

(19) 日本国特許庁(JP)

再公表特許(A1)

(11) 国際公開番号

W02013/001613

発行日 平成27年2月23日 (2015. 2. 23)

(43) 国際公開日 平成25年1月3日 (2013. 1. 3)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 13/372 (2006.01)	G06F 13/372 B	5B061
G06F 3/06 (2006.01)	G06F 3/06 301A	
G06F 13/10 (2006.01)	G06F 13/10 340A	
G06F 12/00 (2006.01)	G06F 12/00 535Z	
G06F 9/50 (2006.01)	G06F 9/46 465C	

審査請求 有 予備審査請求 未請求 (全 28 頁)

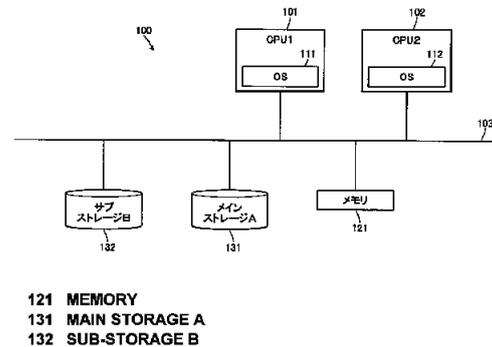
出願番号 特願2013-522397 (P2013-522397)	(71) 出願人 000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(21) 国際出願番号 PCT/JP2011/064841	(74) 代理人 100104190 弁理士 酒井 昭徳
(22) 国際出願日 平成23年6月28日 (2011. 6. 28)	(72) 発明者 栗原 康志 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
(81) 指定国 AP (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW	(72) 発明者 山下 浩一郎 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 鈴木 貴久 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

最終頁に続く

(54) 【発明の名称】 スケジューリング方法およびシステム

(57) 【要約】

システムのスケジューラは、第1CPU(101)によりタスクが第1タスク種類に属するか否かを判定する。タスクがメモリーリードタスクの第1タスク種類に属するときは、タスクがアクセスする第1アクセス領域がメインストレージ(131)またはサブストレージ(132)にあるか否かを判定する。この判定結果に基づいて、タスクがアクセスするストレージ先をメインストレージ(131)またはサブストレージ(132)に設定する。タスクが第1タスク種類に属さないときは他のタスク種類に属するか否かを判定し、複数のCPU(101, 102)が共有するメインストレージ(131)へのアクセス競合を回避する。



【特許請求の範囲】

【請求項 1】

第 1 CPU は、タスクが第 1 タスク種類に属するか否かを判定し、
前記タスクが前記第 1 タスク種類に属するときは前記タスクがアクセスする第 1 アクセス領域が第 1 メモリまたは第 2 メモリにあるか否かを判定し、
前記判定結果に基づいて前記タスクがアクセスするメモリを前記第 1 メモリまたは前記第 2 メモリに設定すること
を特徴とするスケジューリング方法。

【請求項 2】

前記第 1 タスク種類はメモリリードタスクであること
を特徴とする請求項 1 に記載のスケジューリング方法。

10

【請求項 3】

前記タスクが前記第 1 タスク種類に属さないときは前記タスクが第 2 タスク種類に属するか否かを判定すること
を特徴とする請求項 1 または請求項 2 に記載のスケジューリング方法。

【請求項 4】

前記第 2 タスク種類はメモリリードライトタスクであること
を特徴とする請求項 3 に記載のスケジューリング方法。

【請求項 5】

前記タスクが第 3 タスク種類に含まれるときは、前記タスクが第 1 メモリにアクセスする第 1 アクセス時間と前記タスクが前記第 2 メモリにアクセスする第 2 アクセス時間との比較結果に基づいて、前記タスクがアクセスするメモリを前記第 1 メモリまたは前記第 2 メモリに設定すること
を特徴とする請求項 1 乃至請求項 4 の何れか一に記載のスケジューリング方法。

20

【請求項 6】

前記第 1 アクセス領域が第 3 タスク種類に属するタスクがアクセスする第 2 アクセス領域に含まれるか否かを判定し、
前記判定結果に基づいて前記第 1 アクセス領域が前記第 1 メモリまたは前記第 2 メモリにあるか否かを判定すること
を特徴とする請求項 1 乃至請求項 4 の何れか一に記載のスケジューリング方法。

