

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-125175  
(P2019-125175A)

(43) 公開日 令和1年7月25日(2019.7.25)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 12/00 (2006.01)</b>	G06F 12/00 501B	
<b>G06F 13/14 (2006.01)</b>	G06F 13/14 330A	
<b>G06F 3/06 (2006.01)</b>	G06F 3/06 301E	
	G06F 3/06 302E	

審査請求 未請求 請求項の数 8 O L (全 24 頁)

(21) 出願番号 特願2018-5583 (P2018-5583)  
(22) 出願日 平成30年1月17日 (2018.1.17)

(71) 出願人 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 見尾 和俊  
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 阿部 智明  
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

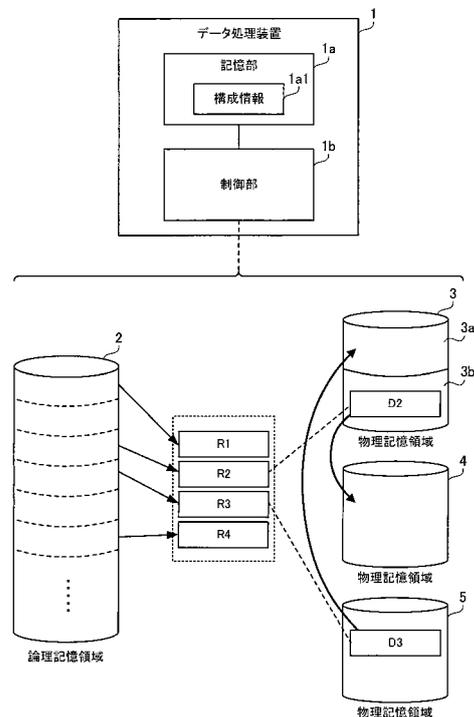
(54) 【発明の名称】 データ処理装置、データ処理システムおよびデータ処理プログラム

(57) 【要約】

【課題】 データ移動処理時間を短縮する。

【解決手段】 論理記憶領域 2 に含まれる各单位領域に対して、アクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される物理記憶領域 3 ~ 5 のいずれかから、記憶領域が割り当てられる。制御部 1 b は、記憶領域の割り当て元の物理記憶領域を変更すべき単位領域 R 1 ~ R 4 の中から、データ移動の候補を順次選択し、選択された単位領域のデータを物理記憶領域間で移動させるように指示する移動制御処理を実行する。この移動制御処理は、単位領域 R 3 を選択したとき、単位領域 R 3 のデータ D 3 の移動先となる物理記憶領域 3 を分割した分割領域 3 a , 3 b のそれぞれにおいて、他の単位領域についてのデータ移動が実行中かを判定し、判定結果に基づいて、単位領域 R 3 についてのデータ移動の指示処理を制御する処理を含む。

【選択図】 図 1



**【特許請求の範囲】****【請求項 1】**

アクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される複数の物理記憶領域の構成を示す構成情報を記憶する記憶部と、

前記構成情報に基づき、論理記憶領域に含まれる複数の単位領域のそれぞれに対して、前記複数の物理記憶領域のいずれかから記憶領域が割り当てられた状態において、前記複数の単位領域の中から、記憶領域の割り当て元の物理記憶領域を変更すべき複数の第 1 単位領域が特定されたとき、前記複数の第 1 単位領域の中からデータ移動の候補を順次選択し、前記候補として選択された第 1 単位領域のデータを物理記憶領域間で移動させるように指示する移動制御処理を実行する制御部と、

を有するデータ処理装置であって、

前記移動制御処理は、

前記複数の第 1 単位領域の中から前記候補として第 2 単位領域を選択したとき、前記第 2 単位領域のデータの移動先となる物理記憶領域を分割した複数の分割領域のそれぞれにおいて、移動指示済みの単位領域についてのデータ移動が実行中であるかを判定し、

前記判定の結果に基づいて、前記第 2 単位領域についてのデータ移動の指示処理を制御する、

処理を含むデータ処理装置。

**【請求項 2】**

前記指示処理の制御では、前記複数の分割領域の中に、移動指示済みの単位領域についてのデータ移動が実行されていない一の分割領域が存在した場合、前記第 2 単位領域のデータを前記一の分割領域に移動させるように指示する、

請求項 1 記載のデータ処理装置。

**【請求項 3】**

前記指示処理の制御では、前記複数の分割領域のすべてにおいて、移動指示済みの単位領域についてのデータ移動が実行中であった場合、前記第 2 単位領域についてのデータ移動を指示待ち状態とし、前記複数の第 1 単位領域の中から第 3 単位領域を前記候補として選択する、

請求項 1 または 2 記載のデータ処理装置。

**【請求項 4】**

前記複数の第 1 単位領域のそれぞれにはデータ移動の実行優先度が設定されており、

前記候補は、前記複数の第 1 単位領域の中から前記実行優先度の高い順に選択され、

前記移動制御処理は、前記候補として前記第 2 単位領域を選択したとき、前記第 2 単位領域における直近のアクセス頻度が所定値未満である場合には、前記複数の第 1 単位領域の中から、前記第 2 単位領域と同一の前記実行優先度が設定され、かつデータ移動が指示されていない第 4 単位領域を特定し、前記第 4 単位領域の中に、直近のアクセス頻度が前記所定値以上である第 5 単位領域が存在する場合には、前記第 2 単位領域よりも先に前記第 5 単位領域についてのデータ移動を指示する処理をさらに含む、

請求項 1 乃至 3 のいずれか 1 項に記載のデータ処理装置。

**【請求項 5】**

前記複数の第 1 単位領域のそれぞれについてのデータ移動元およびデータ移動先の物理記憶領域は、決定タイミングより前の期間に計測された、前記複数の第 1 単位領域のそれぞれにおけるアクセス頻度に基づいて決定され、

前記第 2 単位領域および前記第 4 単位領域についての前記直近のアクセス頻度は、前記決定タイミングの後の期間において計測されたアクセス頻度である、

請求項 4 記載のデータ処理装置。

**【請求項 6】**

前記複数の物理記憶領域が複数組存在し、前記論理記憶領域が複数存在し、前記複数の物理記憶領域のそれぞれを用いて、複数の前記論理記憶領域のうちそれぞれ個別の論理記憶領域を用いて前記移動制御処理が実行され、それぞれの前記移動制御処理において前記

10

20

30

40

50

単位領域のサイズが互いに異なり、

前記複数の第1単位領域のそれぞれにはデータ移動の実行優先度が設定されており、前記候補は、前記複数の第1単位領域の中から前記実行優先度の高い順に選択され、

前記移動制御処理は、前記候補として前記第2単位領域を選択したとき、複数の前記論理記憶領域に含まれる前記第1単位領域の中に、前記第2単位領域と同一の前記実行優先度が設定され、かつ前記第2単位領域より小さいサイズの第6単位領域が存在する場合には、前記第2単位領域よりも先に前記第6単位領域についてのデータ移動を指示する処理をさらに含む、

請求項1乃至5のいずれか1項に記載のデータ処理装置。

【請求項7】

10

複数の記憶装置と、

論理記憶領域に含まれる複数の単位領域のそれぞれに対して、前記複数の記憶装置のうちアクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される複数の物理記憶領域のいずれかから、記憶領域が割り当てられた状態において、前記複数の単位領域の中から、記憶領域の割り当て元の物理記憶領域を変更すべき複数の第1単位領域が特定されたとき、前記複数の第1単位領域の中からデータ移動の候補を順次選択し、前記候補として選択された第1単位領域のデータを物理記憶領域間で移動させるように指示する移動制御処理を実行するデータ処理装置と、

を有するデータ処理システムであって、

前記移動制御処理は、

20

前記複数の第1単位領域の中から前記候補として第2単位領域を選択したとき、前記第2単位領域のデータの移動先となる物理記憶領域を分割した複数の分割領域のそれぞれにおいて、移動指示済みの単位領域についてのデータ移動が実行中であるかを判定し、

前記判定の結果に基づいて、前記第2単位領域についてのデータ移動の指示処理を制御する、

処理を含むデータ処理システム。

【請求項8】

コンピュータに、

論理記憶領域に含まれる複数の単位領域のそれぞれに対して、アクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される複数の物理記憶領域のいずれかから、記憶領域が割り当てられた状態において、前記複数の単位領域の中から、記憶領域の割り当て元の物理記憶領域を変更すべき複数の第1単位領域が特定されたとき、前記複数の第1単位領域の中からデータ移動の候補を順次選択し、前記候補として選択された第1単位領域のデータを物理記憶領域間で移動させるように指示する移動制御処理、

30

を実行させるデータ処理プログラムであって、

前記移動制御処理は、

前記複数の第1単位領域の中から前記候補として第2単位領域を選択したとき、前記第2単位領域のデータの移動先となる物理記憶領域を分割した複数の分割領域のそれぞれにおいて、移動指示済みの単位領域についてのデータ移動が実行中であるかを判定し、

前記判定の結果に基づいて、前記第2単位領域についてのデータ移動の指示処理を制御する、

40

処理を含むデータ処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、データ処理装置、データ処理システムおよびデータ処理プログラムに関する。

【背景技術】

【0002】

ストレージ装置内の記憶装置をアクセス性能に応じて階層化し、階層間でデータを再配

50

置するストレージ階層制御が知られている。例えば、アクセス頻度の高いデータはアクセス速度の高い記憶装置に配置され、アクセス頻度の低いデータはアクセス速度の低い記憶装置に配置される。このようにデータを配置することで、迅速なアクセス処理を実現できる。

【0003】

また、ストレージ階層制御技術を用いたシステムの一例として、次のような計算機システムが提案されている。この計算機システムでは、データベース管理システム上のオブジェクト（例えば、テーブル、インデックスなど）のデータの格納先が制御される。各オブジェクトにはその種別に応じて優先度が設定され、優先度の高いオブジェクトのデータが優先的に再配置される。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2014-199596号公報

【特許文献2】国際公開第2013/164878号

【発明の概要】

【発明が解決しようとする課題】

【0005】

ストレージ階層制御では、例えば、データごとのアクセス頻度に基づいて、階層間で移動させる移動対象のデータが特定される。ここで、移動対象のデータが複数特定された場合、これらのデータの移動処理にかかる全体の時間を短縮することで、各データを早期に適切な階層に配置でき、その結果、各データに対するアクセス性能を向上させることができる。しかし、各データの移動の実行をどのように制御すれば、全体の移動処理時間を短縮できるかという点に課題がある。

【0006】

1つの側面では、本発明は、データ移動処理時間を短縮することが可能なデータ処理装置、データ処理システムおよびデータ処理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】

