 は、HDD 19 とのインターフェース制御 (アクセス制御) をおこなう。たとえば、RAM は、コントローラモジュール 13 の主記憶装置として使用される。RAM には、プログラムやファームウェア、アプリケーションプログラムの少なくとも一部が一時的に格納される。また、RAM には、プロセッサ 32 による処理に必要な各種データが格納される。また、RAM は、各種データの格納に用いるメモリと別体にキャッシュメモリを含むものであってもよい。10

【0037】

不揮発性メモリは、RAID 装置 12 の電源遮断時においても記憶内容を保持する。不揮発性メモリは、たとえば、EEPROM (Electrically Erasable and Programmable Read Only Memory) やフラッシュメモリなどの半導体記憶装置や、HDD などである。不揮発性メモリには、オペレーティングシステムのプログラムやファームウェア、アプリケーションプログラム、および各種データが格納される。20

【0038】

以上のようなハードウェア構成によって、第 2 の実施形態の RAID 装置 12、コントローラモジュール 13 の処理機能を実現することができる。なお、第 1 の実施形態に示したストレージ制御装置 1 も、図示した RAID 装置 12、コントローラモジュール 13 と同様のハードウェアにより実現することができる。

【0039】

次に、第 2 の実施形態の RAID 制御部 16 が実行する RAID グループ追加処理について図 6 を用いて説明する。図 6 は、第 2 の実施形態の RAID グループ追加処理のフローチャートを示す図である。RAID グループ追加処理は、新規の RAID グループが追加されたときに RAID 制御部 16 により、より詳しくは I/O 発行可能数算出部 25 により実行される。RAID グループ追加処理は、新規に追加された RAID グループについて I/O 発行可能数を決定し、決定した I/O 発行可能数のもとで I/O 制御を開始する処理である。30

【0040】

[ステップ S11] RAID 制御部 16 は、新規に追加された RAID グループの RAID レベルを取得する。RAID レベルは、たとえば、RAID 1、RAID 5、RAID 6、RAID 1+0、RAID 5+0 などがあるが、その他の RAID レベルであってもよい。

【0041】

[ステップ S12] RAID 制御部 16 は、ディスク種別を取得する。ディスク種別は、HDD の回転数である。HDD の回転数は、たとえば、15 k rpm (revolution per minute)、10 k rpm、7.2 k rpm などがある。なお、ディスク種別は、ディスクの性能を区別する指標であれば、HDD の回転数に限らず、HDD のサイズや記録容量、バッファ容量、最大読み出し速度や最大書き込み速度、MTBF (平均故障間隔) などであってもよい。また、ディスクの性能を区別する指標は、1 つに限らず複数あってもよい。また、ディスクは、HDD に限らず、SSD などであってもよい。40

【0042】

[ステップ S13] RAID 制御部 16 は、ディスク実構成数を取得する。ディスク実構成数は、新規に追加された RAID グループに割り当てられたディスクの数である。

[ステップ S14] RAID 制御部 16 は、性能情報テーブルを参照する。性能情報テ50

ーブルは、あらかじめ R A I D レベル、ディスク種別に用意される。たとえば、性能情報テーブルは、メモリ 33 に保持されるが、ホストコンピュータ 11 など外部装置から取得されるものであってもよい。

【 0 0 4 3 】

ここで、性能情報テーブルについて図 7 を用いて説明する。図 7 は、第 2 の実施形態の性能情報テーブルの一例を示す図である。性能情報テーブル 100 は、あらかじめ用意されている性能情報テーブルの一例である。性能情報テーブル 100 は、R A I D レベル、ディスク種別（回転数）、基本構成ディスク数、転送単位（k b y t e）、および要求性能（M B / s）を項目として有する。R A I D レベル、ディスク種別（回転数）は、あらかじめ用意されている性能情報テーブルのうちから対応する性能情報テーブルを参照するためのインデックスとなる項目である。性能情報テーブル 100 は、新規に追加された R A I D グループが、R A I D レベル「R A I D 5」、ディスク種別「15 k r p m」であるときに参照される。基本構成ディスク数は、R A I D レベルに応じたディスク数であり、R A I D レベル「R A I D 5」の基本構成ディスク数は、「5」である。10

