

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4824458号  
(P4824458)

(45) 発行日 平成23年11月30日(2011.11.30)

(24) 登録日 平成23年9月16日(2011.9.16)

(51) Int.Cl. F I  
**G 0 6 F 3/06 (2006.01)** G O 6 F 3/06  
**G 1 1 B 19/00 (2006.01)** G 1 1 B 19/00 I O O H

請求項の数 18 (全 26 頁)

<p>(21) 出願番号 特願2006-118870 (P2006-118870)</p> <p>(22) 出願日 平成18年4月24日 (2006.4.24)</p> <p>(65) 公開番号 特開2007-293479 (P2007-293479A)</p> <p>(43) 公開日 平成19年11月8日 (2007.11.8)</p> <p>審査請求日 平成21年1月8日 (2009.1.8)</p> <p>(出願人による申告) 国等の委託研究の成果に係る特許出願 (平成17年度、文部科学省、科学技術試験研究委託費 [1] ログ転送によるリモートサーバレス方式のDBヘッド化技術の実現方式検討、2] クエリプラン利用先読み技術の詳細評価、3] ストレージリソースとI/O性能の見積・予測技術の基本仕様検討における実現方式検討、4] データ配置最適化技術の評価、5] ボトルネック検出技術の基本仕様検討における実現方式検討) 産業活力再生特別措置法第30条の適用を受けるもの)</p>	<p>(73) 特許権者 000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号</p> <p>(74) 代理人 110000279 特許業務法人ウィルフォート国際特許事務所</p> <p>(72) 発明者 出射 英臣 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所 システム開発研究所内</p> <p>(72) 発明者 西川 記史 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所 システム開発研究所内</p>
--	---

最終頁に続く

(54) 【発明の名称】 ストレージシステムの消費電力を低減する計算機システム及び方法

(57) 【特許請求の範囲】

【請求項1】

記憶メディアとしてのディスクを備えたディスク装置を複数台と、前記複数のディスク装置へのアクセスを制御するディスクアクセス制御部とを備え、前記複数のディスク装置にデータベース(DB)を記憶したストレージシステムと、

前記DBを管理するためのDB管理情報を記憶したDB管理情報記憶域を備え、前記DBを操作するデータベース管理システム(DBMS)を実行する計算機と、

前記複数のディスク装置の回転を制御し、アクセスが発生しないディスク装置のディスクを、0以上であって、アクセス時の回転速度である第一の速度よりも遅い第二の速度で回転させるようにさせるようになっているディスク回転制御部と

を備え、

前記DBは、前記DBMSにアクセスされる要素であるDBスキーマを複数個含んでおり、

前記DB管理情報は、どのDBスキーマがどの位置に記憶されているかを表す情報を含んでおり、

前記DBMSは、クエリを受け付け、前記受け付けたクエリのクエリプランを作成し、該クエリプランに従って、前記DB管理情報から特定されるDBスキーマに対するアクセス要求を発行するようになっており、

前記ディスク回転制御部が、前記DBMSがクエリを受け付けた場合に、該クエリのクエリプランと前記DB管理情報とから、該クエリの処理の際にアクセスされるDBスキーマ

マの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを前記第一の速度で回転させ、

前記DBMSは、前記クエリの処理が終了すると、処理が終了した前記クエリの情報を、終了要求として前記ディスク回転制御部に送り、

前記ディスク回転制御部は、前記DBMSからの前記終了要求を受信した場合に、前記クエリの処理の際にアクセスされた全てのディスク装置のディスクを停止させる、

計算機システム。

【請求項2】

前記ディスク回転制御部が、前記DBMSがクエリを受け付けた場合に、該クエリのクエリプランと前記DB管理情報とから、該クエリの処理の際にアクセスされる全てのDBスキーマにそれぞれ対応した全ての記憶位置を特定し、該特定された全ての記憶位置をそれぞれ備える全てのディスク装置のディスクを前記第一の速度で回転させる、

請求項1記載の計算機システム。

【請求項3】

前記ディスク回転制御部が、前記DBMSから発行されるアクセス要求を基に、該クエリの処理進行状況を把握し、該把握された処理進行状況から、次段階でアクセスされるDBスキーマの記憶位置を特定し、該記憶位置を備えたディスク装置のディスクを前記第一の速度で回転させる、

請求項1記載の計算機システム。

【請求項4】

前記DB管理情報には、各DBスキーマ毎にアクセスサイズが含まれており、  
前記ディスク回転制御部は、前記DBMSから発行されるアクセス要求を用いて、該アクセス要求に従ってアクセスされるDBスキーマに対応したアクセスサイズを前記DB管理情報から特定し、特定されたアクセスサイズと、前記把握された処理進行状況とに基づいて、前記次段階でアクセスされるDBスキーマを記憶したディスク装置のディスクを前記第一の速度で回転させる開始タイミングを制御する、

請求項3記載の計算機システム。

【請求項5】

前記ディスク回転制御部が、前記次段階でアクセスされるDBスキーマを記憶したディスク装置について、単位時間当たりのデータ転送量と、ディスクの回転速度が前記第一の速度になるのに要するディスク回転開始時間長とを取得し、前記アクセス要求と、現段階でアクセスされるDBスキーマに対応したアクセスサイズとを基に、現段階の進捗を特定し、該特定された進捗と、前記取得した単位時間当たりのデータ転送量とに基づいて、現段階の処理に要する残り時間長を推測し、該推測された残り時間長が、前記取得されたディスク回転開始時間以下の場合に、前記次段階でアクセスされるDBスキーマを記憶したディスク装置のディスクを前記第一の速度で回転させる、

請求項4記載の計算機システム。

【請求項6】

前記第二の速度は、0であり、

前記ディスク回転制御部が、前記把握された処理進行状況から、次々段階でアクセスされるDBスキーマの記憶位置を特定し、該記憶位置を備えたディスク装置のディスクを、前記第二の速度よりも速く前記第一の速度よりも遅い第三の速度で回転させる、

請求項3記載の計算機システム。

【請求項7】

前記DB管理情報には、各DBスキーマ毎にアクセスサイズが含まれており、

前記ディスク回転制御部が、前記次段階でアクセスされるDBスキーマのアクセスサイズを前記DB管理情報から特定し、該特定されたアクセスサイズが所定値以上の場合には、前記次々段階でアクセスされるDBスキーマを記憶したディスク装置のディスクを回転させない、

請求項6記載の計算機システム。

10

20

30

40

50

## 【請求項 8】

前記ディスク回転制御部が、ディスク装置のディスクを停止させる場合に、停止命令を発行し、

前記ディスクアクセス制御部は、前記ディスク回転制御部から発行された前記停止命令に基づいて、前記ディスク装置のディスクを停止する、

請求項 1 記載の計算機システム。

## 【請求項 9】

前記ディスク回転制御部が、クエリプランを構成するステップ単位ではなく、スキャン単位で、ディスク装置のディスクを前記第一の速度で回転させる、

請求項 1 記載の計算機システム。

10

## 【請求項 10】

前記ディスク回転制御部が、クエリプランを構成するステップ単位ではなく、スキャン単位で、ディスク装置のディスクを前記第二の速度で回転させる、

請求項 1 記載の計算機システム。

## 【請求項 11】

前記ディスク回転制御部は、DB 情報取得部と、ディスク回転指示部とを備え、

前記 DB 情報取得部は、前記 DBMS が受け付けたクエリのクエリプランに関する情報と、前記 DB 管理情報とを取得し、それらの情報を基に、該クエリでのスキャンの手順と、各スキャンでアクセスする DB スキーマとその記憶位置とを特定し、特定した内容を表す情報である DB 処理管理情報をディスク回転指示部に送信し、

20

前記ディスク回転指示部が、前記 DB 情報取得部から DB 処理管理情報を受信し、該 DB 処理管理情報を用いて、前記受け付けたクエリの処理の際のスキャンでアクセスされる DB スキーマの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを前記第一の速度で回転させる、

請求項 1 記載の計算機システム。

## 【請求項 12】

前記 DB 処理管理情報には、各 DB スキーマの ID と各 DB スキーマの記憶位置とを表した DB スキーマ位置情報と、どんな順番でスキャンし且つどのスキャン段階でどの DB スキーマにアクセスするかを表した DB 処理情報とが含まれ、

前記 DB 情報取得部は、前記 DBMS が起動した場合或いは前記 DB に変更があった場合に、前記 DB スキーマ位置情報を作成して前記ディスク回転指示部に送信し、前記 DBMS がクエリを受け付けた場合に、前記 DB 処理情報を作成して前記ディスク回転指示部に送信する、

30

請求項 1 1 記載の計算機システム。

## 【請求項 13】

前記複数のディスク装置には、二以上の論理ユニットが構築され、

前記ストレージシステムは、論理ユニット ID とディスク装置 ID との対応を表す記憶領域管理情報を記憶した記憶域を備え、

前記 DB 管理情報には、各 DB スキーマの ID と各 DB スキーマの記憶位置とを表し前記 DBMS が管理する DBMS 管理情報と、前記記憶位置と論理ユニット ID との対応を表し前記計算機のオペレーティングシステム (OS) が管理する OS 管理情報とが含まれ、

40

前記ディスク回転制御部は、前記 DBMS がクエリを受け付けた場合に、該クエリの処理の際にアクセスされる DB スキーマの記憶位置を前記 DBMS から取得し、該特定された記憶位置に対応する論理ユニット ID を前記 OS から取得し、前記取得した論理ユニット ID を含んだ第一の回転制御命令を発行し、

前記ディスクアクセス制御部が、前記発行された第一の回転制御命令を受信し、該受信した第一の回転制御命令中の論理ユニット ID に対応したディスク装置 ID を前記記憶領域管理情報から特定し、特定されたディスク装置 ID に対応したディスク装置のディスクを前記第一の速度で回転させる、

50

請求項 1 記載の計算機システム。

【請求項 1 4】

前記ディスク回転制御部が、前記 DBMS から発行されるアクセス要求を基に、該クエリの処理進行状況を把握し、該把握された処理進行状況から、現段階の処理を終了したことを特定した場合に、該現段階の処理でアクセスされたディスク装置のディスクを前記第二の速度で回転させる、

請求項 1 記載の計算機システム。

【請求項 1 5】

前記ディスク回転制御部が、前記現段階の処理が終了して次段階の処理が現段階の処理となる場合に、該終了した現段階の処理が次段階の処理となる場合には、該現段階の処理でアクセスされたディスク装置のディスクの回転速度を前記第二の速度としないようにする、

10

請求項 1 4 記載の計算機システム。

【請求項 1 6】

記憶メディアとしてのディスクを備えたディスク装置を複数台を備え、前記複数のディスク装置へのアクセスを制御し、前記複数のディスク装置にデータベース (DB) を記憶したストレージシステムと、

前記 DB を管理するための DB 管理情報を記憶し、前記 DB を操作するデータベース管理システム (DBMS) を実行する計算機と、

前記複数のディスク装置の回転を制御し、アクセスが発生しないディスク装置のディスクを、0 以上であって、アクセス時の回転速度である第一の速度よりも遅い第二の速度で回転させるようにさせるようになっているディスク回転制御部と

20

を備えた計算機システムで実現する方法であって、

前記 DB は、前記 DBMS にアクセスされる要素である DB スキーマを複数個含み、

前記 DB 管理情報は、どの DB スキーマがどの位置に記憶されているかを表す情報を含み、

前記 DBMS は、クエリを受け付け、前記クエリのクエリプランを作成し、該クエリプランに従って、前記 DB 管理情報から特定される DB スキーマに対するアクセス要求を発行し、

ディスク回転制御部は、前記 DBMS がクエリを受け付けた場合に、該クエリのクエリプランと前記 DB 管理情報とから、該クエリの処理の際にアクセスされる DB スキーマの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを、アクセス時の回転速度である第一の速度で回転させ、