【0007】

1つの案では、記憶部と制御部とを有する次のようなデータ処理装置が提供される。このデータ処理装置において、記憶部は、アクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される複数の物理記憶領域の構成を示す構成情報を記憶する。制御部は、構成情報に基づき、論理記憶領域に含まれる複数の単位領域のそれぞれに対して、複数の物理記憶領域のいずれかから記憶領域が割り当てられた状態において、複数の単位領域の中から、記憶領域の割り当て元の物理記憶領域を変更すべき複数の第1単位領域が特定されたとき、複数の第1単位領域の中からデータ移動の候補を順次選択し、候補として選択された第1単位領域のデータを物理記憶領域間で移動させるように指示する移動制御処理を実行する。また、この移動制御処理は、複数の第1単位領域の中から候補として第2単位領域を選択したとき、第2単位領域のデータの移動先となる物理記憶領域を分割した複数の分割領域のそれぞれにおいて、移動指示済みの単位領域についてのデータ移動が実行中であるかを判定し、判定の結果に基づいて、第2単位領域についてのデータ移動の指示処理を制御する、処理を含む。

【0008】

また、1つの案では、複数の記憶装置と、データ処理装置とを有するデータ処理システムが提供される。データ処理装置は、論理記憶領域に含まれる複数の単位領域のそれぞれに対して、複数の記憶装置のうちアクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される複数の物理記憶領域のいずれかから、記憶領域が割り当てられた状態において、複数の単位領域の中から、記憶領域の割り当て元の物理記憶領域を変更すべき複数の第1単位領域が特定されたとき、複数の第1単位領域の中からデータ移動の候補を順次選択し、候補として選択された第1単位領域のデータを物理記憶領域間で移動させるように指

10

20

30

40

50

示す移動制御処理を実行する。また、この移動制御処理は、複数の第1単位領域の中から候補として第2単位領域を選択したとき、第2単位領域のデータの移動先となる物理記憶領域を分割した複数の分割領域のそれぞれにおいて、移動指示済みの単位領域についてのデータ移動が実行中であるかを判定し、判定の結果に基づいて、第2単位領域についてのデータ移動の指示処理を制御する、処理を含む。

【0009】

さらに、1つの案として、上記のデータ処理装置と同様の処理をコンピュータに実行させるデータ処理プログラムが提供される。

【発明の効果】

【0010】

1つの側面では、データ移動処理時間を短縮できる。

【図面の簡単な説明】

【0011】

【図1】第1の実施の形態に係るデータ処理装置の構成例および処理例を示す図である。

【図2】第2の実施の形態に係るストレージシステムの構成例を示す図である。

【図3】階層化された記憶領域の設定例を示す図である。

【図4】サブプールに対するRAIDグループの設定例を示す図である。

【図5】CMおよび管理サーバが備える処理機能の構成例を示すブロック図である。

【図6】CMに記憶される単位領域管理テーブルのデータ構成例を示す図である。

【図7】CMに記憶されるプール管理テーブルのデータ構成例を示す図である。

【図8】優先度設定用テーブルのデータ構成例を示す図である。

【図9】アクセス頻度の収集処理手順を示すシーケンス図の例である。

【図10】データの配置先決定処理手順を示すフローチャートの例である。

【図11】データ移動処理手順を示すフローチャートの例(その1)である。

【図12】データ移動処理手順を示すフローチャートの例(その2)である。

【発明を実施するための形態】

【0012】

以下、本発明の実施の形態について図面を参照して説明する。

〔第1の実施の形態〕

図1は、第1の実施の形態に係るデータ処理装置の構成例および処理例を示す図である。図1に示すデータ処理装置1は、記憶部1aと制御部1bを有する。なお、記憶部1aは、RAM(Random Access Memory)やHDD(Hard Disk Drive)など、データ処理装置1が備える記憶装置の記憶領域によって実現される。制御部1bは、例えば、データ処理装置1が備えるプロセッサとして実現される。

【0013】

このデータ処理装置1は、論理記憶領域2に含まれる複数の単位領域のそれぞれに対して、物理記憶領域を割り当てる。また、各単位領域への割り当て元となる物理記憶領域は複数用意される。本実施の形態では、例として3つの物理記憶領域3~5が用意される。これらの物理記憶領域3~5は、アクセス性能が互いに異なる記憶装置を用いてそれぞれ実現される。そして、論理記憶領域2の各単位領域に対しては、物理記憶領域3~5のいずれかから記憶領域が割り当てられる。ただし、論理記憶領域2の単位領域のうち、少なくともデータが格納されている単位領域に対してのみ、物理記憶領域3~5のいずれかから記憶領域が割り当てられればよい。

【0014】

記憶部1aには、物理記憶領域3~5の構成を示す構成情報1a1が記憶される。例えば、構成情報1a1には、物理記憶領域3~5の容量や、物理記憶領域3~5を実現する記憶装置に関する情報が含まれる。

【0015】

制御部1bは、上記のように、論理記憶領域2の複数の単位領域に対して、物理記憶領域3~5のいずれかから記憶領域が割り当てられている状態において、次のような処理を

10

20

30

40

50

実行する。論理記憶領域 2 の単位領域の中から、記憶領域の割り当て元の物理記憶領域を変更すべき複数の第 1 単位領域が特定されると、制御部 1 b は、これらの複数の第 1 単位領域の中からデータ移動の候補を順次選択する。そして、制御部 1 b は、候補として選択された第 1 単位領域のデータを物理記憶領域間で移動させるように指示する。このように、データ移動の候補を順次選択し、選択された候補のデータの移動指示を行う処理を、ここでは「移動制御処理」と記載する。

【0016】

図 1 では例として、第 1 単位領域として単位領域 R 1 ~ R 4 が特定されたとする。制御部 1 b は、例えば、単位領域 R 1 , R 2 , R 3 , R 4 の順にデータ移動の候補を選択し、選択された単位領域についての移動制御処理を実行する。この移動制御処理では、詳細には、次のような処理が実行される。

10

【0017】

制御部 1 b は、第 1 単位領域の中から候補として 1 つの単位領域（第 2 単位領域とする）を選択したとき、第 2 単位領域のデータの移動先となる物理記憶領域を分割した複数の分割領域のそれぞれにおいて、移動指示済みの単位領域についてのデータ移動が実行中であるかを判定する。

【0018】

例えば、データ移動の候補として単位領域 R 3 が選択されたとする。また、単位領域 R 3 については、記憶領域の割り当て元を物理記憶領域 5 から物理記憶領域 3 へ変更するように要求されており、単位領域 R 3 のデータ D 3 の移動先は物理記憶領域 3 であるとする。さらに、物理記憶領域 3 は、2 つの分割領域 3 a , 3 b に分割されているとする。なお、分割領域 3 a , 3 b は、それぞれ個別の記憶装置によって実現される。例えば、分割領域 3 a , 3 b は、個別の R A I D (Redundant Arrays of Inexpensive Disks) グループとして実現される。

20

【0019】

この場合、制御部 1 b は、分割領域 3 a , 3 b のそれぞれにおいて、移動指示済みの単位領域（このケースでは単位領域 R 1 , R 2 のいずれか）についてのデータ移動が実行中であるかを判定する。そして、制御部 1 b は、その判定結果に基づいて、単位領域 R 3 についてのデータ移動の指示処理を制御する。これにより、データ移動処理時間を短縮できる。

30

【0020】

例えば、分割領域 3 b において、単位領域 R 2 のデータ D 2 の移動が実行中であったとする。図 1 では例として、データ D 2 を分割領域 3 b から物理記憶領域 4 へ移動する処理が実行中であったとする。この場合に、単位領域 R 3 のデータ D 3 を分割領域 3 b へ移動させようとする、データ D 2 の移動が完了するまでデータ D 3 は移動待ちになってしまう。そこで、制御部 1 b は例えば、データ D 3 を、データ移動が実行中でない分割領域 3 a を移動させるように指示する。これにより、データ D 3 の移動待ちが発生することなく、データ D 2 の移動とデータ D 3 の移動とが並列に実行される。その結果、データ移動にかかる全体の時間を短縮できる。

【0021】

また、図示しないが、例えば、分割領域 3 a , 3 b のそれぞれにおいて、単位領域 R 3 以外の単位領域についてのデータ移動が実行中であったとする。この場合、制御部 1 b は例えば、単位領域 R 3 についてのデータ移動を指示待ち状態とし、データ移動の候補として次の単位領域 R 4 を選択して、単位領域 R 4 についての移動制御処理を実行する。これにより、単位領域 R 3 のデータ D 3 の移動待ちによって新たな単位領域 R 4 のデータ移動が実行されなくなる無駄な時間の発生確率を低減でき、データ移動効率を向上させることができる。その結果、データ移動にかかる全体の時間を短縮できる。

40

【0022】

このように、分割領域 3 a , 3 b のそれぞれにおいてデータ移動が実行中であるかを判定し、その判定結果に基づいて単位領域 R 3 についてのデータ移動の指示処理を制御する

50

ことで、データ移動処理時間を短縮できる。

【0023】

〔第2の実施の形態〕

次に、第2の実施の形態に係るストレージシステムについて説明する。

図2は、第2の実施の形態に係るストレージシステムの構成例を示す図である。図2に示すストレージシステムは、ストレージ装置100、ホストサーバ200、管理サーバ300および管理端末400を含む。ストレージ装置100、ホストサーバ200、管理サーバ300および管理端末400は、ネットワーク500を介して相互に接続されている。ネットワーク500は、例えば、LAN (Local Area Network) でもよいし、WAN (Wide Area Network) やインターネットなどの広域ネットワークでもよい。また、ストレージ装置100とホストサーバ200との間は、SAN (Storage Area Network) を通じて通信が行われてもよい。

10

【0024】

ストレージ装置100は、CM (Controller Module) 110とドライブ部120を有する。CM 110は、ホストサーバ200からの要求に応じて、ドライブ部120に搭載された記憶装置にアクセスするストレージ制御装置である。例えば、CM 110は、ドライブ部120に搭載された記憶装置の記憶領域を用いた論理ボリュームを設定し、ホストサーバ200から論理ボリュームに対するアクセス要求を受け付ける。なお、CM 110は、図1に示したデータ処理装置1の一例である。

【0025】

ドライブ部120には、ホストサーバ200からのアクセス対象となる記憶装置121a, 121b, 121c, …が搭載されている。ドライブ部120には、アクセス性能の異なる複数種類の記憶装置が搭載されている。例えば、ドライブ部120には、アクセス性能の低い順に、ニアラインHDD、オンラインHDD、SSD (Solid State Drive) が搭載されている。

20

【0026】

ホストサーバ200は、例えば、種々の業務処理を実行するコンピュータである。ホストサーバ200は、論理ボリュームに対するアクセス要求をCM 110に送信することで、論理ボリュームにアクセスする。

【0027】

管理サーバ300は、ストレージ装置100の運用を管理するサーバコンピュータである。例えば、管理サーバ300は、ストレージ装置100における階層制御を行う。具体的には、管理サーバ300は、論理ボリューム上のデータのアクセス頻度をCM 110から定期的に収集し、論理ボリューム上のデータがそのアクセス頻度に応じた適切なアクセス性能を有する記憶装置に配置されるように制御する。

30

【0028】

管理端末400は、ストレージシステムの管理者が利用する端末装置である。例えば、管理端末400は、管理者の操作により、論理ボリュームの設定や、記憶領域の階層化に関する設定などを実行する。