【 0 0 4 4 】

転送単位は、R A I D グループに設定される転送単位である。たとえば、転送単位は、基本転送単位 N を基準にして 7 段階に分けられている。まず、第 1 段階が「N × 8 より大きい範囲」、第 2 段階が「N × 8 から N × 4 の範囲」、第 3 段階が「N × 4 から N + 1 の範囲」、第 4 段階が「N」、第 5 段階が「N - 1 から N / 4 の範囲」、第 6 段階が「N / 4 から N / 8 の範囲」、第 7 段階が「N / 8 より小さい範囲」である。なお、転送単位は、たとえば、5 段階や 10 段階など任意の段階に設定してもよい。また、各段階の刻み幅も任意の刻み幅に設定してもよい。20

【 0 0 4 5 】

なお、転送単位の初期値は、第 4 段階の N（たとえば、メインフレーム用途で「8」k b y t e、バックアップ用途で「1000」k b y t e など）が設定される。転送単位は、R A I D グループの稼働状態に応じて所要のタイミングで更新される。要求性能は、転送単位ごとに要求される性能である。たとえば、転送単位 N のとき、要求性能は、「200」M B / s である。

【 0 0 4 6 】

[ステップ S 15] R A I D 制御部 16 は、R A I D グループにおける単位時間当たりの I / O 処理数を算出する。R A I D 制御部 16 は、性能情報テーブル 100 の要求性能を転送単位で除することにより単位時間当たりの I / O 処理数を算出する。たとえば、要求性能「200」M B / s、転送単位 N のとき、単位時間当たりの I / O 処理数は、「200 / N」である。30

【 0 0 4 7 】

[ステップ S 16] R A I D 制御部 16 は、R A I D グループにおける構成ディスク比率を算出する。R A I D 制御部 16 は、ステップ S 13 で取得したディスク実構成数を性能情報テーブル 100 の基本構成ディスク数で除することにより構成ディスク比率を算出する。たとえば、ディスク実構成数「5」、基本構成ディスク数「5」のとき、構成ディスク比率は、「1」であり、ディスク実構成数「6」、基本構成ディスク数「5」のとき、構成ディスク比率は、「1.2」である。構成ディスク比率は、「1」以上の値をとり、値が大きいほど I / O 発行可能数を大きくすることができる。40

【 0 0 4 8 】

[ステップ S 17] R A I D 制御部 16 は、管理時間当たりの I / O 発行可能数を算出する。R A I D 制御部 16 は、ステップ S 15 で算出した単位時間当たりの I / O 処理数と、ステップ S 16 で算出した構成ディスク比率と、管理時間の積により管理時間当たりの I / O 発行可能数を算出する。たとえば、単位時間当たりの I / O 処理数「200 / N」、構成ディスク比率「1」、管理時間「1(秒)」のとき、管理時間当たりの I / O 発行可能数は、「200 / N (= (200 / N) × 1 × 1)」である。なお、管理時間は、あらかじめ設定される値である。50

【0049】

[ステップS18] RAID制御部16は、I/O発行数管理テーブルを生成する。I/O発行数管理テーブルは、RAIDグループごとに生成される。

ここで、I/O発行数管理テーブルについて図8を用いて説明する。図8は、第2の実施形態のI/O発行数管理テーブルの一例を示す図である。I/O発行数管理テーブル110は、RAID制御部16がRAIDグループごとに生成するI/O発行数管理テーブルの一例である。I/O発行数管理テーブル110は、RAIDグループ識別情報、転送単位(kbyte)、およびI/O発行可能数を項目として有する。RAIDグループ識別情報は、RAIDグループを一意に特定可能な識別情報である。I/O発行数管理テーブル110は、RAIDグループ「#1」のI/O発行数管理テーブルである。転送単位(kbyte)は、所定の転送単位を示す。I/O発行数管理テーブル110の転送単位は、性能情報テーブル100の転送単位と同じである。I/O発行可能数は、ステップS17で算出した値を、転送単位「N」に対応するI/O発行可能数「M」に設定する。その他の転送単位のI/O発行可能数は、転送単位「N」に対応するI/O発行可能数「M」に所定の計数を乗じた値が設定される。たとえば、転送単位「N×4からN+1の範囲」に対応するI/O発行可能数は、「k3×M」である。所定の計数「k1」、「k2」、「k3」、「k4」、「k5」、「k6」は、転送単位ごとにあらかじめ設定される。