30

前記 DBMS は、前記クエリの処理が終了すると、処理が終了した前記クエリの情報を、終了要求として前記ディスク回転制御部に送り、

前記ディスク回転制御部は、前記 DBMS からの前記終了要求を受信した場合に、前記クエリの処理の際にアクセスされた全てのディスク装置のディスクを停止させる、

ストレージシステムの消費電力低減方法。

【請求項 1 7】

記憶メディアとしてのディスクを備えたディスク装置を複数台と、前記複数のディスク装置へのアクセスを制御するディスクアクセス制御部とを備え、前記複数のディスク装置にデータベース (DB) を記憶したストレージシステムと接続された計算機であって、

40

前記 DB を管理するための DB 管理情報を記憶した DB 管理情報記憶域と、

前記 DB を操作するデータベース管理システム (DBMS) と、

前記複数のディスク装置の回転を制御し、アクセスが発生しないディスク装置のディスクを、0 以上であって、アクセス時の回転速度である第一の速度よりも遅い第二の速度で回転させるようにさせるようになっているディスク回転制御部と

を備え、

前記 DB は、前記 DBMS にアクセスされる要素である DB スキーマを複数個含んでおり、

50

前記DB管理情報は、どのDBスキーマがどの位置に記憶されているかを表す情報を含んでおり、

前記DBMSは、クエリを受け付け、受け付けたクエリのクエリプランを作成し、該クエリプランに従って、前記DB管理情報から特定されるDBスキーマに対するアクセス要求を発行し、

前記ディスク回転制御部が、前記DBMSがクエリを受け付けた場合に、該クエリのクエリプランと前記DB管理情報とから、該クエリの処理の際にアクセスされるDBスキーマの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを、アクセス時の回転速度である第一の速度で回転させ、

前記DBMSは、前記クエリの処理が終了すると、処理が終了した前記クエリの情報を、終了要求として前記ディスク回転制御部に送り、

前記ディスク回転制御部は、前記DBMSからの前記終了要求を受信した場合に、前記クエリの処理の際にアクセスされた全てのディスク装置のディスクを停止させる、  
計算機。

#### 【請求項18】

記憶メディアとしてのディスクを備えたディスク装置を複数台と、前記複数のディスク装置へのアクセスを制御するディスクアクセス制御部とを備え、前記複数のディスク装置にデータベース(DB)を記憶したストレージシステムと接続され、

前記DBを管理するためのDB管理情報を記憶し、

前記DBを操作するデータベース管理システム(DBMS)と、

前記複数のディスク装置の回転を制御し、アクセスが発生しないディスク装置のディスクを、0以上であって、アクセス時の回転速度である第一の速度よりも遅い第二の速度で回転させるようにさせるようになっているディスク回転制御部と、  
を備えた計算機で実現する方法であって、

前記DBは、前記DBMSにアクセスされる要素であるDBスキーマを複数個含み、

前記DB管理情報は、どのDBスキーマがどの位置に記憶されているかを表す情報を含み、

前記DBMSは、クエリを受け付け、受け付けたクエリのクエリプランを作成し、該クエリプランに従って、前記DB管理情報から特定されるDBスキーマに対するアクセス要求を発行し、

前記ディスク回転制御部が、前記DBMSがクエリを受け付けた場合に、該クエリのクエリプランと前記DB管理情報とから、該クエリの処理の際にアクセスされるDBスキーマの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを、アクセス時の回転速度である第一の速度で回転させ、

前記DBMSは、前記クエリの処理が終了すると、処理が終了した前記クエリの情報を、終了要求として前記ディスク回転制御部に送り、

前記ディスク回転制御部は、前記DBMSからの前記終了要求を受信した場合に、前記クエリの処理の際にアクセスされた全てのディスク装置のディスクを停止させる、  
ストレージシステムの消費電力低減方法。

#### 【発明の詳細な説明】

##### 【技術分野】

##### 【0001】

本発明は、データベース(以下「DB」)を記憶するストレージシステムの消費電力を低減するための技術に関する。

##### 【背景技術】

##### 【0002】

現在、DBを基盤とする多くのアプリケーションが存在し、DBに関する一連の処理/管理を行うデータベースマネジメントシステム(以下「DBMS」)は極めて重要なものとなっている。DBの特徴の一つは、膨大な量のデータを扱うことである。そのため、

10

20

30

40

50

DBMSが稼動する計算機システム（以下、DBシステムとも言う）の多くにおいては、DBMSが動作する計算機（以下、DBサーバ）に、複数台の記憶装置を備えた（つまり大容量の）ストレージシステムを接続し、そのストレージシステムにDBのデータを記憶させるシステム形態が一般的である。

【0003】

DBのデータは増加傾向にあり、それに伴ってストレージシステムの更なる大容量化が進んでいる。また、ストレージシステムに搭載される記憶装置としては、例えば、磁気テープ装置、或いは、ディスク（例えば、ハードディスク或いはDVD（Digital Versatile Disk））を搭載できる記憶装置（以下、「ディスク装置」と略記）がある。アクセス頻度の低いデータ（例えばアーカイブデータ）は、ディスク装置よりも価格の安い磁気テープ装置に記憶されることがあるが、ディスク装置の低価格化に伴い、ディスク装置に記憶されるケースが増えつつある。

10

【0004】

ストレージシステムの大容量化の方法の一つとして、搭載するディスク装置の台数を増やす方法がある。しかし、ディスクの回転には多くの電力が消費されるため、ディスク装置の台数の増加によって、ストレージシステム全体の消費電力が増大する。消費電力を制御するための技術として、例えば、特開平9-282057号公報（以下、文献1）及び特開2000-293314号公報（以下、文献2）が知られている。

【0005】

文献1に開示の技術は、ディスクアレイ装置において、上位装置からアクセスが無くなり予め定めた時間経過後、ディスク装置の節電（電源オンオフや節電モードの選択）を制御する。また、上位装置からアクセスがあった場合、アクセス先のディスク装置の電源がオフになっていれば、そのディスク装置の電源を投入した後、アクセスを実行する。

20

【0006】

文献2に開示の技術は、周辺装置を含むコンピュータシステムを構成する電源制御対象装置の電源を制御するための技術である。コンピュータシステムにおいてジョブが実行され、システム電源制御装置が、ジョブの実行計画や稼動実績に基づいて、各電源制御対象装置毎に、電源制御計画を作成し、作成した電源制御計画に従って、その電源制御計画に対応する電源制御対象装置の電源を制御する。

【0007】

【特許文献1】特開2000-293314号公報

【特許文献2】特開平9-282057号公報

【発明の開示】

【発明が解決しようとする課題】

【0008】

前述したように、DBのデータを記憶するためのストレージシステムのディスク装置の搭載台数は、増加傾向にあるため、そのようなストレージシステムの消費電力の低減のための技術が望まれる。

【0009】

ストレージシステムの消費電力を低減するための技術として、文献1に開示の技術を利用する技術、例えば、ディスク装置の電源をオフにしておき、アクセスがあった場合に、そのアクセス先のディスク装置の電源を投入する技術が考えられる。しかし、この技術では、例えば、ストレージシステムが上位装置からリード要求を受信した場合に、ディスク装置の電源を投入してからデータを読み出すことになるため、上位装置に対する読出しの応答が遅くなるという課題がある。

40

【0010】

一方、文献2に開示の技術によれば、文献2で言うコンピュータシステムは、DBサーバとストレージシステムとを含んだDBシステムに相当するため、電源制御対象装置は、DBサーバやストレージシステムとなる。このため、文献2に開示の技術では、ストレージシステム内のディスク装置の電源制御はできない。

50

## 【 0 0 1 1 】

仮に、文献 1 の技術と文献 2 の技術を組み合わせることができたとしても、ストレージシステム内のディスク装置を制御することによりストレージシステム全体の消費電力を低減することと、ストレージシステムの応答速度の劣化を抑えることの両方を実現することはできない。なぜなら、文献 1 と文献 2 の組み合わせから導き出せるのは、ストレージシステム本体の電源を投入し、ストレージシステムにアクセスが発生した場合に、アクセス先のディスク装置の電源がオフになっていればその電源を投入することにすぎないためである。

## 【 0 0 1 2 】

従って、本発明の目的は、DB のデータを記憶するストレージシステム内のディスク装置を制御することによりストレージシステム全体の消費電力を低減することと、ストレージシステムの応答速度の劣化を抑えることの両方を実現することにある。

## 【 課題を解決するための手段 】

## 【 0 0 1 3 】

本発明に従う計算機システムは、ストレージシステムと、計算機と、ディスク回転制御部とを備える。ストレージシステムは、記憶メディアとしてのディスクを備えたディスク装置を複数台と、前記複数のディスク装置へのアクセスを制御するディスクアクセス制御部とを備え、前記複数のディスク装置にデータベース (DB) を記憶する。計算機は、前記 DB を管理するための DB 管理情報を記憶した DB 管理情報記憶域を備え、前記 DB を操作するデータベース管理システム (DBMS) を実行する。ディスク回転制御部は、前記複数のディスク装置の回転を制御する。この計算機システムは、アクセスが発生しないディスク装置のディスクを、0 以上であって、アクセス時の回転速度である第一の速度よりも遅い第二の速度で回転させるようになっている。これは、ディスク回転制御部からの命令によりされても良いし、ストレージシステム内での制御により実現されても良い。前記 DB は、前記 DBMS にアクセスされる要素である DB スキーマを複数個含んでいる。前記 DB 管理情報は、どの DB スキーマがどの位置に記憶されているかを表す情報を含んでいる。前記 DBMS は、クエリを受け付け、受け付けたクエリのクエリプランを作成し、該クエリプランに従って、前記 DB 管理情報から特定される DB スキーマに対するアクセス要求を発行するようになっている。前記ディスク回転制御部が、前記 DBMS がクエリを受け付けた場合に、該クエリのクエリプランと前記 DB 管理情報とから、該クエリの処理の際にアクセスされる DB スキーマの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを前記第一の速度で回転させる。

## 【 0 0 1 4 】

第一の実施態様では、前記ディスク回転制御部が、前記 DBMS がクエリを受け付けた場合に、該クエリのクエリプランと前記 DB 管理情報とから、該クエリの処理の際にアクセスされる全ての DB スキーマにそれぞれ対応した全ての記憶位置を特定し、該特定された全ての記憶位置をそれぞれ備える全てのディスク装置のディスクを前記第一の速度で回転させることができる。

## 【 0 0 1 5 】

第二の実施態様では、前記ディスク回転制御部が、前記 DBMS がクエリの処理を終了した場合に、該クエリの処理でアクセスされた全てのディスク装置のディスクを前記第二の速度で回転させることができる。

## 【 0 0 1 6 】

第三の実施態様では、前記ディスク回転制御部が、前記 DBMS から発行されるアクセス要求を基に、該クエリの処理進行状況を把握し、該把握された処理進行状況から、次段階でアクセスされる DB スキーマの記憶位置を特定し、該記憶位置を備えたディスク装置のディスクを前記第一の速度で回転させることができる。

## 【 0 0 1 7 】

第四の実施態様では、前記第三の実施態様において、前記 DB 管理情報には、各 DB スキーマ毎にアクセスサイズが含まれていてもよい。前記ディスク回転制御部は、前記 DB

10

20

30

40

50

MSから発行されるアクセス要求を用いて、該アクセス要求に従ってアクセスされるDBスキーマに対応したアクセスサイズを前記DB管理情報から特定し、特定されたアクセスサイズと、前記把握された処理進行状況とに基づいて、前記次段階でアクセスされるDBスキーマを記憶したディスク装置のディスクを前記第一の速度で回転させる開始タイミングを制御することができる。

【0018】

第五の実施態様では、前記第四の実施態様において、前記ディスク回転制御部が、前記次段階でアクセスされるDBスキーマを記憶したディスク装置について、単位時間当たりのデータ転送量と、ディスクの回転速度が前記第一の速度になるのに要するディスク回転開始時間長とを取得することができる。また、前記ディスク回転制御部は、前記アクセス要求と、現段階でアクセスされるDBスキーマに対応したアクセスサイズとを基に、現段階の進捗を特定し、該特定された進捗と、前記取得した単位時間当たりのデータ転送量とに基づいて、現段階の処理に要する残り時間長を推測することができる。さらに、前記ディスク回転制御部は、該推測された残り時間長が、前記取得されたディスク回転開始時間以下の場合に、前記次段階でアクセスされるDBスキーマを記憶したディスク装置のディスクを前記第一の速度で回転させることができる。