【0029】

次に、図2を用いて、CM 110のハードウェア構成例について説明する。図2に示すように、CM 110は、プロセッサ111、RAM 112、SSD 113、ネットワークインタフェース (I/F) 114およびドライブインタフェース (I/F) 115を有する。

40

【0030】

プロセッサ111は、CM 110全体を統括的に制御する。プロセッサ111は、例えば、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。また、プロセッサ111は、CPU、MPU、DSP、ASIC、PLDのうちの2以上の要素の組み合わせであっても

50

よい。

【0031】

R A M 1 1 2 は、C M 1 1 0 の主記憶装置として使用される。R A M 1 1 2 には、プロセッサ 1 1 1 に実行させる O S (Operating System) プログラムや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、R A M 1 1 2 には、プロセッサ 1 1 1 による処理に必要な各種データが格納される。

【0032】

S S D 1 1 3 は、C M 1 1 0 の補助記憶装置として使用される。S S D 1 1 3 には、O S プログラム、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。なお、補助記憶装置としては、H D D などの他の種類の不揮発性記憶装置を使用することもできる。

10

【0033】

ネットワークインタフェース 1 1 4 は、ネットワーク 5 0 0 を介してホストサーバ 2 0 0 や管理サーバ 3 0 0 と通信するための通信インタフェースである。なお、例えば、ホストサーバ 2 0 0 と S A N などを介して接続する通信インタフェースと、管理サーバ 3 0 0 と L A N などを介して接続する通信インタフェースとが、個別に設けられていてもよい。

【0034】

ドライブインタフェース 1 1 5 は、ドライブ部 1 2 0 に搭載された記憶装置 1 2 1 a , 1 2 1 b , 1 2 1 c , . . . と通信するための通信インタフェースである。

なお、図示しないが、ホストサーバ 2 0 0 、管理サーバ 3 0 0 および管理端末 4 0 0 も、プロセッサや主記憶装置、補助記憶装置などを有するコンピュータとして実現可能である。

20

【0035】

次に、ストレージシステムにおける階層制御について説明する。

図 3 は、階層化された記憶領域の設定例を示す図である。前述のように、ストレージ装置 1 0 0 には、ホストサーバ 2 0 0 からのアクセス対象となる論理ボリュームが設定される。また、ストレージ装置 1 0 0 には、論理ボリュームに対応する物理記憶領域の割り当て元となる T i e r プールが設定される。T i e r プールは、ドライブ部 1 2 0 に搭載された 1 以上の記憶装置によって実現される記憶領域である。そして、論理ボリュームを所定サイズごとに分割した単位領域ごとに、T i e r プールから記憶領域が割り当てられる。ただし、本実施の形態では、論理ボリューム上の単位領域のうち、ホストサーバ 2 0 0 からの要求に応じてデータが格納されている単位領域に対してのみ、T i e r プールから記憶領域が割り当てられる。

30

【0036】

T i e r プールは、互いにアクセス速度の異なる記憶装置によって実現される複数のサブプールに分割される。これにより、T i e r プールは、アクセス速度の異なる複数の記憶領域に階層化される。そして、単位領域には、ホストサーバ 2 0 0 からのアクセス頻度に応じて、いずれかのサブプールから記憶領域が割り当てられる。すなわち、アクセス頻度が高い単位領域には、アクセス速度の高い記憶装置によって実現されるサブプールが割り当てられ、アクセス頻度が低い単位領域には、アクセス速度の低い記憶装置によって実現されるサブプールが割り当てられる。

40

【0037】

これにより、ストレージ装置 1 0 0 は、アクセス頻度が高いデータほど、高速でアクセスできるようになる。したがって、ホストサーバ 2 0 0 からのアクセス頻度が高いデータほど、ホストサーバ 2 0 0 からのアクセス要求に対する応答速度が高くなる。その結果、全体としてホストサーバ 2 0 0 からのアクセス要求に対する応答性能が向上する。

【0038】

以下の説明では、アクセス速度が互いに異なる記憶装置として、アクセス速度の低い順にニアライン H D D 、オンライン H D D 、S S D が用いられるものとする。なお、ニアライン H D D によって実現されるサブプールの階層を「低階層」、オンライン H D D によ

50

て実現されるサブプールの階層を「中間階層」、SSDによって実現されるサブプールの階層を「高階層」と記載する場合がある。

【0039】

図3では例として、論理ボリュームLV1が設定され、論理ボリュームLV1に対応する物理記憶領域の割り当て元として、TierプールTP1が設定されている。また、TierプールTP1には、3階層のサブプールSP1-L, SP1-M, SP1-Hが設定されている。サブプールSP1-Lは、1台以上のニアラインHDDによって実現され、サブプールSP1-Mは、1台以上のオンラインHDDによって実現され、サブプールSP1-Hは、1台以上のSSDによって実現される。したがって、サブプールSP1-H, SP1-M, SP1-Lの順でアクセス速度が高い。

10

【0040】

さらに、ストレージ装置100には、単位領域のサイズが異なる複数のTierプールを設定可能になっている。そして、各論理ボリュームに対して、いずれか1つのTierプールから記憶領域を割り当てることができるようになっている。図3の例では、TierプールTP1に加えて、TierプールTP2, TP3が設定されている。また、論理ボリュームLV1に加えて、論理ボリュームLV2~4が設定されている。そして、論理ボリュームLV2には、TierプールTP2から記憶領域が割り当てられ、論理ボリュームLV3, LV4には、TierプールTP3から記憶領域が割り当てられる。

【0041】

また、TierプールTP2, TP3も、複数のサブプールに分割される。図3の例では、TierプールTP2には、1台以上のニアラインHDDによって実現されるサブプールSP2-Lと、1台以上のSSDによって実現されるSP2-Hが設定されている。また、TierプールTP3には、1台以上のニアラインHDDによって実現されるサブプールSP3-Lと、1台以上のオンラインHDDによって実現されるサブプールSP3-Mと、1台以上のSSDによって実点されるサブプールSP3-Hが設定されている。

20

【0042】

ここで、データ移動の管理単位となる単位領域のサイズを小さくして、より小さなサイズに分割したデータごとに配置先の階層を決定するほど、アクセス頻度に応じた適切な階層にデータが配置されやすくなる。その反面、管理対象となる単位領域の数が大きくなるので、管理効率が低下する。したがって、単位領域のサイズは、論理ボリュームに格納するデータの性質などに応じて、ユーザによって任意に設定されればよい。

30

【0043】

また、他の例として、Tierプールの容量が、そのTierプールから割り当て可能な単位領域の最大数によって決定されてもよい。この場合、単位領域のサイズを大きくすることで、Tierプールの容量を拡張することができる。この方法によれば、単位領域のサイズに関係なく、Tierプールごとに割り当て先として管理すべき単位領域の最大数が同一となる。このため、単位領域のサイズに関係なく、Tierプールの管理方法を共通化でき、管理を効率化できる。

【0044】

図4は、サブプールに対するRAIDグループの設定例を示す図である。サブプールには、1以上のRAIDグループが設定される。RAIDグループは、データが複数の記憶装置に冗長化されるように書き込みが制御される論理記憶領域である。サブプールに対してRAIDグループが設定されることで、そのサブプールの物理記憶領域が設定される。換言すると、サブプールの記憶領域は、RAIDグループに組み込まれた、RAIDによって制御される複数の記憶装置によって実現される。

40

【0045】

図4の例では、サブプールSP1-LにはRAIDグループRG1, RG2が設定され、サブプールSP1-MにはRAIDグループRG3, RG4が設定され、サブプールSP1-HにはRAIDグループRG5, RG6が設定されている。また、サブプールSP2-LにはRAIDグループRG11, RG12が設定され、サブプールSP2-Hには

50

R A I Dグループ R G 1 3 , R G 1 4 が設定されている。さらに、サブプール S P 3 - L には R A I Dグループ R G 2 1 ~ R G 2 3 が設定され、サブプール S P 3 - M には R A I Dグループ R G 2 4 ~ R G 2 6 が設定され、サブプール S P 3 - H には R A I Dグループ R G 2 7 ~ R G 2 9 が設定されている。

【 0 0 4 6 】

次に、ストレージ装置 1 0 0 の C M 1 1 0 および管理サーバ 3 0 0 で実行される処理の詳細について説明する。

図 5 は、C M および管理サーバが備える処理機能の構成例を示すブロック図である。

【 0 0 4 7 】

C M 1 1 0 は、記憶部 1 3 0、アクセス制御部 1 4 1、アクセス頻度送信部 1 4 2 および移動処理部 1 4 3 を有する。なお、記憶部 1 3 0 は、R A M 1 1 2 などの C M 1 1 0 が備える記憶装置の記憶領域として実現される。また、アクセス制御部 1 4 1、アクセス頻度送信部 1 4 2 および移動処理部 1 4 3 の処理は、例えば、プロセッサ 1 1 1 が所定のプログラムを実行することで実現される。

10

【 0 0 4 8 】

記憶部 1 3 0 には、論理ボリューム管理テーブル 1 3 1、単位領域管理テーブル 1 3 2、プール管理テーブル 1 3 3 および R A I Dグループ管理テーブル 1 3 4 が記憶される。

論理ボリューム管理テーブル 1 3 1 には、論理ボリュームに関する設定情報が登録される。例えば、論理ボリューム管理テーブル 1 3 1 には、論理ボリュームごとに、ボリューム名、容量、記憶領域が割り当てられる T i e r プールを示すプール名などが登録される。

20

【 0 0 4 9 】

単位領域管理テーブル 1 3 2 には、論理ボリュームに含まれる各単位領域に関する情報が登録される。例えば、単位領域管理テーブル 1 3 2 には、単位領域ごとに、T i e r プールから割り当てられた記憶領域を示すサブプール名、R A I Dグループ名およびオフセットや、単位期間におけるホストサーバ 2 0 0 からのアクセス回数などが登録される。

【 0 0 5 0 】

プール管理テーブル 1 3 3 には、T i e r プールに関する情報が登録される。例えば、プール管理テーブル 1 3 3 には、T i e r プールごとに、単位領域のサイズ、設定されたサブプールを示す情報、サブプールに設定された R A I Dグループを示す情報、空き領域を示す情報などが登録される。

30

【 0 0 5 1 】

R A I Dグループ管理テーブル 1 3 4 には、設定された R A I Dグループに関する情報が登録される。例えば、R A I Dグループ管理テーブル 1 3 4 には、R A I Dグループごとに、R A I Dレベル、使用される記憶装置を示す情報などが登録される。