10

【0050】

RAID制御部16は、生成したI/O発行数管理テーブルをメモリ33に保持する。したがって、RAID制御部16は、RAIDグループごとにI/O発行数管理テーブルをメモリ33に保持する。

20

【0051】

[ステップS19] RAID制御部16は、I/O発行数管理テーブルを参照して、設定されている転送単位に対応するI/O発行可能数を取得して設定する。たとえば、RAID制御部16は、I/O発行数管理テーブル110を参照して、初期値として設定されている転送単位「N」に対応するI/O発行可能数「M」を取得し、RAIDグループ「#1」のI/O発行可能数に「M」を設定する。

【0052】

[ステップS20] RAID制御部16は、新規に追加されたRAIDグループのI/O制御を開始し、RAIDグループ追加処理を終了する。

30

次に、第2の実施形態のRAID制御部16が実行するRAIDグループI/O受付処理について図9を用いて説明する。図9は、第2の実施形態のRAIDグループI/O受付処理のフローチャートを示す図である。RAIDグループI/O受付処理は、RAID制御部16がホストI/O制御部15からI/O要求を受け付けることにより実行される。RAIDグループI/O受付処理は、ホストI/O制御部15から受け付けたI/O要求をI/O発行待ちキューにキューイングするとともに、RAIDグループ追加処理で設定したI/O発行可能数を更新する処理である。

【0053】

[ステップS21] RAID制御部16(より詳しくはI/O発行可能数算出部25)は、I/O受付部22がホストI/O制御部15から受け付けたI/O要求を、I/O要求に対応するRAIDグループごとのI/O発行待ちキューにキューイングする。

40

【0054】

[ステップS22] RAID制御部16は、RAIDグループごとのI/O要求の統計情報を更新する。統計情報は、所定の監視時間当たりの転送単位使用頻度を含み、統計情報管理テーブルに記録される。統計情報管理テーブルは、メモリ33に保持される。ここで、統計情報管理テーブルについて図10を用いて説明する。図10は、第2の実施形態の統計情報管理テーブルの一例を示す図である。

【0055】

統計情報管理テーブル120は、RAID制御部16がRAIDグループごとに管理する統計情報管理テーブルの一例である。統計情報管理テーブル120は、RAIDグルー

50

プロセス情報、転送単位(k b y t e)、および受付 I / O 数を項目として有する。 R A I D グループ識別情報は、 R A I D グループを一意に特定可能な識別情報である。

【 0 0 5 6 】

統計情報管理テーブル 120 は、 R A I D グループ「 #1 」の I / O 発行数管理テーブルである。転送単位(k b y t e)は、所定の転送単位を示す。統計情報管理テーブル 120 の転送単位は、性能情報テーブル 100 の転送単位、および I / O 発行数管理テーブル 110 の転送単位と同じである。受付 I / O 数は、所定の監視時間当たりの転送単位使用頻度である。より詳しくは、受付 I / O 数は、 I / O 受付部 22 が受け付けた I / O 要求のデータサイズに対応する転送単位ごとの I / O 要求の受付数である。たとえば、転送単位「 N 」の受付 I / O 数「 30 」は、所定の監視時間の間に、データサイズ「 N 」の I / O 要求を 30 回受け付けたことを示す。
10

【 0 0 5 7 】

なお、所定の監視時間は、たとえば、 60 分など統計的に有意な情報を収集できる時間を任意に設定することができる。監視時間は、 R A I D グループで共通の値としてもよいし、 R A I D グループごとに異なる値としてもよい。また、監視時間は、あらかじめ設定する固定値としてもよいし、時間帯や時間当たりの I / O 要求数に応じた可変値としてもよい。

【 0 0 5 8 】

[ステップ S 23] R A I D 制御部 16 は、 I / O 発行可能数の更新タイミングであるか否かを判定する。更新タイミングは、あらかじめ設定した更新周期(たとえば、 1 分など)であってもよいし、所定のトリガ検出(たとえば、統計情報管理テーブル 120 の受付 I / O 数のいずれかが所定値を超えたなど)であってもよい。 R A I D 制御部 16 は、 I / O 発行可能数の更新タイミングである場合にステップ S 24 にすすみ、 I / O 発行可能数の更新タイミングでない場合に R A I D グループ I / O 受付処理を終了する。
20