10

【0019】

第六の実施態様では、前記第三の実施態様において、前記第二の速度は、0とすることができる。前記ディスク回転制御部が、前記把握された処理進行状況から、次々段階でアクセスされるDBスキーマの記憶位置を特定し、該記憶位置を備えたディスク装置のディスクを、前記第二の速度よりも速く前記第一の速度よりも遅い第三の速度で回転させることができる。

20

【0020】

第七の実施態様では、前記第六の実施態様において、前記DB管理情報には、各DBスキーマ毎にアクセスサイズが含まれていてもよい。前記ディスク回転制御部が、前記次段階でアクセスされるDBスキーマのアクセスサイズを前記DB管理情報から特定し、該特定されたアクセスサイズが所定値以上の場合には、前記次々段階でアクセスされるDBスキーマを記憶したディスク装置のディスクを回転させないようにすることができる。

【0021】

第八の実施態様では、前記ストレージシステムが、前記複数のディスク装置にそれぞれ対応したカウント値を記憶する記憶域を備えることができる。前記ディスク回転制御部が、ディスク装置のディスクを前記第一の速度で回転させる場合に、第一の回転制御命令を発行することができる。前記ディスクアクセス制御部が、前記ディスク回転制御部から前記第一の回転制御命令を受けた場合に、該第一の回転制御命令に従うディスク装置のディスクを回転させ、且つ、該第一の回転制御命令に対応するカウント値を増やし、所定のタイミングで、該カウント値を減らし、更新後のカウント値が所定値になった場合に、該更新後のカウント値に対応するディスク装置のディスクを第二の速度で回転させることができる。

30

【0022】

第九の実施態様では、前記第八の実施態様において、前記ディスク回転制御部が、ディスク装置のディスクを前記第二の速度で回転させる場合に、第二の回転制御命令を発行することができる。前記所定のタイミングは、前記ディスクアクセス制御部が前記ディスク回転制御部から前記第二の回転制御命令を受けたときとすることができる。

40

【0023】

第十の実施態様では、前記ディスク回転制御部が、クエリプランを構成するステップ単位ではなく、スキャン単位で、ディスク装置のディスクを前記第一の速度で回転させることができる。

【0024】

第十一の実施態様では、前記ディスク回転制御部が、クエリプランを構成するステップ単位ではなく、スキャン単位で、ディスク装置のディスクを前記第二の速度で回転させる

50



ことができる。

【 0 0 2 5 】

第十二の実施態様では、前記ディスク回転制御部は、D B 情報取得部と、ディスク回転指示部とを備えることができる。前記D B 情報取得部は、前記D B M S が受け付けたクエリのクエリプランに関する情報と、前記D B 管理情報とを取得し、それらの情報を基に、該クエリでのスキンの手順と、各スキンでアクセスするD B スキーマとその記憶位置とを特定し、特定した内容を表す情報であるD B 処理管理情報をディスク回転指示部に送信することができる。前記ディスク回転指示部が、前記D B 情報取得部からD B 処理管理情報を受信し、該D B 処理管理情報を用いて、前記受け付けたクエリの処理の際のスキンでアクセスされるD B スキーマの記憶位置を特定し、該特定された記憶位置を備えたディスク装置のディスクを前記第一の速度で回転させることができる。

10

【 0 0 2 6 】

第十三の実施態様では、前記第十二の実施態様において、前記D B 処理管理情報には、各D B スキーマのI D と各D B スキーマの記憶位置とを表したD B スキーマ位置情報と、どんな順番でスキンし且つどのスキン段階でどのD B スキーマにアクセスするかを表したD B 処理情報とが含まれる。前記D B 情報取得部は、前記D B M S が起動した場合或いは前記D B に変更があった場合に、前記D B スキーマ位置情報を作成して前記ディスク回転指示部に送信し、前記D B M S がクエリを受け付けた場合に、前記D B 処理情報を作成して前記ディスク回転指示部に送信することができる。

【 0 0 2 7 】

20

第十四の実施態様では、前記複数のディスク装置には、二以上の論理ユニットが構築される。前記ストレージシステムは、論理ユニットI D とディスク装置I D との対応を表す記憶領域管理情報を記憶した記憶域を備えることができる。前記D B 管理情報には、各D B スキーマのI D と各D B スキーマの記憶位置とを表し前記D B M S が管理するD B M S 管理情報と、前記記憶位置と論理ユニットI D との対応を表し前記計算機のオペレーティングシステム(O S )が管理するO S 管理情報とが含まれる。前記ディスク回転制御部は、前記D B M S がクエリを受け付けた場合に、該クエリの処理の際にアクセスされるD B スキーマの記憶位置を前記D B M S から取得し、該特定された記憶位置に対応する論理ユニットI D を前記O S から取得し、前記取得した論理ユニットI D を含んだ第一の回転制御命令を発行することができる。前記ディスクアクセス制御部が、前記発行された第一の回転制御命令を受信し、該受信した第一の回転制御命令中の論理ユニットI D に対応したディスク装置I D を前記記憶領域管理情報から特定し、特定されたディスク装置I D に対応したディスク装置のディスクを前記第一の速度で回転させることができる。

30

【 0 0 2 8 】

第十五の実施態様では、前記ディスク回転制御部が、前記D B M S から発行されるアクセス要求を基に、該クエリの処理進行状況を把握し、該把握された処理進行状況から、現段階の処理を終了したことを特定した場合に、該現段階の処理でアクセスされたディスク装置のディスクを前記第二の速度で回転させることができる。

【 0 0 2 9 】

第十六の実施態様では、前記第十五の実施態様において、前記ディスク回転制御部が、前記現段階の処理が終了して次段階の処理が現段階の処理となる場合に、該終了した現段階の処理が次段階の処理となる場合には、該現段階の処理でアクセスされたディスク装置のディスクの回転速度を前記第二の速度としないようにすることができる。

40

【 0 0 3 0 】

上述した各部は、各手段と言い換えてもよい。各部は、ハードウェア(例えば回路)、コンピュータプログラム、或いはそれらの組み合わせ(例えば、コンピュータプログラムを読み込んで実行する一又は複数のC P U )によって実現することもできる。各コンピュータプログラムは、コンピュータマシンに備えられる記憶資源(例えばメモリ)から読み込むことができる。その記憶資源には、C D - R O M やD V D (Digital Versatile Disk)等の記録媒体を介してインストールすることもできるし、インターネットやL A N等の

50

通信ネットワークを介してダウンロードすることもできる。

【発明の効果】

【0031】

本発明によれば、DBのデータを記憶するストレージシステム内のディスク装置を制御することによりストレージシステム全体の消費電力を低減することと、ストレージシステムの応答速度の劣化を抑えることの両方を実現することができる。

【発明を実施するための最良の形態】

【0032】

以下、本発明の実施の形態を説明する。まず、実施の形態の概要を説明する。

【0033】

本実施形態に係るDBシステムは、一種の計算機システムであり、DBMSが稼動する計算機（以下、DBサーバ）と、DBのデータを記憶するストレージシステムとが備えられる。また、ストレージシステムにディスクの回転に関する指示を与えるディスク回転指示プログラムが備えられる。ストレージシステムは、ディスク回転プログラムからの指示に従って、ディスクの回転制御を行う。

【0034】

本実施形態の特徴の一つとして、DBMSが受け付けたクエリに従う処理（以下、DB処理）が、そのクエリのプラン（以下、クエリプラン）に沿って進められる点に着目している。DB処理では、クエリプランに従う順番で、表或いは索引といったDBスキーマにアクセスが行われるが、本実施形態では、各DBスキーマと各DBスキーマの記憶位置との対応付けを表す情報（以下、DBスキーマ位置情報）が構築される。

【0035】

ディスク回転指示プログラムは、DBMSが受け付けたクエリのクエリプランに関する情報と、DBスキーマ位置情報とに基づいて、やがてアクセスされるDBスキーマとその記憶位置とを特定し、特定された記憶位置に該当するディスク装置のディスクの回転開始命令をストレージシステムに発行する。また、ディスク回転指示プログラムは、一つのクエリについてのDB処理が終了した時点で、そのDB処理で回転させたディスクの回転停止命令をストレージシステムに発行する。

【0036】

以下、本実施形態について詳細に説明する。尚、本実施形態に本発明が限定されるものではない。

【0037】

図1は、本発明の一実施形態に係るDBシステムの構成例を示した図である。

【0038】

このDBシステム（計算機システム）では、通信ネットワーク130を介してDBサーバ100とストレージシステム140とが接続される。DBサーバ100は、ストレージシステム140に格納されているDBを管理するDBMS118を実行する。ストレージシステム150は、DB等、DBサーバ100が使用するデータを格納する。通信ネットワーク130は、ローカルエリアネットワーク（LAN）やワイドエリアネットワーク（WAN）等のネットワークでも、ファイバチャネル等で構成されるネットワーク（ストレージエリアネットワーク：SAN）でも良い。尚、図1において、DBサーバ100、ストレージシステム140はそれぞれ1台のみ記載しているが、複数台でも良い。

【0039】

DBサーバ100は、例えば一般的な計算機とすることができる。具体的には、例えば、DBサーバ100は、内部バス102で接続されたCPU（制御プロセッサ）104、入出力装置106、記憶装置108、メモリ110及びI/F112（通信ネットワーク130とのインタフェース）を有する。メモリ110及び記憶装置108の少なくとも一方には、オペレーティングシステム（以下「OS」）114、DBMS118、DB情報取得プログラム122及びディスク回転指示プログラム124が格納される。

【0040】

10

20

30

40

50

OS 114、DBMS 118、DB情報取得プログラム 122、及びディスク回転指示プログラム 124は、CPU 104によって実行される。OS 110は、ローデバースとストレージシステム 140上の論理的な記憶領域とを対応付けるローデバース情報 116を保持する。尚、ローデバースとは、OSで管理する論理的な記憶装置単位である。DBMS 118は、DBの運用管理に必要となるDB管理情報 120を保持する。DBMS 118は、図示しないアプリケーションからクエリを受け付け、受け付けたクエリのクエリプランを作成し、そのクエリプランに従って、DBMS管理情報 120から特定されるアクセス先に対するアクセス要求（I/O要求とも言う）を発行する。DB情報取得プログラム 122は、DBに関する種々の情報を取得するための処理を実行する。ディスク回転指示プログラム 124は、ディスク回転の指示に必要となるDB処理管理情報 126や、  
10 ストレージシステム 140のディスクに関するディスク情報 128を保持する。

【0041】

ストレージシステム 140は、複数のディスク装置 152と、それらディスク装置 152に接続されたコントローラ 851とを備える。コントローラ 851は、例えば、内部バス 142で接続されたI/F 144（通信ネットワーク 130とのインタフェース）、CPU（制御プロセッサ）146、キャッシュメモリ 148及びメモリ 150を有する。メモリ 150には、ストレージシステム 140を制御する制御プログラム 154が格納され、CPU 146によって実行される。また、制御プログラム 154は、ストレージシステム 150の論理的な記憶領域とディスク装置 152が有する物理的な記憶領域を対応付ける記憶領域管理情報 156と、ディスク装置 152の回転状態を管理するためのディスク  
20 回転管理情報 158とを保持する。尚、ディスク装置 152は、例えば、ハードディスクドライブであり、ストレージシステム 150は、複数のディスク装置をRAID（Redundant Array of Independent (or Inexpensive) Disks）構成にしているも良い。尚、上述した各種プログラムは、可搬記憶媒体あるいは通信ネットワークを介して各装置にインストールされても良い。