【 0 0 5 2 】

アクセス制御部 1 4 1 は、ホストサーバ 2 0 0 からの論理ボリュームに対するアクセス要求に応じたアクセス制御を実行する。例えば、アクセス制御部 1 4 1 は、論理ボリューム内のいずれかの単位領域に格納されたデータの読み出し要求を受信すると、単位領域管理テーブル 1 3 2 を参照して、そのデータが格納されている R A I Dボリューム ( R A I Dグループに対応する論理記憶領域) 上の記憶領域を特定する。アクセス制御部 1 4 1 はさらに、R A I Dグループ管理テーブル 1 3 4 を参照して、特定された記憶領域に対応する物理記憶領域 (ドライブ部 1 2 0 内の記憶装置の記憶領域) を特定し、特定された物理記憶領域からデータを読み出してホストサーバ 2 0 0 に送信する。

40

【 0 0 5 3 】

また、アクセス制御部 1 4 1 は、論理ボリューム内のいずれかの単位領域に対する新規データの書き込み要求を受信すると、論理ボリューム管理テーブル 1 3 1 を参照して、この論理ボリュームに対応付けられた T i e r プールを特定する。アクセス制御部 1 4 1 は、プール管理テーブル 1 3 3 を参照して、特定された T i e r プールから空き領域を特定し、特定された空き領域にデータを書き込む。また、アクセス制御部 1 4 1 は、単位領域

50

管理テーブル 1 3 2 におけるこの単位領域に対応するレコードに、書き込み先の記憶領域を示す情報を登録する。なお、書き込み先とする空き領域は、例えば、特定された T i e r プール内のサブプールのうち、あらかじめ決められたサブプール（例えば高階層のサブプール）から特定される。

**【 0 0 5 4 】**

アクセス頻度送信部 1 4 2 は、一定時間間隔で単位領域ごとのアクセス頻度を収集し、収集されたアクセス頻度をアクセス頻度情報として管理サーバ 3 0 0 に送信する。アクセス頻度情報は、例えば、各単位領域におけるアクセス頻度の値がリスト化された情報として送信される。

**【 0 0 5 5 】**

移動処理部 1 4 3 は、管理サーバ 3 0 0 から移動指示情報を受信し、移動指示情報に基づいて、階層間（サブプール間）でのデータ移動（再配置）を実行する。移動指示情報は、例えば、移動対象の単位領域とその移動先階層とを示す情報がリスト化された情報として受信される。なお、後述するように、移動指示情報には、単位領域ごとに優先度が設定されている。

**【 0 0 5 6 】**

一方、管理サーバ 3 0 0 は、記憶部 3 1 0、アクセス頻度収集部 3 2 1 および配置先決定部 3 2 2 を有する。なお、記憶部 3 1 0 は、管理サーバ 3 0 0 が備える図示しない記憶装置の記憶領域として実現される。また、アクセス頻度収集部 3 2 1 および配置先決定部 3 2 2 の処理は、例えば、管理サーバ 3 0 0 が備える図示しないプロセッサが所定のプログラムを実行することで実現される。

**【 0 0 5 7 】**

記憶部 3 1 0 には、単位領域管理テーブル 3 1 1、プール管理テーブル 3 1 2 および優先度設定用テーブル 3 1 3 が記憶される。

単位領域管理テーブル 3 1 1 には、論理ボリュームに含まれる各単位領域に関する情報が登録される。例えば、単位領域管理テーブル 3 1 1 には、単位領域ごとに、T i e r プールから割り当てられた記憶領域を示すサブプール名や、単位期間におけるホストサーバ 2 0 0 からのアクセス回数などが登録される。なお、C M 1 1 0 側の単位領域管理テーブル 1 3 2 とは異なり、単位領域管理テーブル 3 1 1 では、R A I D グループ単位での記憶領域の位置は管理されない。

**【 0 0 5 8 】**

プール管理テーブル 3 1 2 には、T i e r プールに関する情報が登録される。例えば、プール管理テーブル 3 1 2 には、T i e r プールごとに、単位領域のサイズ、設定されたサブプールを示す情報、空き領域を示す情報などが登録される。ただし、C M 1 1 0 側のプール管理テーブル 1 3 3 とは異なり、プール管理テーブル 3 1 2 では、R A I D グループ単位での空き領域の管理は行われない。

**【 0 0 5 9 】**

優先度設定用テーブル 3 1 3 には、移動対象となった単位領域に対して優先度を設定するための条件が設定されている。優先度設定用テーブル 3 1 3 では、基本的に、移動先の階層とアクセス頻度との組み合わせに応じて優先度が設定されている。

**【 0 0 6 0 】**

アクセス頻度収集部 3 2 1 は、単位領域ごとのアクセス頻度を一定時間間隔で C M 1 1 0 から収集する。収集された情報は単位領域管理テーブル 3 1 1 に記録される。

配置先決定部 3 2 2 は、収集されたアクセス頻度に基づいて、単位領域のデータを配置すべき階層（サブプール）を決定する。そして、配置先決定部 3 2 2 は、決定された配置先が現在の配置先と異なる単位領域を移動対象として抽出し、移動対象の単位領域のデータを階層間で移動させるように指示する移動指示情報を、C M 1 1 0 に送信する。

**【 0 0 6 1 】**

配置先決定部 3 2 2 は、移動対象とされた複数の単位領域についての移動を、まとめて指示する。このとき、配置先決定部 3 2 2 は、複数の単位領域のそれぞれに対して、優先

10

20

30

40

50

度設定用テーブル 3 1 3 に基づいて優先度を設定する。送信される移動指示情報には、複数の単位領域のそれぞれに対して、移動先階層を示す情報と優先度が設定されている。

【 0 0 6 2 】

図 6 は、C M に記憶される単位領域管理テーブルのデータ構成例を示す図である。C M 1 1 0 に記憶される単位領域管理テーブル 1 3 2 は、単位領域ごとのレコードを含む。各レコードは、単位領域名、論理ボリューム名、L B A ( Logical Block Address )、サブプール名、R A I D グループ名、オフセット、アクセス頻度の各項目を含む。

【 0 0 6 3 】

単位領域名は、単位領域の識別情報を示す。論理ボリューム名は、単位領域が属する論理ボリュームの識別情報を示す。L B A は、その論理ボリュームにおける単位領域の先頭の論理アドレスを示す。サブプール名は、単位領域に割り当てられたサブプールの識別情報を示す。R A I D グループ名は、単位領域に割り当てられた記憶領域が属する R A I D グループの識別情報を示す。オフセットは、単位領域に割り当てられた記憶領域についての、R A I D グループに対応する R A I D ボリューム上の位置を示す。

10

【 0 0 6 4 】

アクセス頻度は、一定時間においてホストサーバ 2 0 0 から単位領域に対してアクセスされた回数を示す。

本実施の形態では、一例として、管理サーバ 3 0 0 のアクセス頻度収集部 3 2 1 は、アクセス頻度を収集する収集期間を一定の単位期間に分割し、単位期間ごとにアクセス頻度を C M 1 1 0 から収集するものとする。単位領域管理テーブル 1 3 2 のアクセス頻度の項目には、単位期間におけるアクセス回数が登録される。具体的には、アクセス頻度の値は、ホストサーバ 2 0 0 から単位領域に対する書き込みおよび単位領域からの読み出しが要求されるたびに、アクセス制御部 1 4 1 によってカウントアップされる。そして、単位期間に相当する一定時間が経過して、アクセス頻度収集部 3 2 1 からアクセス頻度情報の送信が要求されると、アクセス頻度送信部 1 4 2 は、アクセス頻度の項目の値をアクセス頻度情報に含めてアクセス頻度収集部 3 2 1 に送信する。これとともに、アクセス頻度送信部 1 4 2 は、アクセス頻度の項目に登録された値を 0 にリセットする。

20

【 0 0 6 5 】

このような処理により、C M 1 1 0 に記憶される単位領域管理テーブル 1 3 2 では、単位期間が終了するタイミングにおいて、アクセス頻度の項目にその単位期間におけるアクセス頻度が保持されるようになる。なお、上記処理に鑑み、本実施の形態では、「アクセス頻度」は「単位期間におけるアクセス回数」として定義されるものとする。

30

【 0 0 6 6 】

一方、図示しないが、管理サーバ 3 0 0 に記憶される単位領域管理テーブル 3 1 1 も、単位領域ごとのレコードを含む。ただし、各レコードは、単位領域名、論理ボリューム名、L B A、サブプール名、アクセス頻度の各項目を含むが、R A I D グループ名およびオフセットの各項目を含まない。単位領域名、論理ボリューム名、L B A、サブプール名の各項目には、C M 1 1 0 の単位領域管理テーブル 1 3 2 における対応する項目と同じ値が登録される。

【 0 0 6 7 】

また、アクセス頻度収集部 3 2 1 は、C M 1 1 0 からアクセス頻度を収集したとき、収集されたアクセス頻度の値が現在アクセス頻度の項目に登録されている値より大きい場合には、収集された値によって登録されている値を更新する。これにより、アクセス頻度の項目には、収集期間が終了したとき、その収集期間における単位期間ごとのアクセス頻度のうちの最大値が登録されるようになる。

40

【 0 0 6 8 】

図 7 は、C M に記憶されるプール管理テーブルのデータ構成例を示す図である。C M 1 1 0 に記憶されるプール管理テーブル 1 3 3 は、設定された T i e r プールごとのレコードを含む。各レコードは、単位領域サイズの項目を含むとともに、低階層、中間階層、高階層のそれぞれについてサブプール名、ディスク種別、R A I D グループ名の各項目を含

50

む。

【0069】

単位領域サイズは、Tierプールにおける単位領域のサイズを示す。サブプール名は、Tierプールに設定されたサブプールの識別情報を示す。ディスク種別は、サブプールを実現する記憶装置の種別を示す。RAIDグループ名は、サブプールに設定されたRAIDグループの識別情報を示す。

【0070】

また、図示しないが、管理サーバ300に記憶されるプール管理テーブル312も、設定されたTierプールごとのレコードを含む。ただし、各レコードは、単位領域サイズ、サブプール名の各項目を含むが、少なくともRAIDグループ名の項目を含まない。すなわち、管理サーバ300は、Tierプールの構成をサブプール単位（すなわち、階層単位）では認識するものの、RAIDグループ単位では認識しない。なお、単位領域サイズ、サブプールの各項目には、CM110のプール管理テーブル133における対応する項目と同じ値が登録される。

【0071】

なお、図示しないが、実際には、プール管理テーブル133, 312にはさらに、各サブプールのサイズを示す情報や、各サブプールにおける空き領域を管理するための情報が保持される。

【0072】

次に、優先度に基づく階層間のデータ移動制御について説明する。

前述のように、管理サーバ300の配置先決定部322は、移動対象の単位領域を抽出すると、CM110に対して、抽出された複数の単位領域についてまとめて移動を指示する。このとき、配置先決定部322は、各単位領域に対して優先度を設定する。CM110の移動処理部143は、基本的に、移動が指示された複数の単位領域について、優先度の高い順に階層間のデータ移動を実行する。