【 0 0 5 9 】

[ステップ S 24] R A I D 制御部 16 は、統計情報管理テーブルを参照して受付 I / O 数のボリュームゾーンを選択する。たとえば、 R A I D 制御部 16 は、統計情報管理テーブル 120 を参照して、受付 I / O 数のボリュームゾーンとして受付 I / O 数が最大の「 1000 」となる転送単位「 N / 4 から N / 8 の範囲」を選択する。なお、ボリュームゾーンとして選択する転送単位は、受付 I / O 数が最大となる場合に限らず、平均値や中央値など所定の選択基準に従い選択されるものであってもよい。
30

【 0 0 6 0 】

[ステップ S 25] R A I D 制御部 16 は、 I / O 発行数管理テーブルを参照して、選択した転送単位に対応する I / O 発行可能数を取得して設定する。たとえば、 R A I D 制御部 16 は、転送単位「 N / 4 から N / 8 の範囲」を選択した場合、 I / O 発行数管理テーブル 110 を参照して、転送単位「 N / 4 から N / 8 の範囲」に対応する I / O 発行可能数「 k 5 × M 」を取得し、 R A I D グループ「 #1 」の I / O 発行可能数に「 k 5 × M 」を設定する。 R A I D 制御部 16 は、 I / O 発行可能数を設定した後、 R A I D グループ I / O 受付処理を終了する。

【 0 0 6 1 】

このように、 I / O 発行可能数は、受け付けた I / O 要求に応じて適宜更新される。

次に、第 2 の実施形態の R A I D 制御部 16 が実行する R A I D グループ I / O 発行数管理処理について図 11 を用いて説明する。図 11 は、第 2 の実施形態の R A I D グループ I / O 発行数管理処理のフローチャートを示す図である。 R A I D グループ I / O 発行数管理処理は、 R A I D グループの I / O 制御の開始を受けて実行される。 R A I D グループ I / O 受付処理は、 I / O 発行待ちキューにキューイングされている I / O 要求を I / O 発行可能数の範囲に制限しながら R A I D グループに I / O 要求を発行する処理である。

【 0 0 6 2 】

[ステップ S 31] R A I D 制御部 16 (より詳しくは I / O 発行数管理部 23) は、

50

I / O 発行待ちキューにキューイングされている I / O 要求の数（受付 I / O 発行数）が「0」より大きいか否かを判定する。すなわち、RAID 制御部 16 は、I / O 発行待ちキューにキューイングされている I / O 要求があるか否かを判定する。RAID 制御部 16 は、I / O 発行待ちキューにキューイングされている I / O 要求がある場合にステップ S32 にすすみ、I / O 要求がない場合に I / O 発行待ちキューへのキューイングを待ち受ける。

【0063】

[ステップ S32] RAID 制御部 16 は、受付 I / O 発行数が RAID グループに設定されている I / O 発行可能数を超えているか否かを判定する。RAID 制御部 16 は、受付 I / O 発行数が I / O 発行可能数を超えている場合にステップ S33 にすすみ、受付 I / O 発行数が I / O 発行可能数を超えていない場合にステップ S34 にすすむ。
10

【0064】

[ステップ S33] RAID 制御部 16 は、I / O 発行待ちキューにキューイングされている I / O 要求のうちから I / O 発行可能数だけ I / O 要求をディスク I / O 制御部 17 に向けて発行する。

【0065】

[ステップ S34] RAID 制御部 16 は、I / O 発行待ちキューにキューイングされている I / O 要求をディスク I / O 制御部 17 に向けて発行する。

[ステップ S35] RAID 制御部 16 は、I / O 要求の発行後の管理時間の経過を待つ。管理時間は、RAID グループ追加処理のステップ S17 で I / O 発行可能数を算出した際に用いた管理時間である。RAID 制御部 16 は、I / O 要求の発行後の管理時間を経過している場合にステップ S31 にすすみ、I / O 要求の発行後の管理時間を経過していない場合に管理時間の経過を待つ。
20

【0066】

このようにして、RAID 制御部 16 は、RAID グループに向けて発行する I / O 要求の数を、管理時間当たり I / O 発行可能数の範囲に制限することができる。したがって、RAID グループを構成する一部の HDD 19 の処理性能が過大であっても、RAID グループの I / O 処理数が過大になることがない。