【0042】

ストレージシステム 140がDBサーバ 100からライト要求及びデータを受信した場合、制御プログラム 154は、受信したデータをキャッシュメモリ 148に一時的に記憶させ、その後、キャッシュメモリ 148からそのデータを読み出して、ライト要求に従うアクセス先となるディスク装置 152にそのデータを書込む。ストレージシステム 140  
30 がDBサーバ 100からリード要求を受信した場合、制御プログラム 154は、リード要求に従うアクセス先となるディスク装置 152からデータを読み出してキャッシュメモリ 148に一時的に記憶させ、その後、キャッシュメモリ 148からそのデータを読み出してDBサーバ 100に送信する。

【0043】

以上が、本実施形態に係るDBシステムの構成についての説明である。

【0044】

なお、本実施形態において、ディスク回転指示プログラム 124はDBサーバ 100上で動作するが、ディスク回転指示プログラム 124は、DBサーバ及びストレージシステムと接続された別の計算機上、または、ストレージシステム上で動作しても良い。また、  
40 メモリ 150とキャッシュメモリ 148は、一体であっても良い。

【0045】

また、DBサーバ 100とストレージシステム 140は、一つの筐体に搭載されても良い。具体的には、例えば、DBサーバ 100は、いわゆるブレードサーバとして、ストレージシステム 140のコントローラ 851或いはディスク装置 152が搭載される筐体に  
搭載されても良い。

【0046】

また、コントローラ 851の上述した構成は一例であり、他の構成を採用することもできる。例えば、コントローラ 851は、上記の構成に代えて、外部の装置（例えばDBサーバ 100）との通信を制御する一以上の第一の制御部（例えば制御回路基板）と、ディ  
50

スク装置 152 との通信を制御する一以上の第二の制御部（例えば制御回路基板）と、外部の装置とディスク装置 152 との間で授受されるデータを記憶することができるキャッシュメモリと、ストレージシステム 140 を制御するためのデータを記憶することができる制御メモリと、各第一の制御部、各第二の制御部、キャッシュメモリ及び制御メモリを接続する接続部（例えば、クロスバスイッチなどのスイッチ）とを備えることができる。この場合、第一の制御部と第二の制御部の一方が、又は双方が協働して、コントローラ 851 としての処理を行うことができる。制御メモリは無くても良く、その場合、キャッシュメモリに、制御メモリが記憶する情報を記憶する領域が設けられても良い。

【0047】

さて、以下、前述した各種情報について詳細に説明する。

10

【0048】

図 2 (a) は、ストレージシステム 140 の制御プログラム 154 が保持する記憶領域管理情報 156 の構成例を示した図である。

【0049】

記憶領域情報 156 は、ストレージシステム 140 が有する論理的な記憶領域（以下「論理ユニット」と）と、実際にデータが記憶される物理的な記憶領域を対応付ける情報であり、論理ユニット毎にエントリを有する。各エントリは、論理ユニットを識別するための論理ユニット番号を登録するフィールド 200 と、論理ユニット内における論理ブロックの番号（以下「論理ブロックアドレス」）を登録するフィールド 202 と、論理ブロックと対応する物理ブロックが存在するディスク装置の番号（物理ディスク番号）を登録する

20

フィールド 204 と、論理ブロックと対応する物理ブロックの番号（以下「物理ブロックアドレス」）を登録するフィールド 206 とを有する。

【0050】

尚、ブロックとは、記憶領域を扱う際の単位を示すものであり、一般的に 512 バイトであることが多い。また、論理ブロックは物理ブロックと一対一で対応している必要があるが、全ての論理ブロックが同一のディスク上にある必要はない。ストレージシステム 140 は、DBサーバ 100 から論理ユニット及び論理ブロックアドレスによるアクセスを受け付ける。そして、ストレージシステム 140 は、論理ユニット及び論理ブロックに対応する物理ブロックの物理ディスク番号と物理ブロックアドレスを記憶領域情報 156 から割り出し、実際にアクセスを行う。

30

【0051】

図 2 (b) は、ストレージシステム 140 の制御プログラム 154 が保持するディスク回転管理情報 158 の構成例を示した図である。

【0052】

ディスク回転管理情報 158 は、ストレージシステム 140 が有するディスク装置 152 の回転状態を個別（物理ディスク単位）に管理する情報であり、ディスク装置 152 毎に対応するエントリを有する。各エントリは、個々のディスク装置を識別するための物理ディスク番号を登録するフィールド 220 と、ディスク装置の回転状態を示すためのカウント値（以下、回転状態カウント値）を登録するフィールド 222 とを有する。

【0053】

40

図 3 は、DBサーバ 100 の OS 114 が保持するローデバース情報 116 の構成例を示した図である。

【0054】

ローデバース情報 116 は、OS 114 が持つローデバースと、そのローデバースに割り当てられたストレージシステム 140 上の記憶領域とを対応付ける情報である。ローデバース情報 116 は、ローデバース毎に対応するエントリを有する。各エントリは、ローデバースのファイル名を登録するフィールド 300 と、ローデバースに対して割り当てられた記憶領域が存在するストレージシステムのアドレスを登録するフィールド 302 と、割り当てられた記憶領域が存在する論理ユニットの論理ユニット番号を登録するフィールド 304 と、割り当てられた記憶領域の論理ユニット上での先頭論理ブロックアドレスを

50

登録するフィールド 306 と、割り当てられた記憶領域の論理ブロック数を登録するフィールド 308 とを有する。

【0055】

図 4 ( a ) は、DBMS 118 が保持する DB 管理情報 120 の構成例を示した図である。尚、DBMS 管理情報 120 は、DB の初期設定時や運用時などに、管理者等によって設定される。

【0056】

DBMS 管理情報 120 は、DB システムファイルの設定情報である DB システムファイル情報 400 と、データ領域の設定情報であるデータ領域情報 420 と、テーブルやインデックスといった DB スキーマの設定情報である DB スキーマ情報 440 とを有する。

10

【0057】

図 4 ( b ) は、DB システムファイル情報 400 の構成例を示した図である。

【0058】

DB システムファイル情報 400 は、DB システムファイルと、前述のローデバイスとを対応付ける情報である。DB システムファイル情報 400 は、DB システムファイル毎にエントリを有する。各エントリは、DB システムファイルのファイル名を登録するフィールド 402 と、DB システムファイルが登録されるローデバイスのファイル名を登録するフィールド 404 とを有する。尚、DB システムファイルは、ローデバイスに作られる一つの大きなファイルである。DBMS 118 は、この DB システムファイルに DB のデータを書き込むことで DB を構築する。また、DB システムファイルに書き込まれた DB のデータについて、OS 114 が、実際にデータを書き込むストレージシステム 140 の論理的な記憶領域を前述のローデバイス情報 116 から割り出す。そして割り出した記憶領域に対するデータの書込み命令を OS 114 がストレージシステム 150 に発行する。

20

【0059】

図 4 ( c ) は、データ領域情報 420 の構成例を示した図である。

【0060】

データ領域情報 420 は、DBMS 118 が管理するデータ領域の設定情報である。データ領域情報 420 は、データ領域毎にエントリを有する。各エントリは、データ領域を識別するためのデータ領域 ID を登録するフィールド 422 と、データ領域の名前を登録するフィールド 424 と、データ領域が作られる DB システムファイルのファイル名を登録するフィールド 426 と、データ領域に割り当てた領域サイズを登録するフィールド 428 とを有する。尚、データ領域とは、DB のデータを記憶するデータ領域であり、DB システムファイル上に作成される。

30

【0061】

図 4 ( d ) は、DB スキーマ情報 440 の構成例を示した図である。

【0062】

DB スキーマ情報 440 は、DBMS 114 が管理する表 ( テーブル ) や索引 ( インデックス ) といった DB スキーマに関する設定情報である。DB スキーマ情報 440 は、DB スキーマ毎にエントリを有する。各エントリは、DB スキーマを識別するための DB スキーマ ID を登録するフィールド 442 と、DB スキーマの名前を登録するフィールド 444 と、DB スキーマの種別を登録するフィールド 446 と、DB スキーマが書き込まれるデータ領域の ID を登録するフィールド 448 と、DB スキーマのサイズを登録するフィールド 450 と、DB スキーマが書き込まれるデータ領域上のオフセットを登録するフィールド 452 とを有する。

40

【0063】

図 5 ( a ) は、ディスク回転指示プログラム 124 が保持する DB 処理管理情報 126 の構成例を示した図である。尚、DB 処理管理情報 126 は、DB 情報取得プログラム 122 が DBMS 118 や OS 114 から必要な情報を取得し、ディスク回転指示プログラム 124 に送信することによって自動的に作成される。

【0064】

50

DB 処理管理情報 1 2 6 は、各 DB スキーマの論理的な位置情報である DB スキーマ位置情報 5 0 0 と、DB 処理に関する情報である DB 処理情報 5 2 0 とを有する。

【 0 0 6 5 】

図 5 ( b ) は、DB スキーマ位置情報 5 0 0 の構成例を示した図である。

【 0 0 6 6 】

DB スキーマ位置情報 5 0 0 は、前述の DB スキーマとその記憶位置を示す情報である。DB スキーマ位置情報 5 0 0 は、DB スキーマ毎にエントリを有する。各エントリは、DB スキーマを識別するためのスキーマ ID を登録するフィールド 5 0 2 と、DB スキーマを有する DBMS を識別するための DBMS ID を登録するフィールド 5 0 4 と、DB スキーマの種別を登録するフィールド 5 0 6 と、DB スキーマが書き込まれたストレージシステムのアドレスを登録するフィールド 5 0 8 と、DB スキーマが書き込まれたストレージシステム 1 4 0 の論理ユニット番号を登録するフィールド 5 1 0 とを有する。尚、DBMS ID は管理者等によってあらかじめ各 DBMS に一意の値として割り当てられている値である。

【 0 0 6 7 】

図 5 ( c ) は、DB 処理情報 5 2 0 の構成例を示した図である。

【 0 0 6 8 】

DB 処理情報 5 2 0 は、DBMS が実行するクエリ、及びそのクエリ内で実行される検索（テーブルスキャンやインデックススキャン等。以下「スキャン」ともいう）に関する情報である。DB 処理情報 5 2 0 は、クエリのステップ毎にエントリを有する。各エントリは、クエリを実行する DBMS を識別するための DBMS ID を登録するフィールド 5 2 2 と、実行するクエリを識別するためのクエリ ID を登録するフィールド 5 2 4 と、クエリ内で実行されるスキャンを識別するためのスキャン ID を登録するフィールド 5 2 6 と、そのスキャンでアクセスされる DB スキーマの ID を登録するフィールド 5 2 8 と、アクセスされる DB スキーマのサイズを登録するフィールド 5 3 0 と、スキャンの種別を登録するフィールド 5 3 2 とを有する。尚、スキャン ID は、そのクエリの中で実行されるスキャンを一意的な値として割り当てられ、その値が小さいものから順番に実行される。

【 0 0 6 9 】

図 5 ( d ) は、ディスク回転指示プログラム 1 2 4 が保持するディスク情報 1 2 8 の構成例を示した図である。尚、ディスク情報 1 2 8 は、ディスク回転指示プログラム 1 2 4 の起動時などに管理者などが設定する。

【 0 0 7 0 】

ディスク情報 1 2 8 は、DB のデータを記憶しているストレージシステム 1 4 0 のディスク装置 1 5 2 の回転開始時間、及び 1 秒辺りの平均データ転送量に関する情報である。一台のストレージシステムに搭載される複数のディスクは、同一のディスクが搭載されることがほとんどである。そこで、ディスク情報 1 2 8 は、DB サーバ 1 0 0 に接続されるストレージシステム 1 4 0 毎にエントリを有する。各エントリは、ストレージシステム 1 4 0 のアドレスを登録するフィールド 5 4 0 と、ストレージシステム 1 4 0 に搭載されているディスク装置 1 5 2 のディスク回転開始時間を登録するフィールド 5 4 2 と、ストレージシステム 1 4 0 の 1 秒辺りの平均データ転送量を登録するフィールド 5 4 4 とを有する。尚、ディスク回転開始時間とは、ディスクの回転が停止している状態から回転を始め、回転が安定するまでの時間（換言すれば、実際にアクセスされる時の回転速度になるまでの時間）である。なお、一台のストレージシステムに異なる種類（例えば I/F 種類）のディスク装置が混在する場合には（例えば、SCSI I/F を有するディスク装置と ATA I/F を有するディスク装置とが混在する場合には）、ディスク装置毎にエントリが設けられても良い。