30

【請求項 7】

前記第 3 タスク種類はメモリライトタスクであること
を特徴とする請求項 5 または請求項 6 に記載のスケジューリング方法。

【請求項 8】

前記第 1 アクセス領域が前記第 2 アクセス領域に含まれるときは、前記第 2 アクセス領域の前記第 2 メモリから前記第 1 メモリへの書き込みの終了に基づいて前記タスクがアクセスするメモリを前記第 1 メモリまたは前記第 2 メモリに設定すること
を特徴とする請求項 6 または請求項 7 に記載のスケジューリング方法。

【請求項 9】

前記タスクがアクセスするメモリが第 2 メモリに設定されるとき、前記タスクがアクセスした前記第 2 メモリのメモリ領域が前記第 1 メモリに書き込まれること
を特徴とする請求項 1 乃至請求項 8 の何れか一に記載のスケジューリング方法。

40

【請求項 10】

前記タスクが第 1 タスク種類グループに含まれるときは、前記第 1 メモリにアクセスする他のタスクが割り当てられている第 2 CPU に前記第 1 タスク種類グループに含まれるタスクを割り当てること
を特徴とする請求項 1 乃至請求項 9 の何れか一に記載のスケジューリング方法。

【請求項 11】

第 2 CPU に割り当てられるタスクが第 2 タスク種類に属さないときであって前記第 2 CPU で実行されている他のタスクが前記第 2 タスク種類に属するとき、前記他のタスク

50

は前記第2CPUのタスクキューに戻されること

を特徴とする請求項1乃至請求項9の何れかーに記載のスケジューリング方法。

【請求項12】

第2CPUに割り当てられるタスクが第2タスク種類に属するときであって前記第2CPUを含む複数のCPUの少なくとも一のCPUで前記第2タスク種類グループに含まれるタスクが実行されていないとき、前記第2タスク種類に属するつぎのタスクの実行を停止すること

を特徴とする請求項1乃至請求項9の何れかーに記載のスケジューリング方法。

【請求項13】

複数のCPUと複数のメモリを含み、

前記複数のCPUに含まれる第1CPUは、

タスクの種類を判定し、

前記タスクが第1タスク種類に属するときは前記タスクがアクセスする第1アクセス領域が前記複数のメモリのうち、第1メモリまたは第2メモリにあるか否かを判定し、

前記判定結果に基づいて前記タスクがアクセスするメモリを前記第1メモリまたは前記第2メモリに設定すること

を特徴とするシステム。

【請求項14】

前記第1CPUは、

前記複数のメモリのうち同一アクセス先のメモリとなる前記タスクを、前記複数のCPUに含まれる同一のCPUに割り当てることを特徴とする請求項13に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数のCPUが共有するメモリに対するアクセス競合を回避するスケジューリング方法およびシステムに関する。

【背景技術】

【0002】

複数のストレージおよび複数のCPUを備えるマルチコアシステムでは、一つのストレージに対して同時に複数のCPUからのストレージアクセスが生じる可能性がある。これを回避する対策としては、複数のストレージアクセスを切り替えながら処理することが考えられる。

【0003】

たとえば、ディスクドライブに対するアクセス・タスクをグループ化し、タスクのリストにしたがってスレッドを直列に実行することにより、ディスクに対するシークの競合を防止した技術がある（たとえば、下記特許文献1参照。）。また、制限時間つきストレージアクセス要求を対象にし、対象ストレージへの書き込みが制限時間内に終了しない場合は、データを別のストレージへ一旦書き込みをおこない、後ほど本来のストレージへデータを移動させる技術がある（たとえば、下記特許文献2参照。）。

【0004】

また、画像データ検索システムにおいて、集合型光ディスク装置と単盤型光ディスク装置を接続し、画像データの登録を単盤型光ディスク装置に仮登録した後に、集合型光ディスク装置に転送することにより、画像データの登録要求と検索要求とが同時に発生しても画像データへのアクセス競合を防ぐ技術がある（たとえば、下記特許文献3参照。）。また、マルチプロセッサにおいて、HDDへのアクセス要求の種別を判定し、種別毎に異なるアドレス空間を設定し、アクセス要求を制御する調停論理手段を設け、競合したアクセス発生時に、いずれかの処理を先におこなう技術がある（たとえば、下記特許文献4参照。）。