【0073】

優先度の決定方法としては、例えば、どの階層からどの階層へ移動するかを示す「移動経路条件」に基づいて決定する方法がある。例えば、中間階層から高階層への移動が指示された単位領域は、低階層から中間階層への移動が指示された単位領域より、アクセス頻度が高い。このため、中間階層から高階層への移動は、低階層から中間階層への移動より緊急性が高い。したがって、中間階層から高階層への移動に対してより高い優先度が設定される。また、例えば、2つの階層のうち上位階層への移動が指示された単位領域は、下位階層への移動が指示された単位領域より、アクセス頻度が高い。このため、上位階層への移動は、下位階層への移動より緊急性が高い。したがって、上位階層への移動に対してより高い優先度が設定される。

【0074】

また、移動経路条件が同一の単位領域については、さらに、アクセス頻度の高さを示す「アクセス頻度条件」に基づいて優先度が決定されてもよい。すなわち、アクセス頻度が高いほど移動の緊急性が高いので、アクセス頻度が高い単位領域に対してより高い優先度が設定される。

【0075】

しかしながら、実際には、上記の移動経路条件やアクセス頻度条件を用いて優先度を設定するだけでは、実際の状況に応じた最適な階層制御を実現できるとは限らない。例えば、上記各条件以外に、データ移動の緊急性に影響を与える条件がある可能性がある。また、データ移動が実行されるCM110側の処理状況や設定環境によっては、データ移動の緊急性に違いが生じることもあり得る。さらに、同じ階層に属する物理記憶領域の中でも、データを配置する位置によってはCM110の処理効率が低下することも考えられる。

【0076】

このような問題に鑑み、本実施の形態では、管理サーバ300側で、上記各条件に加えてさらに別の条件を用いた下記のような優先度決定制御が実行される。さらに、CM11

10

20

30

40

50

0側でも、下記のようなデータ移動制御(1)~(4)が実行される。

【0077】

<管理サーバでの優先度決定制御>

ある階層から上位階層へデータを移動させる際に、上位階層に空き領域がない場合には、CM110の移動処理部143は、まず上位階層のデータを下位階層へ移動させた後、上位階層へのデータ移動を実行する必要がある。以下、このように上位階層の空き領域を確保するために下位階層へデータを移動することを、「領域確保用の移動」と記載する。

【0078】

上位階層へデータを移動させる際に、下位階層への領域確保用の移動が必要になる場合には、上位階層へのデータ移動より下位階層への領域確保用の移動の方が、緊急性が高い。そこで、管理サーバ300の配置先決定部322は、前述の移動経路条件およびアクセス頻度条件に加えて、領域確保用の移動が否かという条件を用いて、優先度を決定する。

【0079】

図8は、優先度設定用テーブルのデータ構成例を示す図である。本実施の形態では、配置先決定部322は、優先度設定用テーブル313に基づいて優先度を決定する。なお、図8では、Tierプールが3階層である場合の例を示している。

【0080】

優先度設定用テーブル313では、移動内容に応じて大まかな移動種別C1~C7が設定され、移動種別C1~C7のそれぞれに対して異なる優先度が対応付けられている。これらのうち、移動種別C3, C5~C7は、移動経路条件によって区分された移動種別である。中間階層から高階層への移動を示す移動種別C3、低階層から中間階層への移動を示す移動種別C5、高階層から中間階層への移動を示す移動種別C6、中間階層から低階層への移動を示す移動種別C7の順に、高い優先度が設定されている。

【0081】

また、これらの移動種別C3, C5~C7のそれぞれについては、さらにアクセス頻度条件に応じて優先度が細分化されている。すなわち、移動種別C3, C5~C7のそれぞれについて、所定段階数のアクセス頻度範囲ごとに異なる優先度が設定されている。アクセス頻度が高いほど、高い優先度が設定される。

【0082】

さらに、移動種別C1, C2, C4は、領域確保用の移動を示す。移動種別C1は、中間階層から低階層への移動のうち、高階層の領域確保用の移動を示す。高階層の空き領域確保のために高階層から中間階層へデータ移動する際、中間階層に空き領域がない場合には、先に中間階層のデータを低階層に移動させて中間階層に空き領域を確保する必要がある。移動種別C1は、この場合の低階層へのデータ移動を示す。したがって、このようなデータ移動には、最も高い優先度が設定される。移動種別C2は、高階層から中間階層への移動のうち、高階層の領域確保用の移動を示す。この移動種別C2には、移動種別C1の次に高い優先度が設定される。

【0083】

移動種別C4は、中間階層から低階層への移動のうち、中間階層の領域確保用の移動を示す。このデータ移動は、移動種別C5に対応する、低階層から中間階層へのデータ移動のために実行されるものである。このため、移動種別C4には、移動種別C5より高い優先度が設定される。

【0084】

配置先決定部322は、上記のような優先度設定用テーブル313を参照することで、領域確保用の移動が否かという条件を加味して、優先度を適切に決定できる。

なお、Tierプールが2階層である場合には、高階層の領域確保用の移動に対しては、低階層から高階層への移動より高い優先度(最大の優先度を示す値)が設定される。

【0085】

<CMでのデータ移動制御(1)>

図4に例示したように、ある階層のサブプールに対して複数のRAIDグループが設定

10

20

30

40

50

される場合がある。これらの R A I D グループは、それぞれ個別の記憶装置によって実現される。このため、異なる R A I D グループを移動元または移動先としたデータ移動を並列に実行することで、データ移動の処理効率を向上させ、全体のデータ移動にかかる時間を短縮できる。

【 0 0 8 6 】

そこで、C M 1 1 0 の移動処理部 1 4 3 は、あるサブプールへのデータ移動を実行する場合に、そのサブプールに設定されている R A I D グループの中から、データ移動が実行中である R A I D グループ以外の R A I D グループを、データ移動先として選択する。これにより、データ移動を並列に実行できるようになり、その結果、全体のデータ移動にかかる時間を短縮できる。

【 0 0 8 7 】

< C M でのデータ移動制御 ( 2 ) >

移動処理部 1 4 3 は、あるサブプールへの新たなデータ移動を実行しようとしたときに、そのサブプールに設定されているすべての R A I D グループにおいてデータ移動が実行中である場合には、実行中のいずれかのデータ移動の完了を待つ必要がある。いずれかのデータ移動が完了するまで、他のすべての新たなデータ移動を実行待ちにしてしまうと、その待ち時間が無駄になり、処理効率が悪い。

【 0 0 8 8 】

そこで、移動処理部 1 4 3 は、上記のケースでは、実行しようとした新たなデータ移動については一旦実行待ち状態として、次に優先度の高い新たなデータ移動を実行する。これにより、管理サーバ 3 0 0 から移動対象として指示された単位領域のデータ移動を、効率よく実行することができる。すなわち、実行待ちによって新たなデータ移動が実行されない無駄な時間が発生する確率を低減でき、その結果、全体のデータ移動にかかる時間を短縮できる。

【 0 0 8 9 】

< C M でのデータ移動制御 ( 3 ) >

上位階層へのデータ移動は、単位領域のアクセス頻度が上昇した場合に実行される。しかし、管理サーバ 3 0 0 による移動先階層の決定は、その直前におけるアクセス頻度に基づいて行われる。このため、上位階層へ実際にデータを移動する際に、アクセス頻度の状況が変化している可能性がある。

【 0 0 9 0 】

そこで、移動処理部 1 4 3 は、上位階層へのデータ移動を実行しようとする際に、同じ優先度が設定された他の単位領域の中に、現在のアクセス頻度が非常に高く、移動の緊急性が高いと判断される単位領域が存在する場合には、その単位領域のデータ移動を優先的に実行する。

【 0 0 9 1 】

例えば、移動処理部 1 4 3 は、データ移動を実行しようとしている単位領域についての現在のアクセス頻度が所定値未満である場合、同じ優先度が設定された他の単位領域についてのアクセス頻度を確認する。移動処理部 1 4 3 は、他の単位領域の中にアクセス頻度が所定値以上のものがある場合、その単位領域のデータ移動を先に実行する。また、別の処理例としては、移動処理部 1 4 3 は、同じ優先度が設定された他の単位領域の中に、現在データ移動を実行しようとしている単位領域より現在のアクセス頻度が高いものがある場合に、その単位領域のデータ移動を先に実行してもよい。

【 0 0 9 2 】

このような処理により、アクセス頻度の変動に応じて、データ移動の実行順を適正化することができる。すなわち、アクセス頻度の高い単位領域に対してアクセス速度の高い記憶領域を早期に割り当てることができ、その結果、ホストサーバ 2 0 0 からのアクセス要求に対する応答速度を向上させることができる。

【 0 0 9 3 】

< C M でのデータ移動制御 ( 4 ) >

10

20

30

40

50

前述のように、本実施の形態では、単位領域のサイズが可変となっている。一方、CM110は、単位領域ごとのアクセス頻度を、一定の単位期間におけるアクセス回数として計測している。しかし、この方法では、単位期間でのアクセス回数が同一である場合、単位領域のサイズが小さいほど、容量当たりのアクセス頻度（IOPS：Input Output operations Per Second）は高いことになる。このため、単位期間でのアクセス回数が同一である場合、サイズが小さい単位領域の方が、データ移動の緊急性が高い。

【0094】

そこで、移動処理部143は、同一の優先度が設定された複数の単位領域のうち、サイズが小さい単位領域のデータ移動を優先的に実行する。これにより、データ移動の実行順を、アクセス頻度が正確に反映されるように適正化することができる。すなわち、アクセス頻度の高い単位領域に対してアクセス速度の高い記憶領域を早期に割り当てることができ、その結果、ホストサーバ200からのアクセス要求に対する応答速度を向上させることができる。

10

【0095】

次に、管理サーバ300およびCM110の処理の詳細について、シーケンス図およびフローチャートを用いて説明する。

まず、図9は、アクセス頻度の収集処理手順を示すシーケンス図の例である。ストレージシステムにおいては、図9の処理が繰り返し実行される。

【0096】

図9の実行直前の初期状態では、CM110の単位領域管理テーブル132の各レコードでは、アクセス頻度の値が0に初期化されている。また、管理サーバ300の単位領域管理テーブル311の各レコードでは、アクセス頻度の項目に値が登録されていない。

20

【0097】

この状態から、管理サーバ300のアクセス頻度収集部321は、一定時間が経過するまで、すなわち、単位期間が終了するまで待機する（ステップS11）。一定時間が経過するまでの間、CM110のアクセス制御部141は、ホストサーバ200からの要求に応じたアクセス制御を実行する。単位領域に対するアクセスが要求されるたびに、単位領域管理テーブル132におけるその単位領域に対応するアクセス頻度の値がカウントアップされる。

【0098】

一定時間が経過すると、アクセス頻度収集部321は、CM110に対してアクセス頻度の送信を要求する（ステップS12）。CM110のアクセス頻度送信部142は、単位領域管理テーブル132の各レコードからアクセス頻度を取得し、各単位領域のアクセス頻度をリスト化したアクセス頻度情報を作成する（ステップS13）。アクセス頻度送信部142は、作成されたアクセス頻度情報を管理サーバ300に送信する（ステップS14）。これとともに、アクセス頻度送信部142は、単位領域管理テーブル132の各レコードのアクセス頻度の値を0にリセットする（ステップS15）。