【 0 0 7 1 】

図 6 ( a ) は、DB 処理実行情報 6 0 0 の構成例を示した図である。

【 0 0 7 2 】

10

20

30

40

50

DB 処理実行情報 600 は、クエリに従う DB 処理の実行中に DB データへのアクセス要求を DBMS 118 が発行する際に、DB 情報取得プログラム 122 がディスク回転指示プログラム 124 に送信する情報である。DB 処理実行情報 600 は、アクセスの要求元となったクエリを実行している DBMS の ID を登録するフィールド 602 と、アクセスの要求元になったクエリのクエリ ID を登録するフィールド 604 と、アクセスの要求元になったスキンのスキャン ID を登録するフィールド 606 と、そのアクセスの時点における該当スキンの進捗を示す進捗情報を登録するフィールド 608 とを有する。尚、進捗の示し方は種々の方法を採用できる。具体的には、例えば、進捗情報は、そのスキンの読み出したデータ量のスキーマサイズ合計（そのスキンの ID に属する DB スキーマのサイズの和）に対する割り合いでも良いし、そのスキンのアクセスする全体の行のうち、どれだけに行の検索が終わったかを示す値でも良い。

10

## 【0073】

図 6 (b) は、DB 処理終了情報 620 の構成例を示した図である。

## 【0074】

DB 処理終了情報 620 は、DBMS 118 がクエリを終了する際に、DB 情報取得プログラム 122 がディスク回転指示プログラム 124 に送信する情報である。DB 処理終了情報 620 は、終了するクエリを実行していた DBMS の DBMS ID を登録するフィールド 622 と、終了するクエリのクエリ ID を登録するフィールド 624 とを有する。

## 【0075】

20

図 7 (a) は、ディスク回転指示プログラム 124 がストレージシステム 140 に対して発行するディスク回転開始命令 700 の構成例を示した図である。

## 【0076】

ディスク回転開始命令 700 は、ディスクの回転を開始させたい論理ユニット番号のリスト 702 を有する。

## 【0077】

図 7 (b) は、ディスク回転指示プログラム 124 がストレージシステム 140 に対して発行するディスク回転停止命令 710 の構成例を示した図である。

## 【0078】

ディスク回転停止命令 710 は、ディスクの回転を停止させたい論理ユニット番号のリスト 712 を有する。

30

## 【0079】

以下、DB 情報取得プログラム 122 が実行する各種情報の取得送信処理（以下「DB 情報取得送信処理」）、ディスク回転指示プログラム 124 が実行するディスクの回転指示に関する処理（以下「ディスク回転指示処理」）、ストレージ 140 の制御プログラムが実行するディスク回転の制御に関する処理（以下「ディスク回転制御処理」）の手順について説明する。尚、DB は構築済み（DBMS 管理情報 120 は作成済み）であるものとする。

## 【0080】

図 8 は、DB 情報取得送信処理の手順例を示したフロー図である。以下、プログラムが主語になる場合は、実際にはそのプログラムを実行する CPU によって処理が行われるものとする。

40

## 【0081】

システム管理者は、DB サーバ 100 上で DB 情報取得プログラム 122 を起動し、DB 情報取得送信処理を開始する（ステップ 800）。

## 【0082】

DB 情報取得プログラム 122 は、起動直後、スリープ状態（待機状態）に入り、以下に説明する割り込みを待ち続ける（ステップ 802）。

## 【0083】

DBMS 118 が起動した場合（ステップ 804 で Yes）、又は DB に変更があった

50

場合（ステップ804でNo、ステップ806でYes）、DB情報取得プログラム122は、DBMS118が保持するDBMS管理情報120や、OS114が保持するローデバイス情報116から、DBスキーマに関する情報を取得する（ステップ808）。具体的には、例えば、スキーマID、スキーマ種別及びデータ領域IDがDBスキーマ情報440から取得され、そのデータ領域IDに対応するDBシステムファイル名がデータ領域情報420から取得され、そのDBシステムファイル名に対応するローデバイスファイル名がDBシステムファイル情報400から取得され、そのローデバイスファイル名に対応するストレージシステムアドレス及び論理ユニット番号がローデバイス情報116から取得される。

#### 【0084】

DB情報取得プログラム122は、取得した情報と、DBMS管理情報120を保持するDBMS118のIDとを含んだDBスキーマ位置情報500を作成し、作成したDBスキーマ位置情報500をディスク回転指示プログラムに送信する（ステップ810）。尚、DB情報取得プログラム122が起動した時点でDBMS118が動作している場合、DB情報取得プログラム122は、起動後すぐにステップ808及びステップ810の処理を実行する。

#### 【0085】

DBMS118がクエリを受け付けた場合（ステップ812でYes）、DBMS118によって、そのクエリのクエリプランが作成される。DB情報取得プログラム122は、そのクエリプランに関する情報をDBMS118から取得し（ステップ814）、その情報を用いてDB処理情報520を作成して、作成したDB処理情報520をディスク回転指示プログラムに送信する（ステップ816）。以下、これについて、図11に例示するクエリプランの場合を例に採り説明する。図11に例示するクエリプランは、複数のステップが木構造で構成される。このクエリプランによれば、3つのスキャン処理、具体的には、スキーマ名「T1」のテーブルをスキャンする処理、スキーマ名「I2」の索引を使ってスキーマ名「T2」のテーブルをスキャンする処理、及び、スキーマ名「I2」の索引を使ってスキーマ名「T3」のテーブルをスキャンする処理が行われる。DB情報取得プログラム122は、各スキャン処理毎に固有のIDを付与する。また、DB情報取得プログラム122は、各スキーマ名に対応するスキーマID及びスキーマサイズをDBスキーマ情報440から取得する。また、DB情報取得プログラム122は、取得したスキーマIDに対応するDBMS IDをDBスキーマ位置情報500から取得する。そして、DB情報取得プログラム122は、取得したDBMS ID、DBMS118が受けたクエリのID、上記付与したスキャンID、上記取得されたスキーマID及びスキーマサイズ、各スキャン処理のスキャン種別を含んだDB処理情報520を作成し、作成したDB処理情報520をディスク回転指示プログラム124に送信する。DB処理情報520がディスク回転指示プログラム124に入力された場合、ディスク回転指示プログラム124によって、初めのスキャンでのアクセスされるDBスキーマを記憶した論理ユニットが特定され、その論理ユニットの番号を含んだディスク回転開始命令700が、ストレージシステム140に発行される。それにより、初めのスキャンのためのアクセス要求がDBサーバ100からストレージシステム140に届いてディスク装置152にアクセスが発生するよりも前に、そのディスク装置152のディスクを回転させておくことが期待できる。

#### 【0086】

クエリ実行中（つまりDB処理中）、DBMS118がDBのデータにアクセスする場合、DBMS118から、アクセス要求が発行される。そのアクセス要求には、例えば、スキーマIDや、そのスキーマIDに対応するDBスキーマのどこにアクセスするか（例えばページ番号や行番号）などの情報が含まれている。DBサーバ100では、OS114が、DBMS118からアクセス要求を受け、そのアクセス要求に含まれている情報要素を検索キーにローデバイス情報116から必要な情報を取得し、取得した情報を含んだアクセス要求（例えば論理ユニット番号を含んだアクセス要求）をストレージシステム1

10

20

30

40

50



40に発行する。一つのスキャン処理において、DBMS118からアクセス要求が複数回発行されることがある。DBMS118から発行されるアクセス要求に含まれている情報要素（例えば、DBスキーマIDや、DBスキーマのどこにアクセスするか）を参照することにより、一つのスキャン処理における進捗状況を特定したり、別のスキャン処理が開始されることを特定したりすることができる。

【0087】

DBMS118からアクセス要求が発行される場合（ステップ818でYes）、DB情報取得プログラム122は、そのアクセス要求中の情報要素を取得し（ステップ820）、その情報要素を基にDB処理実行情報600を作成し、そのDB処理実行情報600をディスク回転指示プログラム124に送信する（ステップ822）。DBMS ID、クエリID及びスキャンIDは、例えば、アクセス要求中のスキーマIDを検索キーにDB処理情報520から取得することができる。進捗情報は、例えば、アクセス要求中のスキーマIDを検索キーに取得できるスキーマサイズと、アクセス要求中の所定の情報要素（例えばDBスキーマのどこにアクセスするか）とに基づいて、特定することができる。

10

【0088】

DBMS118がクエリの実行を終了した場合（ステップ824でYes）、DB情報取得プログラム122は、終了したクエリに関する情報をDBMS118から取得し（ステップ826）、DB処理終了情報640を作成してディスク回転指示プログラム124に送信する（ステップ828）。

【0089】

以上が、DB情報取得送信処理の手順例についての説明である。なお、DBMS118が受け付けたクエリが複数個ある場合には、それら複数のクエリの各々について、上述した処理が行われる。

20

【0090】

また、DB情報取得プログラム122が、ステップ804でYes、ステップ806でYes、ステップ812でYes、ステップ818でYes、及びステップ824でYesとなったことを検知できるようにするための方法として、種々の方法を採用することができる。具体的には、例えば、DBMS118内部に、下記の（1）乃至（5）のタイミングで、DB情報取得プログラム122に情報を通知するインタフェースを設け、そのインタフェースを通じて、DB情報取得プログラム122は検知することができる。

30

（1）DBMS起動後（この場合、例えば、DBMS118がDB情報取得プログラム122に対し、起動したことの通知を、起動と同時に送信する。）

（2）DBに変更があった時（この場合、例えば、DBMS118がDB情報取得プログラム122に対し、DBを変更したことの通知を、変更と同時に送信する。）

（3）クエリ受付時（この場合、例えば、DBMS118がDB情報取得プログラム122に対し、受け付けられたクエリのクエリプランに関する情報を、クエリプランの作成と同時に通知する。）

（4）DBMSのアクセス要求発行時（この場合、例えば、DBMS118がDB情報取得プログラム122に対し、アクセス先に関する情報を、アクセス要求発行と同時に送信する。）

40

（5）クエリ終了時（この場合、例えば、DBMS118がDB情報取得プログラム122に対し、終了したクエリに関する情報を、クエリの処理の終了と同時に送信する。）

尚、DB情報取得プログラム122は、上記（1）及び（2）を受け付けた場合、前述したように、DBスキーマ位置に関する情報をOS114やDBMS118から取得し、DBスキーマ位置情報500を作成してディスク回転指示プログラム124に送信する。また、DB情報取得プログラム122は、上記（3）、（4）及び（5）を受け付けた場合は、DBMS118から受けた情報を基に、ディスク回転指示プログラム124に送信する情報を作成して送信する。

【0091】

さて、次に、ディスク回転指示処理の手順例について説明する。

50

## 【 0 0 9 2 】

図 9 は、ディスク回転指示処理の手順例を示したフロー図である。

## 【 0 0 9 3 】

システム管理者は、ディスク回転指示プログラム 1 2 4 を起動し、ディスク回転指示処理を開始する（ステップ 9 0 0）。尚、ディスク回転指示プログラム 1 2 4 は、前述の DB 情報取得プログラム 1 2 2 より先に起動しておく必要がある。

## 【 0 0 9 4 】

ディスク回転指示プログラム 1 2 4 は、起動直後、スリープ状態（待機状態）に入り、DB 情報取得プログラム 1 2 2 から DB スキーマ位置情報 5 0 0、DB 処理情報 5 2 0、DB 処理実行情報 6 0 0、或いは DB 処理終了情報 6 2 0 が送られてくるのを待つ（ステップ 9 0 2）。尚、初期の状態として、ストレージシステム 1 4 0 内の全てのディスク装置 1 5 2 の回転は、停止状態（つまり全く回転していない状態）にある。