【先行技術文献】

【特許文献】

10

20

30

40

50

【 0 0 0 5 】

【特許文献 1】特開平 7 - 1 0 4 9 3 9 号公報

【特許文献 2】特開平 1 0 - 9 1 3 5 7 号公報

【特許文献 3】特開平 3 - 0 1 8 9 7 6 号公報

【特許文献 4】特開 2 0 0 9 - 1 8 7 3 2 7 号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 6 】

しかしながら、特許文献 1 ~ 4 に記載の技術では、ストレージへのアクセススケジューリングと、OS のタスクスケジューリングがそれぞれ個別におこなわれることになるため、ストレージアクセスをおこなう際に複数の CPU 間で同一のストレージに対するアクセス競合が発生することが考えられる。このアクセス競合の発生により、ストレージのアクセス性能が低下する。特許文献 2 に記載の技術では、ストレージへのアクセススケジューリングしか考慮されておらず、ストレージへのアクセススケジューリングと、OS のタスクスケジューリングとの連携がなされないために、複数の CPU でそれぞれタスクを実行するときの一つのストレージへのアクセス競合が発生することがあった。

10

【 0 0 0 7 】

開示のスケジューリング方法およびシステムは、上述した問題点を解消するものであり、複数の CPU が共有するメモリへのアクセス競合を回避することを目的とする。

【課題を解決するための手段】

20

【 0 0 0 8 】

上述した課題を解決し、目的を達成するため、開示技術は、第 1 CPU は、タスクが第 1 タスク種類に属するか否かを判定し、前記タスクが前記第 1 タスク種類に属するときは前記タスクがアクセスする第 1 アクセス領域が第 1 メモリまたは第 2 メモリにあるか否かを判定し、前記判定結果に基づいて前記タスクがアクセスするメモリを前記第 1 メモリまたは前記第 2 メモリに設定することを特徴とする。