【0067】

したがって、RAID 装置 12 は、性能の上限が多様な HDD 19 を搭載しても、RAID グループを構成したときに所定の性能値を得ることができる。このような RAID 装置 12 は、搭載する HDD 19 のマルチベンダ化の推進の障害となることがない。また、このような RAID 装置 12 は、搭載する HDD 19 の性能最適化に伴う、RAID 装置ベンダおよび HDD ベンダの作業負担を軽減することができる。
30

【0068】

また、このような RAID 装置 12 は、RAID グループを構成する HDD 19 を故障などにより交換したような場合に、RAID グループの応答性能が著しく変化することができない。したがって、RAID 装置 12 は、メンテナンスの前後においてユーザに対して安定した性能を発揮する RAID グループを提供することができる。

【0069】

また、RAID 装置 12 は、新規に RAID グループを構成したときに、新規に構成した RAID グループとすでにある RAID グループとの間で性能差が過大になることがない。したがって、RAID 装置 12 は、ユーザに対して安定した性能を発揮する RAID グループを提供することができる。
40

【0070】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、ストレージ制御装置 1、RAID 装置 12、コントローラモジュール 13 などが有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピ
50

ユータで読み取り可能な記録媒体としては、磁気記憶装置、光ディスク、光磁気記録媒体、半導体メモリなどがある。磁気記憶装置には、ハードディスク装置（HDD）、フレキシブルディスク（FD）、磁気テープなどがある。光ディスクには、DVD、DVD-RAM、CD-ROM/RWなどがある。光磁気記録媒体には、MO（Magneto-Optical disk）などがある。

【0071】

プログラムを流通させる場合には、たとえば、そのプログラムが記録されたDVD、CD-ROMなどの可搬型記録媒体が販売される。また、プログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。

10

【0072】

プログラムを実行するコンピュータは、たとえば、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、自己の記憶装置に格納する。そして、コンピュータは、自己の記憶装置からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。また、コンピュータは、ネットワークを介して接続されたサーバコンピュータからプログラムが転送されるごとに、逐次、受け取ったプログラムに従った処理を実行することもできる。

【0073】

また、上記の処理機能の少なくとも一部を、DSP、ASIC、PLDなどの電子回路で実現することもできる。

20

【符号の説明】

【0074】

- 1 ストレージ制御装置
- 2 設定部
- 3 発行制御部
- 4 要求蓄積部
- 5 ストレージ装置
- 6 I/O発行可能数
- 7 I/O要求
- 9 ホストコンピュータ
- 10 ストレージシステム
- 11 ホストコンピュータ
- 12 RAID装置
- 13 コントローラモジュール（CM）
- 14 ドライブエンクロージャ（DE）
- 15 ホストI/O制御部
- 16 RAID制御部
- 17 ディスクI/O制御部
- 19 HDD
- 20, 22, 29 I/O受付部
- 21, 28, 30 I/O完了受付部
- 23 I/O発行数管理部
- 24 管理時間監視部
- 25 I/O発行可能数算出部
- 26 性能情報テーブル
- 27 I/O発行可能数
- 31 チャネルアダプタ
- 32 プロセッサ
- 33 メモリ