30

【0099】

管理サーバ300のアクセス頻度収集部321は、アクセス頻度情報を受信し、アクセス頻度情報に含まれるアクセス頻度を単位領域管理テーブル311に保存する（ステップS16）。このとき、アクセス頻度収集部321は、アクセス頻度情報に含まれるある単位領域のアクセス頻度と、単位領域管理テーブル311に登録された、その単位領域に対応するアクセス頻度とを比較する。アクセス頻度収集部321は、前者の値の方が大きい場合、前者の値によって後者の値を更新する。これにより、単位領域管理テーブル311のアクセス頻度の項目には、アクセス頻度の収集期間における単位期間ごとのアクセス頻度のうちの最大値が登録される。

40

【0100】

アクセス頻度収集部321は、ステップS12、S16の処理を所定回数実行したかを判定する（ステップS17）。アクセス頻度収集部321は、これらの処理を所定回数実行していない場合には、一定時間が経過するまで待機し（ステップS11）、一定時間が

50

経過するとステップ S 1 2 の処理を実行する。また、アクセス頻度収集部 3 2 1 は、ステップ S 1 2 , S 1 6 の処理を所定回数実行済みの場合、すなわち、アクセス頻度の収集期間が終了した場合には、各单位領域のデータの配置先決定処理を開始する（ステップ S 1 8 ）。

#### 【 0 1 0 1 】

図 1 0 は、データの配置先決定処理手順を示すフローチャートの例である。

[ ステップ S 2 1 ] 管理サーバ 3 0 0 の配置先決定部 3 2 2 は、各单位領域のデータの配置先階層を決定する。例えば、配置先決定部 3 2 2 は、単位領域をデータの配置先となる T i e r プールごとに分類し、分類された単位領域のグループのそれぞれについて以下の処理を実行する。

10

#### 【 0 1 0 2 】

配置先決定部 3 2 2 は、まず、単位領域管理テーブル 3 1 1 から各单位領域のアクセス頻度を取得し、取得されたアクセス頻度に基づいて、各单位領域のデータの配置先階層を決定する。例えば、T i e r プールの階層数が 2 層の場合、閾値 T H が設定される。そして、アクセス頻度が閾値 T H 未満である単位領域のデータを低階層に配置し、アクセス頻度が閾値 T H 以上である単位領域のデータを高階層に配置するように、配置先階層が決定される。また、T i e r プールの階層数が 3 層の場合、2 つの閾値 T H 1 , T H 2 (ただし、 $T H 1 < T H 2$ ) が設定される。そして、アクセス頻度が閾値 T H 1 未満である単位領域のデータを低階層に配置し、アクセス頻度が閾値 T H 1 以上閾値 T H 2 未満である単位領域のデータを中間階層に配置し、アクセス頻度が閾値 T H 2 以上である単位領域のデータを高階層に配置するように、配置先階層が決定される。

20

#### 【 0 1 0 3 】

また、閾値を用いた上記の決定方法では、ある階層が配置先に決定された単位領域の合計データサイズが、その階層のサイズを超えてしまう場合がある。この場合、配置先決定部 3 2 2 は、その階層が配置先として決定された単位領域の中から、合計データサイズがその階層のサイズ内に収まる範囲で、アクセス頻度が高い順に単位領域を選択し、選択された各单位領域の配置先をその階層に確定する。また、配置先決定部 3 2 2 は、残りの単位領域の配置先を 1 つ下位の階層に変更する。階層が 3 層以上の場合、配置先決定部 3 2 2 は、上位階層から上記処理を順次実行する。

#### 【 0 1 0 4 】

[ ステップ S 2 2 ] 配置先決定部 3 2 2 は、階層間でデータの移動が必要となる移動対象の単位領域を決定する。具体的には、配置先決定部 3 2 2 は、現在配置されている階層と、ステップ S 2 1 で決定された配置先階層とが異なる単位領域を、移動対象に決定する。また、配置先決定部 3 2 2 は、移動対象に決定された単位領域について、ステップ S 2 1 で決定された配置先階層を、移動先階層に決定する。

30

#### 【 0 1 0 5 】

さらに、配置先決定部 3 2 2 は、移動対象の単位領域の中から、下位階層への移動が行われる単位領域を抽出し、抽出された単位領域の中からさらに、上位階層の領域確保用の移動が行われる単位領域を特定する。このような単位領域としては、移動先の階層が、ステップ S 2 1 で決定された配置先階層より低い単位領域が特定される。具体的には、例えば、次のようにして特定される。

40

#### 【 0 1 0 6 】

T i e r プールの階層数が 2 層の場合、高階層から低階層への移動が行われる単位領域のうち、アクセス頻度が前述の閾値 T H 以上である単位領域が、領域確保用の移動が行われる単位領域として特定される。また、T i e r プールの階層数が 3 層の場合、高階層から中間階層への移動が行われる単位領域のうち、アクセス頻度が前述の閾値 T H 2 以上である単位領域が、領域確保用の移動が行われる単位領域として特定される。さらに、中間階層から低階層への移動が行われる単位領域のうち、アクセス頻度が前述の閾値 T H 1 以上である単位領域が、領域確保用の移動が行われる単位領域として特定される。

#### 【 0 1 0 7 】

50

[ステップS 2 3] 配置先決定部 3 2 2 は、優先度設定用テーブル 3 1 3 を参照して、移動対象の単位領域のそれぞれについて、どの階層からどの階層へ移動するかを示す移動経路条件に基づいて優先度を仮決定する。例えば、Tier プールが 3 階層である場合、配置先決定部 3 2 2 は、各単位領域を移動経路条件に応じて移動種別 C 3, C 5 ~ C 7 (図 8 参照) のいずれかに分類する。配置先決定部 3 2 2 は、優先度設定用テーブル 3 1 3 において、分類された移動種別に対して設定されたいずれかの優先度(例えば、最も高い優先度を示す値)を、単位領域に対する優先度として仮決定する。

【0108】

[ステップS 2 4] 配置先決定部 3 2 2 は、移動対象の単位領域を 1 つ選択する。

[ステップS 2 5] 配置先決定部 3 2 2 は、選択された単位領域についてのデータ移動が、下位階層への移動であるかを判定する。配置先決定部 3 2 2 は、下位階層への移動である場合、ステップS 2 6 の処理を実行し、下位階層への移動でない場合、ステップS 2 8 の処理を実行する。

10

【0109】

[ステップS 2 6] 配置先決定部 3 2 2 は、選択された単位領域についてのデータ移動が、上位階層の領域確保用の移動であるかを判定する。配置先決定部 3 2 2 は、領域確保用の移動である場合、ステップS 2 7 の処理を実行し、領域確保用の移動でない場合、ステップS 2 8 の処理を実行する。

【0110】

[ステップS 2 7] 配置先決定部 3 2 2 は、優先度設定用テーブル 3 1 3 に基づいて、選択された単位領域に対応する優先度の値を、より高い優先度を示すように変更する。例えば、Tier プールが 3 階層である場合、中間階層から低階層への移動が行われる単位領域の優先度は、「1」に変更される。高階層から中間階層への移動が行われる単位領域の優先度は、「2」に変更される。中間階層から低階層への移動が行われる単位領域の優先度は、「7」に変更される。

20

【0111】

[ステップS 2 8] 配置先決定部 3 2 2 は、優先度設定用テーブル 3 1 3 を参照し、単位領域のアクセス頻度に基づいて優先度を最終決定する。具体的には、配置先決定部 3 2 2 は、優先度設定用テーブル 3 1 3 において、ステップS 2 3 で分類された移動種別に対して設定されたアクセス頻度範囲と、単位領域のアクセス頻度とを比較することによって、単位領域に対応する優先度を決定する。

30

【0112】

以上のステップS 2 7, S 2 8 のいずれかの処理により、ステップS 2 4 で選択された移動対象の単位領域に対応する優先度が、正式に決定される。

[ステップS 2 9] 配置先決定部 3 2 2 は、移動対象の単位領域のすべてを選択済みであるかを判定する。未選択の単位領域がある場合、配置先決定部 3 2 2 は、処理をステップS 2 4 に進め、未選択の単位領域を 1 つ選択する。一方、単位領域をすべて選択済みの場合、配置先決定部 3 2 2 は、ステップS 3 0 の処理を実行する。

【0113】

[ステップS 3 0] 配置先決定部 3 2 2 は、移動指示情報を作成する。移動指示情報には、移動対象の単位領域のそれぞれについて、単位領域を示す識別情報と、データ移動先の階層またはサブプールを示す識別情報と、決定された優先度とが含まれる。配置先決定部 3 2 2 は、作成された移動指示情報を CM 1 1 0 に送信する。

40

【0114】

[ステップS 3 1] 配置先決定部 3 2 2 は、移動対象の単位領域のデータ移動が CM 1 1 0 で実行されるたびに、移動結果情報を受信する。移動結果情報には、単位領域の識別情報と、データの移動先の階層またはサブプールを示す識別情報とが含まれている。配置先決定部 3 2 2 は、移動結果情報に基づいて、単位領域管理テーブル 3 1 1 のレコードのうち、データ移動が実行された単位領域に対応するレコードを更新する。

【0115】

50

図 1 1、図 1 2 は、データ移動処理手順を示すフローチャートの例である。C M 1 1 0 の移動処理部 1 4 3 が、図 1 0 のステップ S 3 0 で送信された移動指示情報を受信すると、図 1 1 の処理が開始される。

【 0 1 1 6 】

[ ステップ S 4 1 ] 移動処理部 1 4 3 は、受信された移動指示情報に含まれる移動対象の単位領域を、優先度の高い順にソートした移動対象リストを作成する。

[ ステップ S 4 2 ] 移動処理部 1 4 3 は、移動対象リストから、先頭の単位領域を移動候補として選択する。

【 0 1 1 7 】

[ ステップ S 4 3 ] 移動処理部 1 4 3 は、選択された単位領域についてのデータ移動が、上位階層への移動であるかを判定する。移動処理部 1 4 3 は、上位階層への移動である場合、ステップ S 4 4 の処理を実行し、上位階層への移動でない場合、ステップ S 4 7 の処理を実行する。

10

【 0 1 1 8 】

[ ステップ S 4 4 ] 移動処理部 1 4 3 は、単位領域管理テーブル 1 3 2 を参照し、選択された単位領域に対応するレコードに現在登録されているアクセス頻度を取得する。移動処理部 1 4 3 は、取得されたアクセス頻度が所定値 T H 3 以上かを判定し、所定値 T H 3 以上の場合はステップ S 4 7 の処理を実行し、所定値 T H 3 未満の場合はステップ S 4 5 の処理を実行する。