【発明の効果】

【 0 0 0 9 】

開示のスケジューリング方法およびシステムによれば、複数の CPU が共有するメモリへのアクセス競合を回避できるという効果を奏する。

30

【図面の簡単な説明】

【 0 0 1 0 】

【図 1】図 1 は、実施の形態のスケジューリング方法を実行するシステムの構成例を示す図である。

【図 2】図 2 は、システムの内部構成を示すブロック図である。

【図 3 - 1】図 3 - 1 は、データベースの情報を示す図表である（その 1）。

【図 3 - 2】図 3 - 2 は、データベースの情報を示す図表である（その 2）。

【図 3 - 3】図 3 - 3 は、データベースの情報を示す図表である（その 3）。

【図 4 - 1】図 4 - 1 は、アクセス競合の回避の処理概要を示す図である（その 1）。

【図 4 - 2】図 4 - 2 は、アクセス競合の回避の処理概要を示す図である（その 2）。

40

【図 5 - 1】図 5 - 1 は、ストレージに対するアクセススケジューリングの例を示す図である（その 1）。

【図 5 - 2】図 5 - 2 は、ストレージに対するアクセススケジューリングの例を示す図である（その 2）。

【図 5 - 3】図 5 - 3 は、ストレージに対するアクセススケジューリングの例を示す図である（その 3）。

【図 5 - 4】図 5 - 4 は、ストレージに対するアクセススケジューリングの例を示す図である（その 4）。

【図 5 - 5】図 5 - 5 は、ストレージに対するアクセススケジューリングの例を示す図である（その 5）。

50

【図 6 - 1】図 6 - 1 は、ストレージアクセスとタスクアクセスを連動させたスケジューリングの例を示す図である（その 1）。

【図 6 - 2】図 6 - 2 は、ストレージアクセスとタスクアクセスを連動させたスケジューリングの例を示す図である（その 2）。

【図 6 - 3】図 6 - 3 は、ストレージアクセスとタスクアクセスを連動させたスケジューリングの例を示す図である（その 3）。

【図 6 - 4】図 6 - 4 は、ストレージアクセスとタスクアクセスを連動させたスケジューリングの例を示す図である（その 4）。

【図 6 - 5】図 6 - 5 は、ストレージアクセスとタスクアクセスを連動させたスケジューリングの例を示す図である（その 5）。

10

【図 7】図 7 は、書き戻し前のデータへのアクセスの状態を示す図である。

【図 8 - 1】図 8 - 1 は、データの書き戻しにかかる処理を示す図である（その 1）。

【図 8 - 2】図 8 - 2 は、データの書き戻しにかかる処理を示す図である（その 2）。

【図 8 - 3】図 8 - 3 は、データの書き戻しにかかる処理を示す図である（その 3）。

【図 9】図 9 は、ストレージスケジューラがおこなう処理手順を示すフローチャートである。

【図 10】図 10 は、タスクスケジューラがおこなう処理手順を示すフローチャートである。

【図 11】図 11 は、メインスケジューラ、スレーブスケジューラがおこなう処理手順を示すフローチャートである。

20

【図 12】図 12 は、アクセス監視部がおこなう処理手順を示すフローチャートである。

【図 13】図 13 は、実施の形態による競合の回避および全アクセスの処理時間を説明する図である。

【図 14】図 14 は、図 1 に示したコンピュータを用いたシステムの適用例を示す図である。

【発明を実施するための形態】

【0011】

以下に添付図面を参照して、開示技術の好適な実施の形態を詳細に説明する。開示技術は、複数の CPU と複数のメモリを備えてタスクを並列処理するシステム等に適用される。そして、メモリアccessのスケジューリングとタスクスケジューリングとを連携させることにより、タスク実行により、複数のメモリのうち共有される一つのメモリに対するアクセス競合を回避し、メモリへのアクセス処理の効率化を図る。

30

【0012】

（実施の形態）

図 1 は、実施の形態のスケジューリング方法を実行するシステムの構成例を示す図である。システム 100 は、複数の CPU（第 1 CPU 101、第 2 CPU 102）を含み、CPU 101、102 は、それぞれ OS 111、112 を含む。これら CPU 101、102 は、バス 103 を介してメモリ 121 と、複数の CPU によって共有されるメモリにアクセスする。以下の説明では、図 1 に示す 1 台のコンピュータに複数の CPU 101、102 を搭載したマルチコアの構成例を用いて説明をおこなう。

40

【0013】

図 1 に示す例では、複数の CPU によって共有される複数のメモリを有する。この複数のメモリとして、第 1 メモリとしてのメインストレージ（A）131 と、第 2 メモリとしてのサブストレージ（B）132 を含む。たとえば、これらメインストレージ（A）131 と、サブストレージ（B）132 は、ハードディスク装置を用いることができる。以下の実施例では、CPU 101、102 は、いずれも主にメインストレージ 131 にアクセスし、共有使用するものとしている。そして、メインストレージ 131 に対するアクセス競合回避のために、一時的にサブストレージ 132 を利用するものとして説明する。

【0014】

図 2 は、システムの内部構成を示すブロック図である。マスタ側の OS 111 は、マス

50

タスケジューラ 201 と、アクセス監視部 202 と、ウェイトキュー 203 と、タスクキュー 204 とを含む。マスタスケジューラ 201 は、ストレージスケジューラ 205 と、タスクスケジューラ 206 と、タスクディスパッチ部 207 とを含む。

【0015】

スレーブ側の OS 112 は、スレーブスケジューラ 211 と、アクセス監視部 212 と、タスクキュー 214 とを含む。スレーブスケジューラ 211 は、タスクディスパッチ部 217 を含む。メモリ 121 は、データベース (DB) 221 を含む。

【0016】

マスタ OS 111 のストレージスケジューラ 205 は、ウェイトキュー 203 からタスクを取り出し、取り出したタスクのアクセス先ストレージ (メインストレージ 131 あるいはサブストレージ 132) を決定する。この際、各タスクの特性 (Read あるいは Write、帯域利用、処理時間等) により、アクセス先ストレージを決定し、タスクスケジューラ 206 に通知する。

【0017】

マスタ OS 111 のタスクスケジューラ 206 は、ストレージスケジューラ 205 によるストレージアクセスのスケジューリング情報と、タスクのスケジューリング情報に基づき、タスクの割当先を決定し、タスクキュー 204 に挿入する。このタスクスケジューラ 206 は、ストレージのアクセス先が同じであるタスク同士を同一の CPU 101, 102 に割り当てる。タスクディスパッチ部 207 は、CPU 101 のディスパッチ状況を基に、タスクキュー 204 に挿入されているタスクのタスクディスパッチ、タスクスイッチを制御する。

【0018】

マスタ OS 111 のアクセス監視部 202 は、ストレージに対するリード、ライトの発生時に、実行中タスクの情報に基づき、タスクがアクセスするストレージ (メインストレージ 131 あるいはサブストレージ 132) を指定する。