10

## 【 0 0 9 5 】

DB スキーマ位置情報 5 0 0 を受信した場合（ステップ 9 0 4 で Y e s）、ディスク回転指示プログラム 1 2 4 は、受信した DB スキーマ位置情報 5 0 0 を DB 処理管理情報 1 2 6 の一部としてメモリ 1 1 0 上に格納する（ステップ 9 0 6）。

## 【 0 0 9 6 】

DB 処理情報 5 2 0 を受信した場合（ステップ 9 0 8）、ディスク回転指示プログラム 1 2 4 は、受信した DB 処理情報 5 2 0 を DB 処理管理情報 1 2 6 の一部としてメモリ 1 1 0 上に格納する（ステップ 9 1 0）。ディスク回転指示プログラム 1 2 4 は、ステップ 9 1 0 で格納した DB 処理情報 5 2 0 を参照して、スキャン ID「0」のエントリを検索することで、同エントリのアクセス対象スキーマ ID を特定し、そのアクセス対象スキーマ ID から、スキャン ID「0」に対応するスキャンでアクセスする DB スキーマを識別する。尚、スキャン ID「0」のスキャンは、そのスキャンが実行されるクエリの中で最初に実行されるスキャンを意味する（別の言い方をすれば、スキャン ID が若い順に、スキャンが実行される）。ディスク回転指示プログラム 1 2 4 は、その最初のスキャンでアクセスされるディスク装置の回転開始の指示を行う。具体的には、ディスク回転指示プログラム 1 2 4 は、特定されたアクセス対象スキーマ ID をキーにして DB スキーマ位置情報 5 0 0 を検索し、該当の DB スキーマが記憶されているストレージシステム 1 4 0 のアドレスと、論理ユニット番号とを特定する（ステップ 9 1 2）。ディスク回転指示プログラム 1 2 4 は、ステップ 9 1 2 で特定したアドレスに対応するストレージシステム 1 4 0 に、ステップ 9 1 2 で特定した論理ユニット番号を含むディスク回転開始命令 7 0 0 を発行する（ステップ 9 1 4）。尚、ディスク回転指示プログラム 1 2 4 は、DB 処理情報 5 2 0 を受信した場合に、その DB 処理情報 5 2 0 に対応するクエリの DB 処理においてアクセスする DB スキーマを全て特定し、特定された DB スキーマを記憶する論理ユニット番号を全て特定し、特定された論理ユニット番号を含んだディスク回転開始命令 7 0 0 を、ストレージシステム 1 4 0 に発行しても良い。これにより、その DB 処理においてアクセスされる全てのディスク装置のディスクを予め回転させておくことができる。この場合、ステップ 9 1 6 及びステップ 9 1 8 の処理は不要となり、DB 処理実行情報 6 0 0 の送受信処理も不要となるため、その分のオーバーヘッドが削減され性能向上が見込める。但し、クエリ内のスキャン毎のディスク回転制御ができないため、一つの DB 処理の際の消費電力は、DB 処理の進行に伴って順次にディスク装置を回転させていく場合に比して大きくなる。

20

30

40

## 【 0 0 9 7 】

DB 処理実行情報 6 0 0 を受信した場合（ステップ 9 1 6 で Y e s）、ディスク回転指示プログラム 1 2 4 は、受信した DB 処理実行情報 6 0 0 の DB MS ID、クエリ ID、スキャン ID と同じ値を持つ DB 処理情報 5 2 0 のエントリを検索し、同エントリのスキーマサイズ、ディスク情報 1 2 8 の 1 秒辺りの平均データ転送量、及び、DB 処理実行情報 6 0 0 の進捗情報から、そのスキャン ID に対応するスキャンが終わるまでの残り時間長を推測する。その結果、推測された残り時間長が、アクセス先となるストレージス

50

テム（検索キーとされたスキャンIDに紐付けられたスキーマIDに対応したストレージシステム）140のディスク回転開始時間（ディスク情報128に書かれている時間）よりも短ければ（又は以下であれば）、ディスク回転指示プログラム124は、次のスキャンでアクセスされるDBスキーマの論理ユニット番号を特定し、特定された論理ユニット番号を含んだディスク回転開始命令700を、そのストレージシステム140に発行する（ステップ918）。これにより、次のスキャンでアクセスされるディスク装置152のディスクを、DBサーバ100からそのディスク装置152に対するアクセス要求が発行される前に、回転させておくことができる。

【0098】

DB処理終了情報620を受信した場合（ステップ920でYes）、ディスク回転指示プログラム124は、受信したDB処理終了情報620のDBMS IDとクエリIDの値を持つDB処理情報520のエントリを検索し、同エントリのアクセス対象スキーマIDを特定する。ディスク回転指示プログラム124は、特定されたスキーマIDをキーにしてDBスキーマ位置情報500を検索し、そのスキーマIDに対応するストレージシステムアドレスと論理ユニット番号とを特定する（ステップ922）。ディスク回転指示プログラム124は、ステップ922で特定されたアドレスに対応するストレージシステム140に対して、ステップ922で特定された論理ユニット番号を含んだディスク回転停止命令を発行する（ステップ924）。尚、ステップ922において、条件に合致する全てのエントリに対して、ステップ922、及びステップ924を繰り返す。

【0099】

図10は、ディスク回転制御処理の手順例を示すフロー図である。尚、制御プログラム154はストレージシステム140の制御を行うプログラムであるが、このフロー図ではディスクの回転制御に関する部分だけを示している。

【0100】

ディスク回転指示プログラム124からディスク回転開始命令700を受信した場合（ステップ1002でYes）、制御プログラム154は、記憶領域管理情報156を参照し、当該命令700に含まれている対象論理ユニット番号に対応する物理ディスク番号を特定する（ステップ1004）。制御プログラム154は、ステップ1004で特定した物理ディスク番号をキーに、ディスク回転管理情報158を参照して対応するエントリを検索する。該当エントリの回転状態カウント値が0の場合、制御プログラム154は、対応するディスク装置に対して所定の命令を発行することで、そのディスク装置のディスクの回転を開始させる（ステップ1006）。制御プログラム154は、ステップ1006で検索したエントリの回転状態カウント値に1を加算する（ステップ1008）。尚、1つの論理ユニットに対して複数のディスク装置が割り当てられている場合は、そのディスク装置の数だけ、ステップ1006とステップ1008が繰り返し実行される。また、ステップ1006において、回転状態カウント値が0でない場合（つまり1以上の場合）、該当するディスク装置は回転中なので、回転を開始する処理は不要となり、回転状態カウント値が加算されれば良い。

【0101】

ディスク回転指示プログラム124からディスク回転停止命令710を受信した場合（ステップ1002でNo、ステップ1010でYes）、制御プログラム154は、記憶領域管理情報156を参照し、当該命令710の対象論理ユニット番号に対応する物理ディスク番号を特定する（ステップ1012）。制御プログラム154は、ステップ1012で特定された物理ディスク番号をキーに、ディスク回転管理情報158を参照し対応するエントリを検索し、該当エントリの回転状態カウント値から1を減算する（ステップ1014）。制御プログラム154は、ステップ1006の減算によって回転状態カウント値が0になった場合、対応するディスク装置に所定の命令を発行することで、そのディスク装置のディスクの回転を停止させる（ステップ1016）。尚、1つの論理ユニットに対して複数のディスク装置が割り当てられている場合は、そのディスク装置の数だけ、ステップ1014とステップ1016が繰り返し実行される。

## 【 0 1 0 2 】

以上が、ディスク回転指示処理の手順例及びディスク回転制御処理の手順例についての説明である。なお、DBMS 118が複数のクエリを受け付けた場合には、DB処理情報の受信が複数回行われる。具体的には、例えば、DBMS 118が或るクエリの実行中に、新たにクエリを受け付けた場合には、新たにDB処理情報の受信が行われる。この場合、或る論理ユニットに対するディスク回転開始命令700は、一つのクエリのDB処理では一回しか発生しなくても、別のクエリのDB処理の際に一回以上発行される場合がある。別の言い方をすれば、或る論理ユニットを構築する或るディスク装置152の回転状態カウンタ値(図2(b)参照)は、一つのクエリが終了したことによってディスク回転停止命令710が発行されても、別のクエリのDB処理情報を基にディスク回転開始命令が発行されたことによってカウンタ値が加算されていたために、「0」にならない場合がある。この場合、一つのクエリの処理が終了しても、次のクエリの処理でやがてそのディスク装置152にアクセスが発生するので、その或るディスク装置152のディスクの回転が継続されることになる。このようにして、複数のクエリにも対処することができる。なお、ディスク回転指示プログラム124は、一つのクエリ毎に、DB処理管理情報126を保持し、一つのクエリの処理が終了した場合に(例えば、DB処理終了情報620を受けたときに)、その終了したクエリに対応するDB処理管理情報126を破棄してもよい。また、各クエリ或いは各スキャン処理に優先順位が付与される場合には、その優先順位に従ってディスク回転開始命令が発行されても良い。

10

## 【 0 1 0 3 】

以上、上述した実施形態によれば、DBMS 118がクエリに従うDB処理を実行していない場合には、ストレージシステム140のディスク装置152が停止状態となっている。そして、DBMS 118がクエリを受け付け、そのクエリのクエリプランを基に作成されたDB処理情報520がディスク回転指示プログラム124に入力された場合に、ディスク回転指示プログラム124が、そのDB処理情報520から、やがてアクセスされる論理ユニットを特定し、特定された論理ユニットに対するディスク回転開始命令が発行され、それにより、その論理ユニットを構築するディスク装置152のディスクが回転させられる。そして、そのディスク装置152に対してアクセスが発生しないとわかった場合、具体的には、クエリに従うDB処理が終了した場合に、そのディスク装置152により構築された論理ユニットに対するディスク回転停止命令が発行され、それにより、その論理ユニットを構築するディスク装置152のディスクの回転が停止させられる。

20

30

## 【 0 1 0 4 】

つまり、この実施形態によれば、ディスク装置152に対してアクセスの発生しない状態では、ディスク装置152のディスクの回転は停止状態とされ、クエリプランに基づくDB処理情報520から、やがてアクセスされる論理ユニットが特定された場合に、その論理ユニットを構築するディスク装置152が回転させられる。これにより、ストレージシステム140全体の消費電力を低減することができる。

## 【 0 1 0 5 】

また、ディスク装置152のディスクの回転は、そのディスク装置152に対して実際にアクセスが発生するよりも前に行われる。具体的には、ディスク装置152のディスクに実際にアクセスが行われる頃には、そのディスクの回転が安定した状態になっているようなタイミングで、そのディスク装置152により構築された論理ユニットに対するディスク回転開始命令700が発行される。これにより、ストレージシステム140の応答速度の劣化を抑えることができる。

40

## 【 0 1 0 6 】

従って、本実施形態によれば、ストレージシステム140全体の消費電力を低減することと、ストレージシステム140の応答速度の劣化を抑えることの両方を実現することができる。

## 【 0 1 0 7 】

また、本実施形態によれば、ストレージシステム140の応答速度の劣化が抑えられる

50

ものの、ディスク装置 152 を常時回転させておくストレージシステムに比べれば、応答速度が遅くなり得る。このため、本実施形態に係る DB システムは、頻繁にアクセスが生じる故に応答速度（例えばデータの読出しの速度）が最優先されるようなオンライン系のシステムよりも、あまり頻繁にアクセスが生じないようないわゆるアーカイブシステムに適用することが好ましいと考えられる。