【 0 1 1 9 】

[ ステップ S 4 5 ] 移動処理部 1 4 3 は、移動対象リストから、移動候補として選択された単位領域と同一の優先度が設定された他の単位領域を特定する。移動処理部 1 4 3 は、単位領域管理テーブル 1 3 2 を参照し、特定された他の単位領域に対応するレコードに現在登録されているアクセス頻度を取得する。移動処理部 1 4 3 は、特定された他の単位領域の中に、アクセス頻度が所定値 T H 3 以上のものがあるかを判定する。移動処理部 1 4 3 は、アクセス頻度が所定値 T H 3 以上である他の単位領域がある場合、ステップ S 4 6 の処理を実行し、アクセス頻度が所定値 T H 3 以上である他の単位領域がない場合、ステップ S 4 7 の処理を実行する。

20

【 0 1 2 0 】

[ ステップ S 4 6 ] 移動処理部 1 4 3 は、移動候補を、ステップ S 4 5 でアクセス頻度が所定値 T H 3 以上と判定された他の単位領域に変更する。該当する他の単位領域が複数存在する場合には、例えば、アクセス頻度が最大の単位領域が選択される。なお、変更前の単位領域は、移動対象リストにそのまま残される。

30

【 0 1 2 1 】

[ ステップ S 4 7 ] 移動処理部 1 4 3 は、移動対象リストから、移動候補として選択された単位領域と同一の優先度が設定された他の単位領域を特定する。移動処理部 1 4 3 は、特定された他の単位領域の中に、移動候補として選択された単位領域よりサイズが小さいものがあるかを判定する。移動処理部 1 4 3 は、サイズが小さい他の単位領域がある場合、ステップ S 4 8 の処理を実行し、サイズが小さい他の単位領域がない場合、図 1 2 のステップ S 5 1 の処理を実行する。

40

【 0 1 2 2 】

[ ステップ S 4 8 ] 移動処理部 1 4 3 は、移動候補を、ステップ S 4 7 でサイズが小さいと判定された他の単位領域に変更する。該当する他の単位領域が複数存在する場合には、例えば、サイズが最小の単位領域が選択される。なお、変更前の単位領域は、移動対象リストにそのまま残される。この後、図 1 2 のステップ S 5 1 の処理が実行される。

【 0 1 2 3 】

以下、図 1 2 を用いて説明を続ける。

[ ステップ S 5 1 ] 移動処理部 1 4 3 は、移動候補として選択された単位領域についてのデータ移動先のサブプールを判別する。移動処理部 1 4 3 は、プール管理テーブル 1 3 3 を参照し、判別されたサブプールに設定された R A I D グループを特定する。移動処理

50

部 1 4 3 は、特定された R A I D グループの中から、他の単位領域についての階層間のデータ移動が現在実行されていない R A I D グループを抽出する。

【 0 1 2 4 】

[ ステップ S 5 2 ] 移動処理部 1 4 3 は、ステップ S 5 1 においてデータ移動が実行されていない R A I D グループが抽出された場合、ステップ S 5 3 の処理を実行し、抽出されなかった場合、ステップ S 5 6 の処理を実行する。

【 0 1 2 5 】

[ ステップ S 5 3 ] 移動処理部 1 4 3 は、抽出された R A I D グループの中から、単位領域に対して記憶領域が割り当てられた容量が最も少ない R A I D グループを選択し、選択された R A I D グループをデータの移動先に決定する。

10

【 0 1 2 6 】

[ ステップ S 5 4 ] 移動処理部 1 4 3 は、移動候補として選択された単位領域のデータを、ステップ S 5 3 で決定された移動先の R A I D グループに移動させる。

[ ステップ S 5 5 ] 移動処理部 1 4 3 は、ステップ S 5 4 でのデータ移動内容を示す移動結果情報を、管理サーバ 3 0 0 に送信する。また、移動処理部 1 4 3 は、データ移動が完了した単位領域を、移動対象リストから削除する。

【 0 1 2 7 】

[ ステップ S 5 6 ] 移動処理部 1 4 3 は、移動対象として選択された単位領域を、移動対象リストの末尾に移動させる。これにより、この単位領域は移動待ち状態になる。

[ ステップ S 5 7 ] 移動処理部 1 4 3 は、移動対象リストにデータ移動が実行されていない単位領域が残っているかを判定する。単位領域が残っている場合、移動処理部 1 4 3 はステップ S 4 2 に処理を進め、次の単位領域を移動候補として選択する。一方、単位領域が残っていない場合、移動処理部 1 4 3 は処理を終了する。

20

【 0 1 2 8 】

以上の図 1 1、図 1 2 の処理では、移動先のサブプールに設定されている R A I D グループの中にデータ移動が実行されていない R A I D グループがある場合には、ステップ S 5 4 においてその R A I D グループに対してデータが移動される。これにより、前述のデータ移動制御 ( 1 ) が実現される。

【 0 1 2 9 】

また、移動先のサブプールに設定されているすべての R A I D グループにおいてデータ移動が実行中の場合には、ステップ S 5 6 において、移動候補として選択された単位領域のデータ移動の実行が後回しにされる。そして、その後再度ステップ S 4 2 の処理が実行されて、次の単位領域が移動候補として選択される。これにより、前述のデータ移動制御 ( 2 ) が実現される。

30

【 0 1 3 0 】

また、同一の優先度が設定された他の単位領域の中に、現在のアクセス頻度が非常に高い単位領域が存在する場合には、ステップ S 4 6 において、移動候補がその単位領域に変更されて、その単位領域のデータ移動が優先的に実行されるようになる。これにより、前述のデータ移動制御 ( 3 ) が実現される。

【 0 1 3 1 】

また、同一の優先度が設定された他の単位領域の中に、移動候補の単位領域よりサイズが小さい単位領域が存在する場合には、ステップ S 4 8 において、移動候補がその単位領域に変更されて、その単位領域のデータ移動が優先的に実行されるようになる。これにより、前述のデータ移動制御 ( 4 ) が実現される。

40

【 0 1 3 2 】

なお、上記の各実施の形態に示した装置 ( データ処理装置 1、C M 1 1 0、ホストサーバ 2 0 0、管理サーバ 3 0 0 ) 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、各装置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、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

50

録しておくことができる。コンピュ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、CD（Compact Disc）、DVD（Digital Versatile Disc）、ブルーレイディスク（BD）などがある。光磁気記録媒体には、MO（Magneto-Optical disk）などがある。

【0133】

プログラムを流通させる場合には、例えば、そのプログラムが記録されたDVD、CD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

【0134】

プログラムを実行するコンピュータは、例えば、可搬型記録媒体に記録されたプログラムまた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、プログラムにしたが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、そのプログラムにしたが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したがった処理を実行することもできる。

20

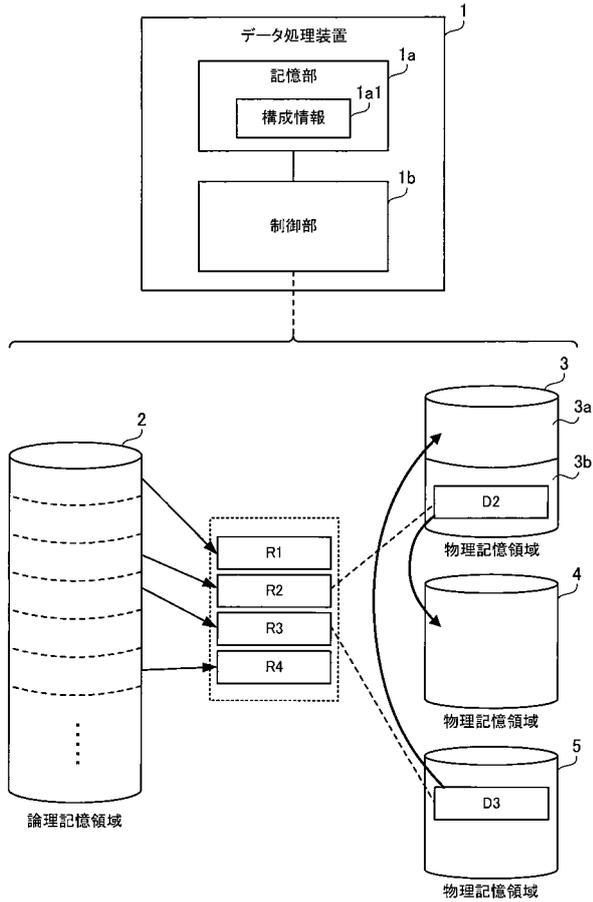
【符号の説明】

【0135】

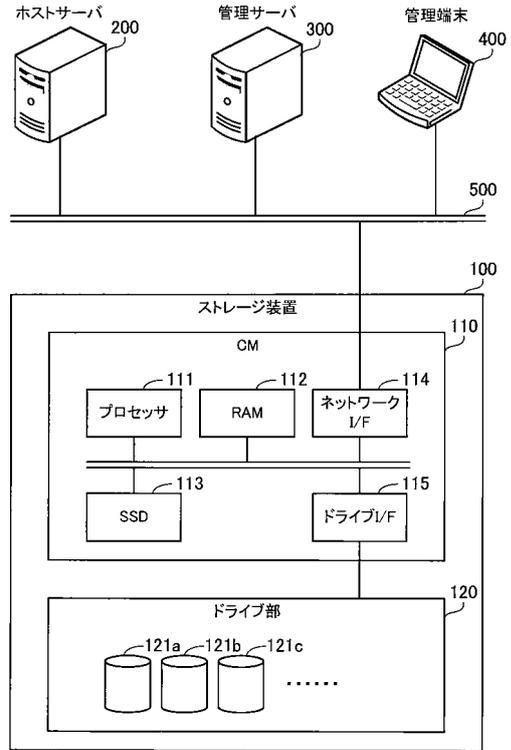
- 1 データ処理装置
- 1 a 記憶部
- 1 a 1 構成情報
- 1 b 制御部
- 2 論理記憶領域
- 3 ~ 5 物理記憶領域
- 3 a , 3 b 分割領域
- D 2 , D 3 データ
- R 1 ~ R 4 単位領域