【0019】

スレーブ OS 112 は、マスタ OS 111 の統括制御を受けて CPU 102 の制御を実行する。スレーブ OS 112 のスレーブスケジューラ 211 は、タスクディスパッチ部 217 を含み、このタスクディスパッチ部 217 は、CPU 102 のディスパッチ状況を基に、タスクキュー 214 に挿入されているタスクのタスクディスパッチ、タスクスイッチを制御する。スレーブ OS 112 のアクセス監視部 212 は、ストレージに対するリード、ライトの発生時に、実行中タスクの情報に基づき、タスクがアクセスするストレージ (メインストレージ 131 あるいはサブストレージ 132) を指定する。

【0020】

メモリ 121 のデータベース 221 には、タスクテーブル、ストレージアクセススケジューリング情報、タスクのスケジューリング情報が保持される。

【0021】

図 3 - 1 ~ 図 3 - 3 は、データベースの情報を示す図表である。図 3 - 1 は、データベース 221 に保存されるタスクテーブル 301 を示す図である。タスクテーブル 301 は、ライト (Write) タスクテーブル 311 と、リード (Read) タスクテーブル 312 と、F タスクテーブル 313 とを含む。この F タスクテーブル 313 の情報は、アクセス競合を避けるために Write タスクが一旦別のストレージに書き込んだ場合に参照される。F タスク (書き戻しタスクに相当) は、実際のアクセス対象のストレージに、一時的にデータを書き込んだストレージからデータの書き戻しをおこなうタスクである。この F タスクは、ストレージにアクセスをおこなう他のタスクに影響を与えないタイミングでディスパッチされる。

【0022】

ライトタスクテーブル 311 は、(1) ライトタスク ID と、(2) 指定書き込み領域と、(3) 書き込み指定領域と、(4) 書き戻し判定フラグと、(5) 一時保存先アクセスリードタスクと、(6) データサイズの各情報を含む。

10

20

30

40

50

【 0 0 2 3 】

(1) ライトタスク ID は、ストレージに書き込むライトタスクの識別子 (ID) である。(2) 指定書き込み領域は、書き込みが指定されているストレージのアドレスである。(3) 書き込み指定領域は、一時的に書き込む先のストレージのアドレスである。(4) 書き戻し判定フラグは、指定書き込み領域のアドレスのデータを、書き込み指定領域に書き込みが終了したかを示すフラグ値である。(5) 一時保存先アクセスリードタスクは、指定書き込み領域のアドレスにアクセスしているタスクの ID である。(6) データサイズは、書き込みをおこなうデータのサイズである。

【 0 0 2 4 】

リードタスクテーブル 3 1 2 は、(1) リードタスク ID と、(2) 指定アクセス領域と、(3) 読み出し指定領域と、(4) データサイズの各情報を含む。(1) リードタスク ID は、ストレージから読み出すリードタスクの識別子 (ID) である。(2) 指定アクセス領域は、読み出しを指定されたストレージのアドレスである。(3) 読み出し指定領域は、実際に読み出しをおこなうストレージのアドレスである。(4) データサイズは、読み出しをおこなうデータのサイズである。

10

【 0 0 2 5 】

F タスクテーブル 3 1 3 は、(1) F タスク ID と、(2) 書き戻しライト (W r i t e) タスク ID とを含む。(1) F タスク ID は、F タスクの ID である。(2) 書き戻しライトタスク ID は、一時的に書き込みをおこなったデータを実際の指定先のストレージに書き戻しをおこなう対象ライト (W r i t e) タスク ID である。

20

【 0 0 2 6 】

図 3 - 2 は、データベース 2 2 1 に保存されるストレージアクセススケジューリングテーブル 3 2 1 と、タスクスケジューリングテーブル 3 2 2 とを示す図である。ストレージアクセススケジューリングテーブル 3 2 1 は、ストレージ ID と、各ストレージ ID 毎の割り当てタスク ID、および終了予測時間とを含む。ストレージ ID は、上記メインストレージ 1 3 1、およびサブストレージ 1 3 2 を識別する ID である。割り当てタスク ID は、各ストレージにアクセスするタスク ID である。終了予測時間は、後述するアクセス終了時間予測式により求められたタスクの終了予測時間である。