30

40

50

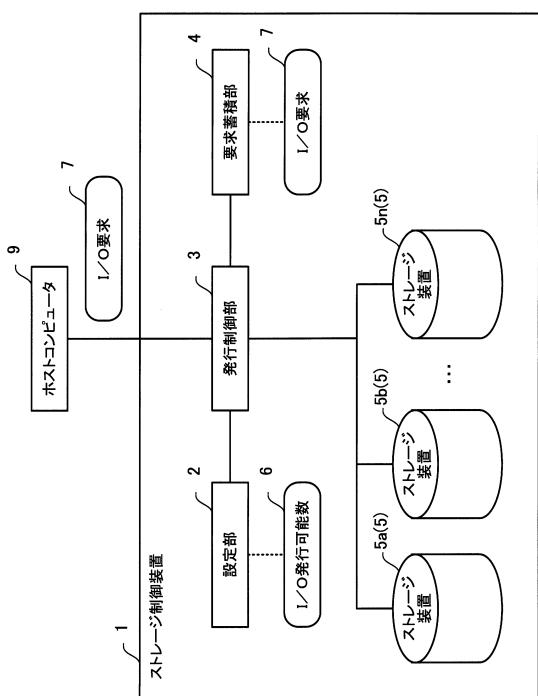
3 4 ディスクアダプタ

1 0 0 性能情報テーブル

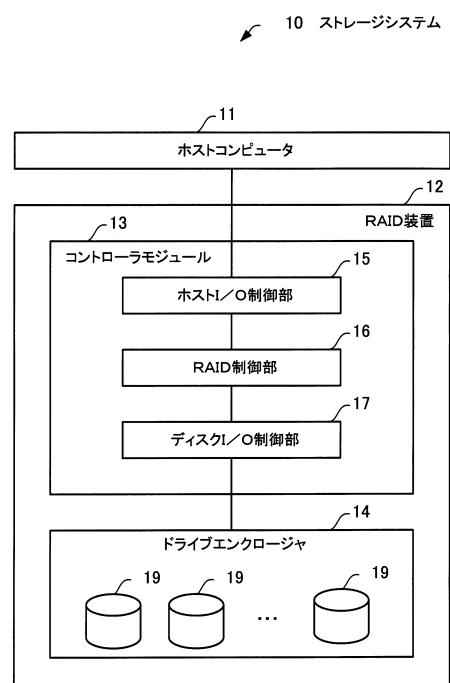
1 1 0 I / O 発行数管理テーブル

1 2 0 統計情報管理テーブル

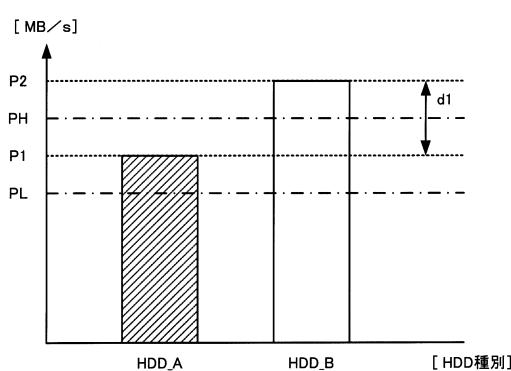
【図 1】



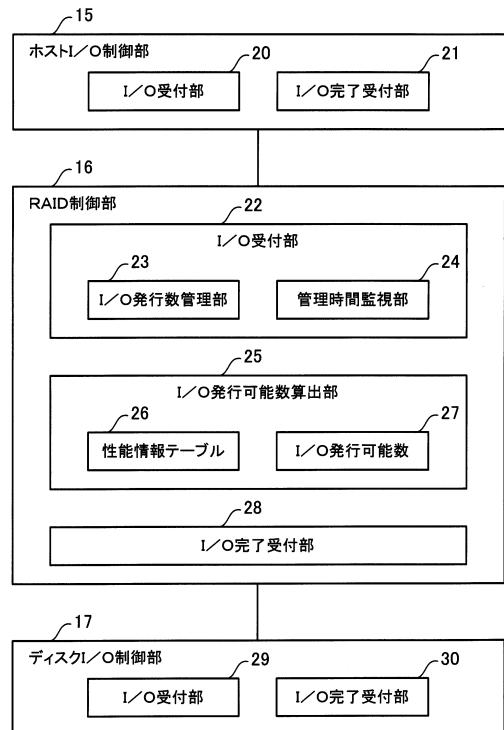
【図 2】



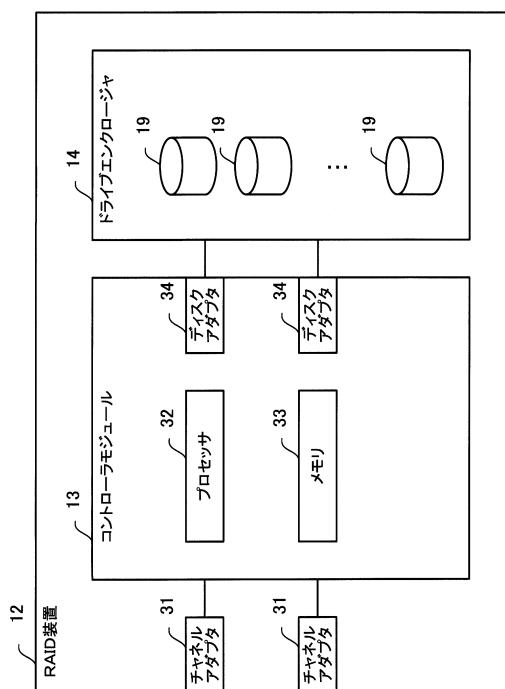
【図3】



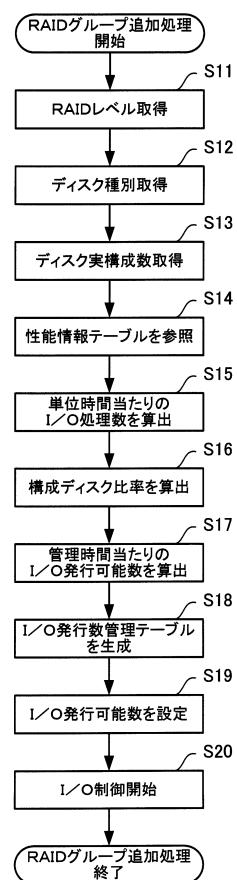
【図4】



【図5】



【図6】



【図7】

100 性能情報テーブル

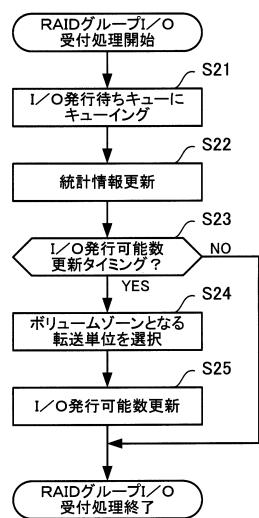
| RAIDレベル | RAIDグループ識別 (ディスク種別 (回転数) | 基本構成 ディスク数 | 転送単位 (kbyte) | 要求性能 (MB/s) |
|---------|--------------------------------|---------------|-----------------|----------------|
| RAID5 | | 5 | >(N×8) | 400 |
| | | | (N×8)～(N+1) | 350 |
| | | | (N×4)～(N×4) | 300 |
| | | | N | 200 |
| | | | (N-1)～(N/4) | 100 |
| | | | (N/4)～(N/8) | 10 |
| | | | <(N/8) | 8 |

【図8】

110 I/O発行数管理テーブル