【0108】

さて、上述した実施形態には、幾つかの変形例が考えられる。以下、各変形例について説明する。

【0109】

第一の変形例では、各ディスク装置 152 が、複数種類の回転速度でディスクを回転することができるようになっている。具体的には、例えば、各ディスク装置 152 が、アクセス時の回転速度である高速回転と、それよりも低速である低速回転の 2 種類の回転を行うことができるようになっている。ディスク回転指示プログラム 124 は、図 12 のステップ 2000 に例示するように、次段階のスキャンでアクセスされる DB スキーマを記憶した論理ユニットに対しては、高速回転を実行するためのディスク回転開始命令を発行し、次々段階のスキャンでアクセスされる DB スキーマを記憶した論理ユニットに対しては、低速回転を実行するためのディスク回転開始命令を発行し、次々段階よりも後段のスキャンでアクセスされる DB スキーマを記憶した論理ユニットに対しては、ディスク回転開始命令を発行しない（このステップ 2000 は、例えば、図 9 のステップ 914 或いは 918 で採用されてもよい）。この場合、次段階のスキャンでアクセスされるディスク装置は、高速回転でアクセス待ちとなり、次々段階のスキャンでアクセスされるディスク装置は、低速回転でアクセス待ちとなり、次々段階よりも後段のスキャンでアクセスされるディスク装置は、停止状態となる。これにより、スキャンが次段階が次々段階に切り替わった場合に、ディスク装置 152 の回転が安定するまでの時間長を短くすることができ、  
10  
20  
30  
40  
50  
60  
70  
80  
90  
100  
110  
120  
130  
140  
150  
160  
170  
180  
190  
200  
210  
220  
230  
240  
250  
260  
270  
280  
290  
300  
310  
320  
330  
340  
350  
360  
370  
380  
390  
400  
410  
420  
430  
440  
450  
460  
470  
480  
490  
500  
510  
520  
530  
540  
550  
560  
570  
580  
590  
600  
610  
620  
630  
640  
650  
660  
670  
680  
690  
700  
710  
720  
730  
740  
750  
760  
770  
780  
790  
800  
810  
820  
830  
840  
850  
860  
870  
880  
890  
900  
910  
920  
930  
940  
950  
960  
970  
980  
990  
1000  
1010  
1020  
1030  
1040  
1050  
1060  
1070  
1080  
1090  
1100  
1110  
1120  
1130  
1140  
1150  
1160  
1170  
1180  
1190  
1200  
1210  
1220  
1230  
1240  
1250  
1260  
1270  
1280  
1290  
1300  
1310  
1320  
1330  
1340  
1350  
1360  
1370  
1380  
1390  
1400  
1410  
1420  
1430  
1440  
1450  
1460  
1470  
1480  
1490  
1500  
1510  
1520  
1530  
1540  
1550  
1560  
1570  
1580  
1590  
1600  
1610  
1620  
1630  
1640  
1650  
1660  
1670  
1680  
1690  
1700  
1710  
1720  
1730  
1740  
1750  
1760  
1770  
1780  
1790  
1800  
1810  
1820  
1830  
1840  
1850  
1860  
1870  
1880  
1890  
1900  
1910  
1920  
1930  
1940  
1950  
1960  
1970  
1980  
1990  
2000  
2010  
2020  
2030  
2040  
2050  
2060  
2070  
2080  
2090  
2100  
2110  
2120  
2130  
2140  
2150  
2160  
2170  
2180  
2190  
2200  
2210  
2220  
2230  
2240  
2250  
2260  
2270  
2280  
2290  
2300  
2310  
2320  
2330  
2340  
2350  
2360  
2370  
2380  
2390  
2400  
2410  
2420  
2430  
2440  
2450  
2460  
2470  
2480  
2490  
2500  
2510  
2520  
2530  
2540  
2550  
2560  
2570  
2580  
2590  
2600  
2610  
2620  
2630  
2640  
2650  
2660  
2670  
2680  
2690  
2700  
2710  
2720  
2730  
2740  
2750  
2760  
2770  
2780  
2790  
2800  
2810  
2820  
2830  
2840  
2850  
2860  
2870  
2880  
2890  
2900  
2910  
2920  
2930  
2940  
2950  
2960  
2970  
2980  
2990  
3000  
3010  
3020  
3030  
3040  
3050  
3060  
3070  
3080  
3090  
3100  
3110  
3120  
3130  
3140  
3150  
3160  
3170  
3180  
3190  
3200  
3210  
3220  
3230  
3240  
3250  
3260  
3270  
3280  
3290  
3300  
3310  
3320  
3330  
3340  
3350  
3360  
3370  
3380  
3390  
3400  
3410  
3420  
3430  
3440  
3450  
3460  
3470  
3480  
3490  
3500  
3510  
3520  
3530  
3540  
3550  
3560  
3570  
3580  
3590  
3600  
3610  
3620  
3630  
3640  
3650  
3660  
3670  
3680  
3690  
3700  
3710  
3720  
3730  
3740  
3750  
3760  
3770  
3780  
3790  
3800  
3810  
3820  
3830  
3840  
3850  
3860  
3870  
3880  
3890  
3900  
3910  
3920  
3930  
3940  
3950  
3960  
3970  
3980  
3990  
4000  
4010  
4020  
4030  
4040  
4050  
4060  
4070  
4080  
4090  
4100  
4110  
4120  
4130  
4140  
4150  
4160  
4170  
4180  
4190  
4200  
4210  
4220  
4230  
4240  
4250  
4260  
4270  
4280  
4290  
4300  
4310  
4320  
4330  
4340  
4350  
4360  
4370  
4380  
4390  
4400  
4410  
4420  
4430  
4440  
4450  
4460  
4470  
4480  
4490  
4500  
4510  
4520  
4530  
4540  
4550  
4560  
4570  
4580  
4590  
4600  
4610  
4620  
4630  
4640  
4650  
4660  
4670  
4680  
4690  
4700  
4710  
4720  
4730  
4740  
4750  
4760  
4770  
4780  
4790  
4800  
4810  
4820  
4830  
4840  
4850  
4860  
4870  
4880  
4890  
4900  
4910  
4920  
4930  
4940  
4950  
4960  
4970  
4980  
4990  
5000  
5010  
5020  
5030  
5040  
5050  
5060  
5070  
5080  
5090  
5100  
5110  
5120  
5130  
5140  
5150  
5160  
5170  
5180  
5190  
5200  
5210  
5220  
5230  
5240  
5250  
5260  
5270  
5280  
5290  
5300  
5310  
5320  
5330  
5340  
5350  
5360  
5370  
5380  
5390  
5400  
5410  
5420  
5430  
5440  
5450  
5460  
5470  
5480  
5490  
5500  
5510  
5520  
5530  
5540  
5550  
5560  
5570  
5580  
5590  
5600  
5610  
5620  
5630  
5640  
5650  
5660  
5670  
5680  
5690  
5700  
5710  
5720  
5730  
5740  
5750  
5760  
5770  
5780  
5790  
5800  
5810  
5820  
5830  
5840  
5850  
5860  
5870  
5880  
5890  
5900  
5910  
5920  
5930  
5940  
5950  
5960  
5970  
5980  
5990  
6000  
6010  
6020  
6030  
6040  
6050  
6060  
6070  
6080  
6090  
6100  
6110  
6120  
6130  
6140  
6150  
6160  
6170  
6180  
6190  
6200  
6210  
6220  
6230  
6240  
6250  
6260  
6270  
6280  
6290  
6300  
6310  
6320  
6330  
6340  
6350  
6360  
6370  
6380  
6390  
6400  
6410  
6420  
6430  
6440  
6450  
6460  
6470  
6480  
6490  
6500  
6510  
6520  
6530  
6540  
6550  
6560  
6570  
6580  
6590  
6600  
6610  
6620  
6630  
6640  
6650  
6660  
6670  
6680  
6690  
6700  
6710  
6720  
6730  
6740  
6750  
6760  
6770  
6780  
6790  
6800  
6810  
6820  
6830  
6840  
6850  
6860  
6870  
6880  
6890  
6900  
6910  
6920  
6930  
6940  
6950  
6960  
6970  
6980  
6990  
7000  
7010  
7020  
7030  
7040  
7050  
7060  
7070  
7080  
7090  
7100  
7110  
7120  
7130  
7140  
7150  
7160  
7170  
7180  
7190  
7200  
7210  
7220  
7230  
7240  
7250  
7260  
7270  
7280  
7290  
7300  
7310  
7320  
7330  
7340  
7350  
7360  
7370  
7380  
7390  
7400  
7410  
7420  
7430  
7440  
7450  
7460  
7470  
7480  
7490  
7500  
7510  
7520  
7530  
7540  
7550  
7560  
7570  
7580  
7590  
7600  
7610  
7620  
7630  
7640  
7650  
7660  
7670  
7680  
7690  
7700  
7710  
7720  
7730  
7740  
7750  
7760  
7770  
7780  
7790  
7800  
7810  
7820  
7830  
7840  
7850  
7860  
7870  
7880  
7890  
7900  
7910  
7920  
7930  
7940  
7950  
7960  
7970  
7980  
7990  
8000  
8010  
8020  
8030  
8040  
8050  
8060  
8070  
8080  
8090  
8100  
8110  
8120  
8130  
8140  
8150  
8160  
8170  
8180  
8190  
8200  
8210  
8220  
8230  
8240  
8250  
8260  
8270  
8280  
8290  
8300  
8310  
8320  
8330  
8340  
8350  
8360  
8370  
8380  
8390  
8400  
8410  
8420  
8430  
8440  
8450  
8460  
8470  
8480  
8490  
8500  
8510  
8520  
8530  
8540  
8550  
8560  
8570  
8580  
8590  
8600  
8610  
8620  
8630  
8640  
8650  
8660  
8670  
8680  
8690  
8700  
8710  
8720  
8730  
8740  
8750  
8760  
8770  
8780  
8790  
8800  
8810  
8820  
8830  
8840  
8850  
8860  
8870  
8880  
8890  
8900  
8910  
8920  
8930  
8940  
8950  
8960  
8970  
8980  
8990  
9000  
9010  
9020  
9030  
9040  
9050  
9060  
9070  
9080  
9090  
9100  
9110  
9120  
9130  
9140  
9150  
9160  
9170  
9180  
9190  
9200  
9210  
9220  
9230  
9240  
9250  
9260  
9270  
9280  
9290  
9300  
9310  
9320  
9330  
9340  
9350  
9360  
9370  
9380  
9390  
9400  
9410  
9420  
9430  
9440  
9450  
9460  
9470  
9480  
9490  
9500  
9510  
9520  
9530  
9540  
9550  
9560  
9570  
9580  
9590  
9600  
9610  
9620  
9630  
9640  
9650  
9660  
9670  
9680  
9690  
9700  
9710  
9720  
9730  
9740  
9750  
9760  
9770  
9780  
9790  
9800  
9810  
9820  
9830  
9840  
9850  
9860  
9870  
9880  
9890  
9900  
9910  
9920  
9930  
9940  
9950  
9960  
9970  
9980  
9990  
10000

【0110】

第二の変形例では、スキャン種別は、テーブルスキャンが索引スキャンであるかに代えて、全検索か一部検索などの他種の種別を採用することができる。或いは、スキーマサイズ及びスキャン種別に代えて、他種のアクセスサイズを採用することができる。要するに、第二の変形例では、ディスク回転指示プログラム 124 は、アクセス対象のアクセスサイズに基づいて、次段階のスキャンでアクセスされる論理ユニットに対するディスク回転開始命令の発行タイミングを制御することができる。例えば、アクセスサイズと 1 秒当たりの平均データ転送量と進捗状態とから、一つのスキャン処理が完了するまでの残り時間長を推測し、その残り時間長が、ディスク回転開始時間よりも短ければ、次段階のスキャンでアクセスされる論理ユニットに対するディスク回転開始命令が発行されても良い。

【0111】

第三の変形例では、一つのクエリの DB 処理が終了した場合にディスク回転停止命令 712 が発行されるのではなく、一つのクエリの DB 処理中に、適時に、ディスク回転停止命令 712 が発行される。具体的には、例えば、ディスク回転指示プログラム 124 は、一つのスキャン処理が終了した場合、そのスキャン処理でアクセスされた論理ユニットが、次段階或いは次々段階よりも後のスキャン処理までアクセスされることが無いことを、DB 処理情報 520 から特定できた場合には、DB 処理中であっても、その論理ユニットに対するディスク回転停止命令を発行しても良い。