30

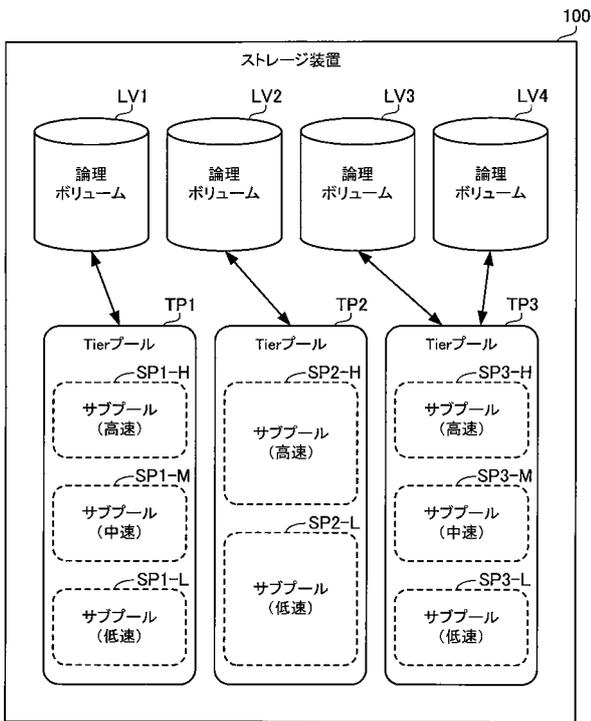
【図1】



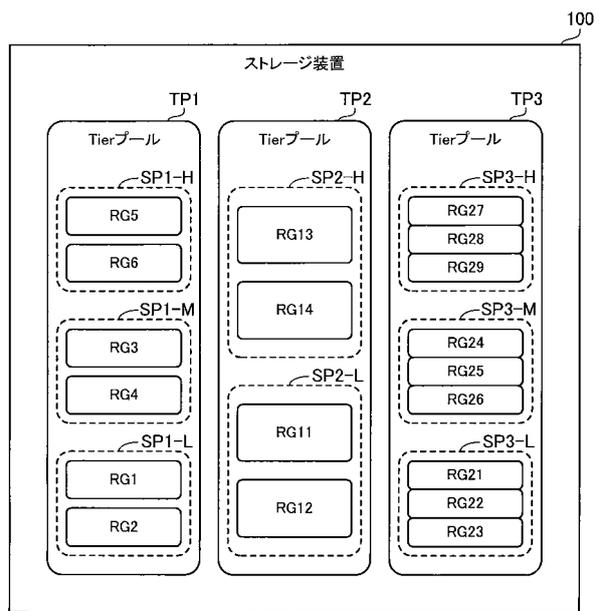
【図2】



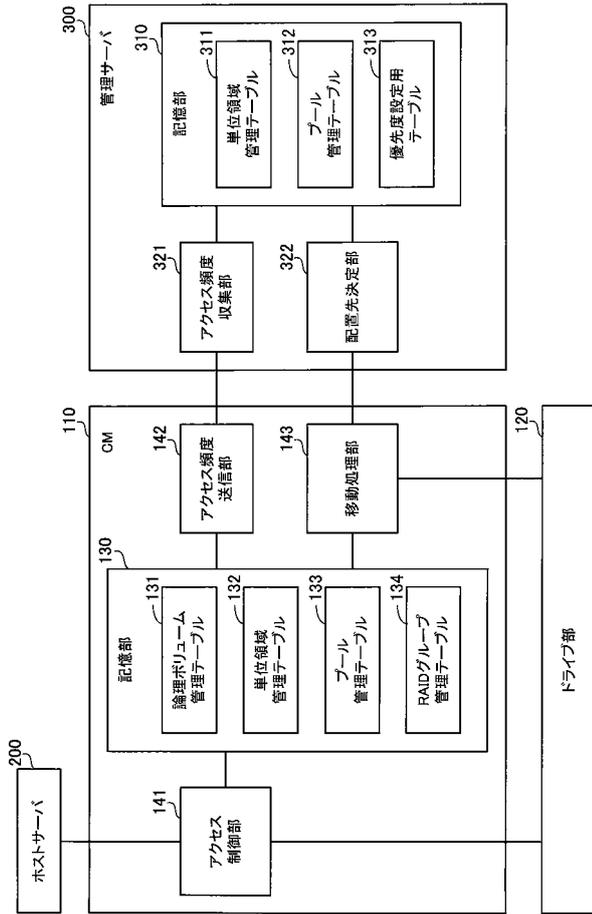
【図3】



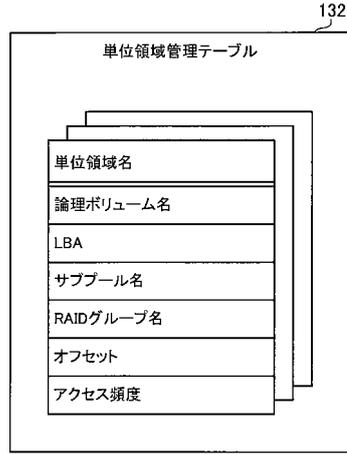
【図4】



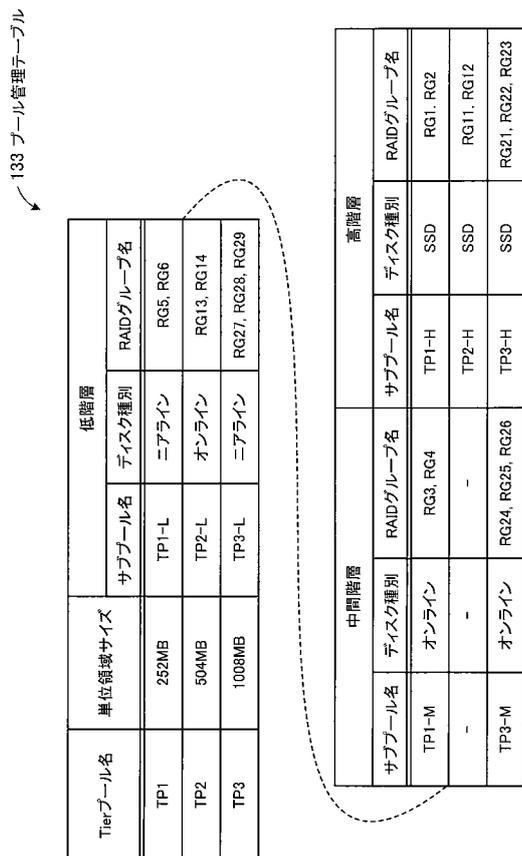
【 図 5 】



【 図 6 】



【 図 7 】

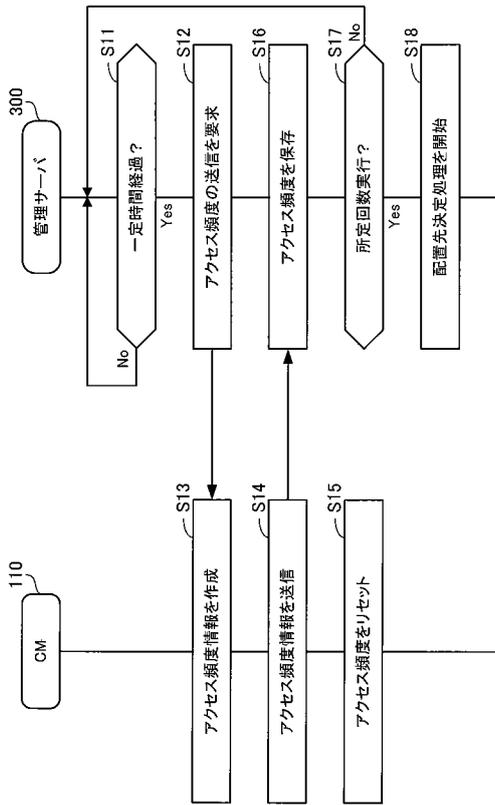


【 図 8 】

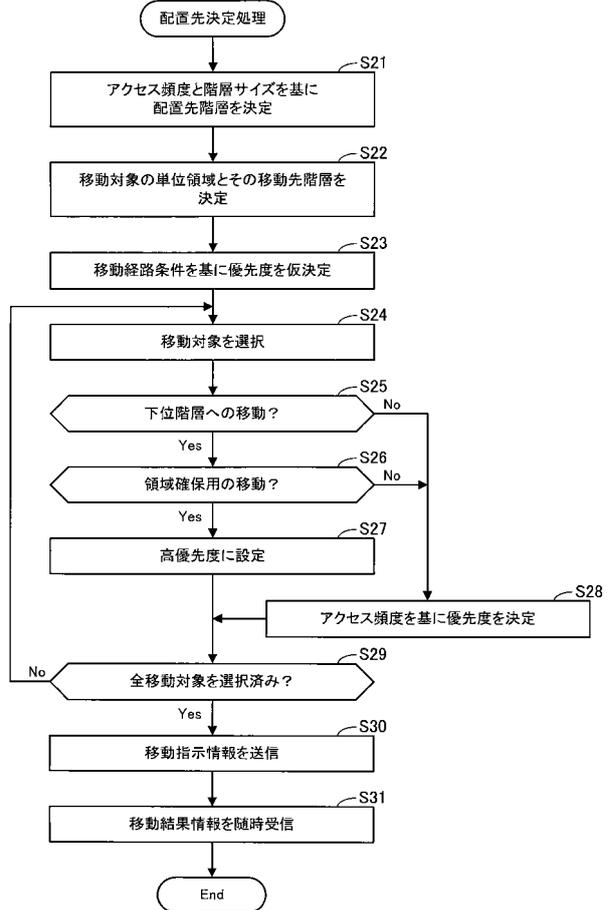
313 優先度設定用テーブル (Priority Setting Table 313)

移動種別	移動内容	アクセス頻度範囲	優先度
C1	低階層へ下げる (高階層の領域確保用)	-	1
C2	中間階層へ下げる (高階層の領域確保用)	-	2
C3	高階層へ上げる	xx以上	3
		xx~xx	4
		...	5
		xx以下	6
C4	低階層へ下げる (中間階層の領域確保用)	-	7
C5	中間階層へ上げる	xx以上	8
		xx~xx	9
		...	10
		xx以下	11
C6	中間階層へ下げる	xx以上	12
		xx~xx	13
		...	14
		xx以下	15
C7	低階層へ下げる	xx以上	16
		xx~xx	17
		...	18
		xx以下	19

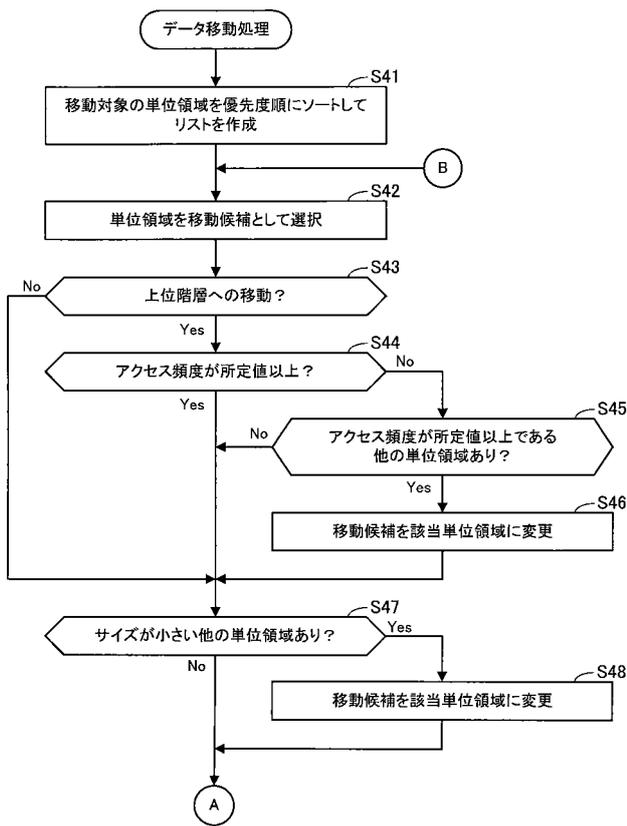
【図9】



【図10】



【図11】



【図12】