| RAIDグループ識別情報 | 転送単位(kbyte) | I/O発行可能数 |
|--------------|-------------|----------|
| #1 | >(N×8) | k1×M |
| | (N×8)～(N×4) | k2×M |
| | (N×4)～(N+1) | k3×M |
| | N | M |
| | (N-1)～(N/4) | k4×M |
| | (N/4)～(N/8) | k5×M |
| | <(N/8) | k6×M |

【図9】

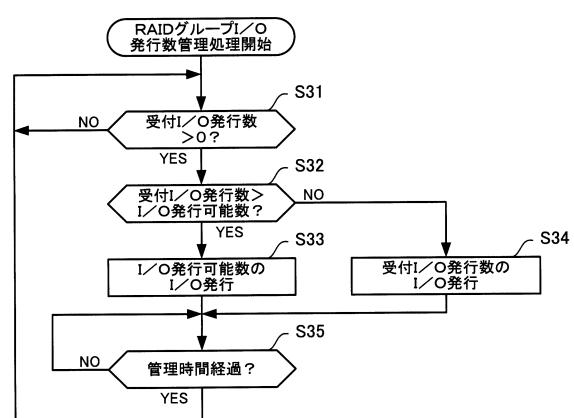


【図10】

120 統計情報管理テーブル

| RAIDグループ識別情報 | 転送単位(kbyte) | 受付I/O数 |
|--------------|-------------|--------|
| #1 | >(N×8) | 0 |
| | (N×8)～(N×4) | 10 |
| | (N×4)～(N+1) | 5 |
| | N | 30 |
| | (N-1)～(N/4) | 5 |
| | (N/4)～(N/8) | 1000 |
| | <(N/8) | 0 |

【図11】



フロントページの続き

(56)参考文献 特開平11-167470(JP,A)
米国特許第6366981(US,B1)
特開2009-294949(JP,A)
米国特許出願公開第2009/0307534(US,A1)
特開2005-332236(JP,A)
特開2008-186211(JP,A)
米国特許出願公開第2015/0261462(US,A1)
米国特許第7552280(US,B1)
米国特許出願公開第2013/0212349(US,A1)

(58)調査した分野(Int.Cl., DB名)

G 06 F 3 / 06 - 3 / 08
G 06 F 13 / 10 - 13 / 14