10

20

30

40

50

## 【 0 1 1 2 】

第四の変形例では、ディスク装置 1 5 2 が、複数種類の回転速度で回転できる場合には、それら複数種類の回転速度を基に、ディスク情報 1 2 8 に書かれるディスク回転開始時間が用意されても良い。具体的には、例えば、停止状態から高速回転になるまでのディスク回転開始時間と、低速回転から高速回転になるまでのディスク回転開始時間とが用意されても良い。この場合、ディスク回転指示プログラム 1 2 4 は、低速回転から高速回転にするのか、或いは、停止状態から高速回転するのかに応じて、参照するディスク回転開始時間の種類を違えても良い。

## 【 0 1 1 3 】

以上、本発明の実施形態及び幾つかの変形例を説明したが、これらの実施形態及び変形例は本発明の説明のための例示にすぎず、本発明の範囲をそれらの実施形態及び変形例にのみ限定する趣旨ではない。本発明は、その要旨を逸脱することなく、その他の様々な態様でも実施することができる。例えば、ディスク装置は基本的に停止状態或いは低速回転状態となるようになっていて、ディスク回転開始プログラム 1 2 4 が、必要に応じて回転指示させてもよい。具体的には、例えば、ディスク回転指示プログラム 1 2 4 からのディスク回転指示命令に従ってディスクの回転を停止することに代えて、ストレージシステム 1 4 0 が、各ディスク装置 1 5 2 を、ディスク回転開始後一定時間経ったらディスクの回転を停止するようになっていて、ディスク回転指示プログラム 1 2 4 が、ディスクの回転を停止してもらいたくない場合には、ディスク回転開始命令を送信し、そのディスクに対応したカウント値を増やすことで、ディスクの回転を継続させても良い。

## 【 図面の簡単な説明 】

## 【 0 1 1 4 】

【 図 1 】 図 1 は、本発明の一実施形態に係る DB システムの構成例を示した図である。

【 図 2 】 図 2 ( a ) は、記憶領域管理情報 1 5 6 の構成例を示した図である。図 2 ( b ) は、ディスク回転管理情報の構成例を示した図である。

【 図 3 】 図 3 は、ローデバイス情報 1 1 6 の構成例を示した図である。

【 図 4 】 図 4 ( a ) は、DBMS 管理情報 1 2 0 の構成例を示した図である。図 4 ( b ) は、DB システムファイル情報 4 0 0 の構成例を示した図である。図 4 ( c ) は、データ領域情報 4 2 0 の構成例を示した図である。図 4 ( d ) は、DB スキーマ情報 4 4 0 の構成例を示した図である。

【 図 5 】 図 5 ( a ) は、DB 処理管理情報 1 2 6 の構成例を示した図である。図 5 ( b ) は、DB スキーマ位置情報 5 0 0 の構成例を示した図である。図 5 ( c ) は、DB 処理情報 5 2 0 の構成例を示した図である。図 5 ( d ) は、ディスク情報 1 2 8 の構成例を示した図である。

【 図 6 】 図 6 ( a ) は、DB 処理実行情報 6 0 0 の構成例を示した図である。図 6 ( b ) は、DB 処理終了情報の構成例を示した図である。

【 図 7 】 図 7 ( a ) は、ディスク回転開始命令 7 0 0 の構成例を示した図である。図 7 ( b ) は、ディスク回転停止命令 7 1 0 の構成例を示した図である。

【 図 8 】 図 8 は、DB 情報取得送信処理の手順例を示したフロー図である。

【 図 9 】 図 9 は、ディスク回転指示処理の手順例を示したフロー図である。

【 図 1 0 】 図 1 0 は、ディスク回転制御処理の手順例を示したフロー図である。

【 図 1 1 】 図 1 1 は、クエリプランの構成例を示した図である。

【 図 1 2 】 図 1 2 は、ディスク装置の回転速度種類に基づくディスク回転開始命令の発行タイミングの一例を示す。

## 【 符号の説明 】

## 【 0 1 1 5 】

1 0 0 ... DB サーバ、 1 3 0 ... 通信ネットワーク、 1 4 0 ... ストレージシステム、 1 1 4 ... OS、 1 1 8 ... DBMS、 1 2 2 ... DB 情報取得プログラム、 1 2 4 ... ディスク回転指示プログラム

10

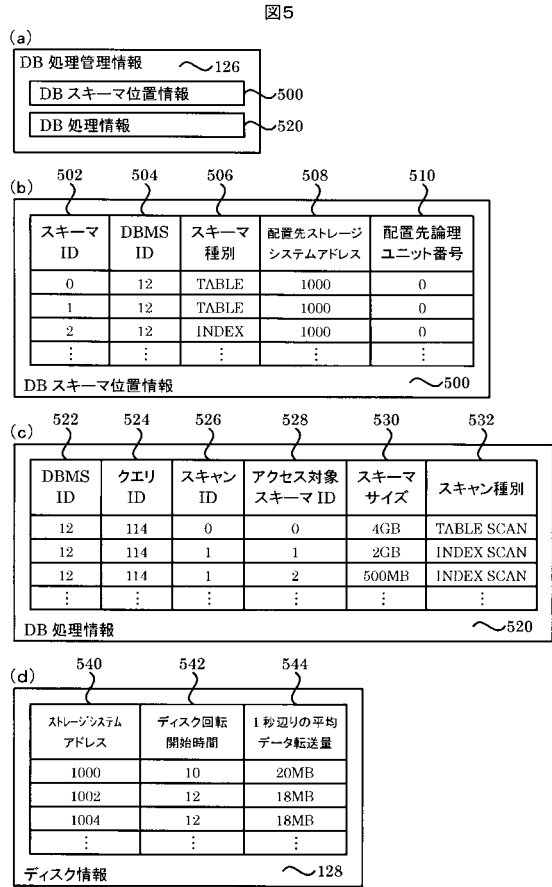
20

30

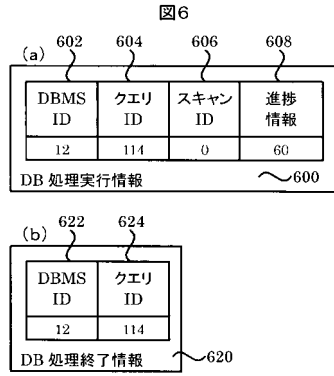
40



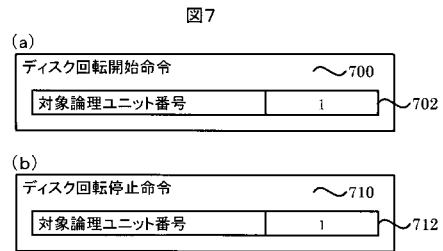
【図5】



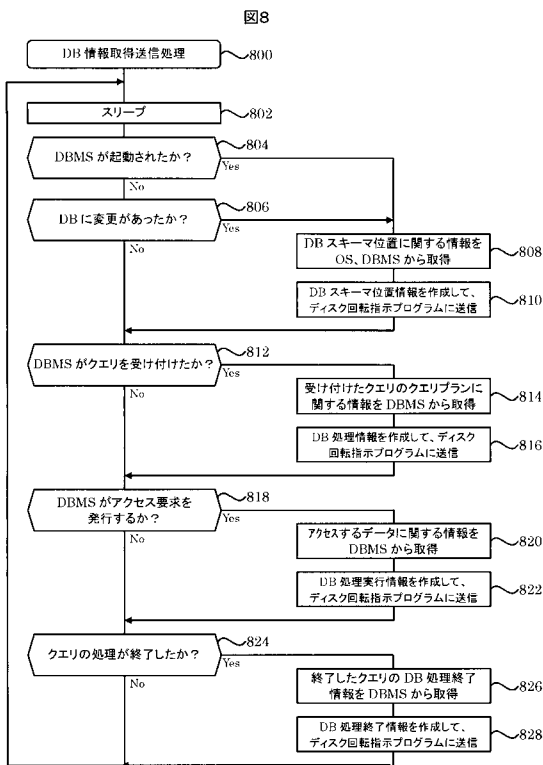
【図6】



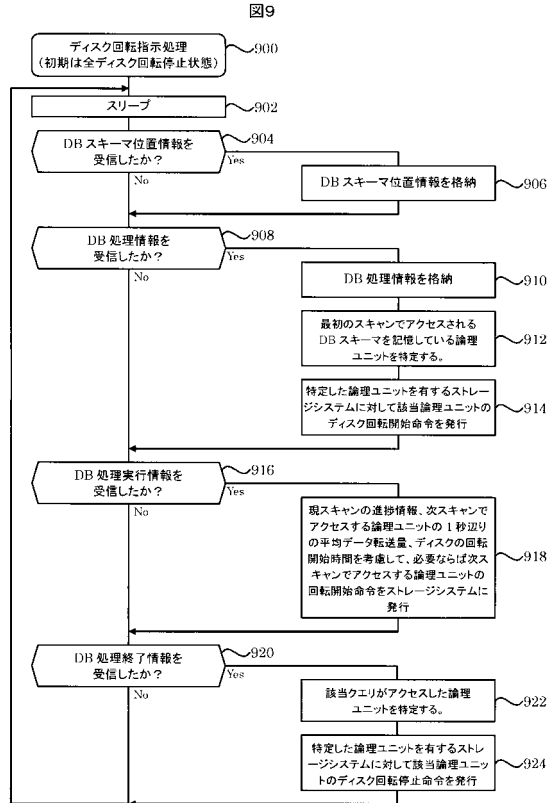
【図7】



【図8】

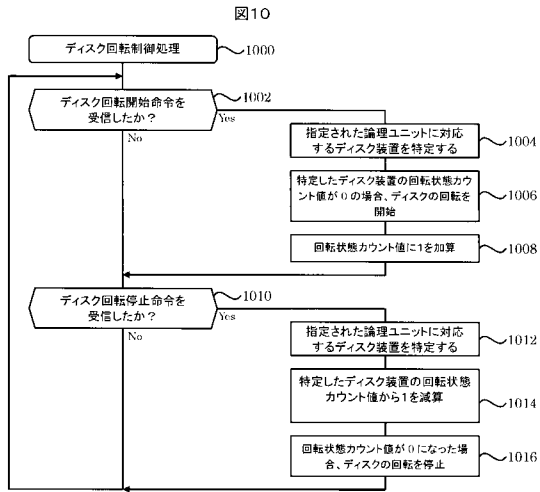


【図9】

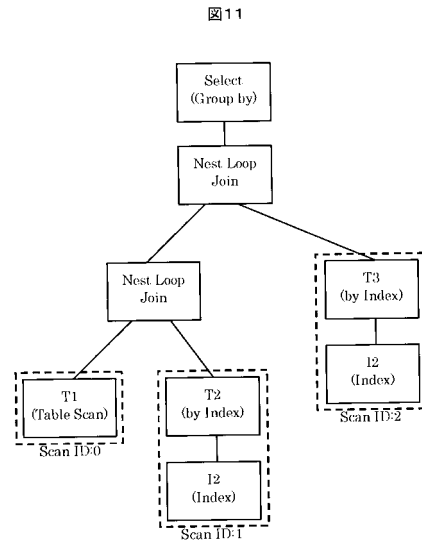




【 図 10 】

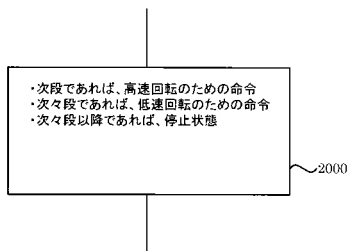


【 図 11 】



【 図 12 】

図12



---

フロントページの続き

(72)発明者 茂木 和彦

神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

審査官 坂東 博司

(56)参考文献 特開平 0 9 - 2 8 2 0 5 7 ( J P , A )  
特開平 0 9 - 3 0 5 2 7 0 ( J P , A )  
特開 2 0 0 5 - 2 5 8 7 3 5 ( J P , A )  
特開 2 0 0 3 - 1 8 7 5 1 7 ( J P , A )  
特開 2 0 0 5 - 1 5 7 7 1 0 ( J P , A )

(58)調査した分野(Int.Cl. , DB名)

G 1 1 B 1 9 / 0 0  
G 0 6 F 3 / 0 6