【 0 0 2 7 】

タスクスケジューリングテーブル 3 2 2 は、CPU ID と、CPU ID 毎の割り当てタスク ID とを含む。CPU ID は、複数の CPU 1 0 1、1 0 2 の識別子 (ID) である。割り当てタスク ID は、複数の CPU 1 0 1、1 0 2 毎に割り当てられているタスクの識別子 (ID) である。

30

【 0 0 2 8 】

図 3 - 3 は、データベース 2 2 1 に保存されるストレージアクセスをおこなうタスクの分類を示す図表である。図示の分類表 3 3 1 に示されるように、タスクは、ストレージへのアクセスタイプと、帯域の利用の仕方、処理時間等により複数 (図示の例では A ~ F) に分類される。

【 0 0 2 9 】

分類 A のタスクは、アクセスタイプがリード、帯域利用が一定帯域であり、ビデオ等のアプリケーション処理時のデータ等である。分類 B のタスクは、アクセスタイプがリード、帯域利用が全帯域であり、アップロードするデータ等である。分類 C のタスクは、アクセスタイプがライト、帯域利用全帯域であり、ダウンロードしたデータ等である。分類 D のタスクは、アクセスタイプがライト、帯域利用が最優先、処理時間が一瞬である、スワップアウトのデータ等である。分類 E のタスクは、アクセスタイプがリード、帯域利用が最優先、処理時間が一瞬である、スワップインのデータ等である。分類 F のタスク (F タスク) は、上記のように、サブストレージ 1 3 2 からメインストレージ 1 3 1 への書き戻しをおこなう処理にかかるタスクであり、アクセスタイプがリードおよびライト、帯域利用が全帯域である。

40

【 0 0 3 0 】

50

(アクセス競合回避の処理の概要)

つぎに、複数のCPUによる一つのストレージに対するアクセス競合を回避する処理の概要について説明する。図4-1、図4-2は、アクセス競合の回避の処理概要を示す図である。はじめに、図4-1に示すように、CPU101がストレージからリードのタスクを実行しているとする。リードタスクは、データの読み出し先であるストレージがあらかじめ決まっている(図示の例では、メインストレージ131である)。

【0031】

このとき、新たにCPU102でライトタスクが発生したとする。このライトタスクのデータは、メインストレージ131にアクセスし、格納されるものであるとする。しかし、メインストレージ131は、CPU101によるリードタスクを実行中であるため、このライトタスクをどのストレージに書き込むかを判断する。この判断は、各ストレージのストレージ帯域M、Sの情報に基づき、メインストレージ131に対するタスクのアクセス終了時間を予測出し、いずれのストレージ131、132に書き込むかを決定する。図示の例では、ライトタスクとリードタスクが同じメインストレージ131にアクセスして生じるアクセス競合を避けるために、CPU102がライトタスクをサブストレージ132に書き込む状態を示している。

10

【0032】

図4-1に示す状態の後、図4-2に示すように、メインストレージ131に対するアクセスがない時期に、サブストレージ132に書き込んだデータをメインストレージ131に書き込む(上記Fタスクの処理)。これにより、同一のストレージに対する複数のCPUからのアクセス競合を回避する。この際、CPU101、102のOS111、112は、メモリ121に格納されているデータベース221を参照して同一ストレージのアクセス先のタスクを同一のCPUに割り当てることにより、同一ストレージへのアクセス競合を回避するようにしている。

20

【0033】

(タスクのアクセス終了時間の予測について)

ここで、上記のタスクのアクセス終了時間の予測の具体例について説明する。この予測処理は、ストレージスケジューラ205が実行する。ストレージスケジューラ205は、ライトタスクであるタスクC、Dをどのストレージに書き込むかを判断するために予測演算する。

30

【0034】

算出時の条件としては、

1. 分類Aのリードタスクは、一定帯域を利用してストレージにアクセスする。この分類Aのタスクは、他の分類B、C、D、E、Fのタスクと並列アクセスが可能である。
2. 分類B、C、D、E、Fのタスクは、利用可能な全帯域を利用してストレージにアクセスする。すなわち、分類Aが使用していない全帯域を利用する。
3. 分類B、C、D、E、Fのタスクは、シーケンシャルに処理する。分類Aのタスクとのみ並列アクセスを許可する。

【0035】

そして、メインストレージ131のストレージ帯域をM、サブストレージ132のストレージ帯域をS、メインストレージ131に割り当てられた分類BのタスクBxのデータ量をBmx、メインストレージ131に割り当てられた分類CのタスクCxのデータ量をCmx、メインストレージ131に割り当てられた分類AのタスクAxの使用帯域をAmx、サブストレージ132に割り当てられた分類BのタスクBxのデータ量をBsx、サブストレージ132に割り当てられた分類CのタスクCxのデータ量をCsx、サブストレージ132に割り当てられた分類AのタスクAxの使用帯域をAsxとする。このとき、各ストレージ割り当て時のアクセス終了時間をTm、Tsとすると、下記式(1)で求めることができる。

40

【0036】

【数 1】

$$T_m = \frac{\sum B_{mx} + \sum C_{mx}}{M - \sum A_{mx}} \quad T_s = \frac{\sum B_{sx} + \sum C_{sx}}{S - \sum A_{sx}} \quad \dots(1)$$

【0037】

そして、 $T_m > T_s$ であれば、ライトタスクC, Dをサブストレージ132へ書き込み、 $T_m < T_s$ であれば、ライトタスクC, Dをメインストレージ131へ書き込み指示する。

10

【0038】

(ストレージアクセススケジューリングについて)

はじめに、ストレージスケジューラ205が実行するストレージアクセススケジューリングについて説明する。図5-1~図5-5は、ストレージに対するアクセススケジューリングの例を示す図である。分類A, C(C1, C2), B, Dの順にスケジューリングする例について説明する。ここで、分類A, Bの対象データは、メインストレージ131に存在し、いずれも利用帯域は10、Bのデータ量は30、C1のデータ量は20、C2のデータ量は30であるとする。また、メインストレージ131、サブストレージ132ともにストレージ帯域は50であるとする。

【0039】

20

図5-1に示すように、はじめに、ストレージスケジューラ205は、ウェイトキュー203から分類Aのリードタスクを取り出す。この分類Aの対象データは、メインストレージ131にあるため、このタスクをメインストレージ131に割り当てる。つぎに、図5-2に示すように、分類C(C1)のライトタスクは、上記式(1)のアクセス終了時刻の予測演算をおこなうと、 $T_m = 20 / (50 - 10) = 0.5$ 、 $T_s = 20 / 50 = 0.4$ と算出されるため、この分類C1のタスクをサブストレージに割り当てる。

【0040】

つぎに、図5-3に示すように、分類C(C2)のライトタスクについては、 $T_m = 30 / (50 - 10) = 0.75$ 、 $T_s = 20 + 30 / 50 = 1$ により、この分類C2のタスクをメインストレージに割り当てる。つぎに、図5-4に示すように、分類Bのリードタスクは、メインストレージ131に存在するので、この分類Bのタスクをメインストレージ131に割り当てる。最後に、図5-5に示すように、分類Dのライトタスクについては、 $T_m = (30 + 30) / (50 - 10) = 1.5$ 、 $T_s = 20 / 50 = 0.4$ と算出されるため、この分類Dのタスクをサブストレージに割り当てる。

30

【0041】

ストレージスケジューラ205は、上記のスケジューリング毎に、割り当てたタスクのIDと、終了予測時間をストレージアクセススケジューリングテーブル321に保存していく。

【0042】

(ストレージアクセスとタスクアクセスの連動について)

40

つぎに、ストレージアクセスと、タスクアクセスの連動について説明する。タスクスケジューラ206は、ストレージスケジューラ205のスケジューリング情報をストレージアクセススケジューリングテーブル321から読み出し、各CPUのタスクキューにタスクのキューイングをおこなう。また、タスクディスパッチ部207, 217は、CPU101, 102のディスパッチ状況を基に、タスクキュー204, 214に挿入されているタスクのタスクディスパッチをおこなう。そして、アクセス監視部202, 212は、ストレージに対するリード、ライトの発生時に、実行中タスクの情報に基づき、タスクがアクセスするストレージとして、メインストレージ131あるいはサブストレージ132を指定する。

【0043】

50

このタスクスケジューラ 206 は、ストレージアクセス先が同一のストレージとなるように、各 CPU 101, 102 のタスクキュー 204, 214 にキューイングする。これにより、同一ストレージに対するアクセス競合を避ける。また、タスクディスパッチ部 207, 217 は、分類 B, C のタスクを各 CPU 101, 102 で同時に一つしかディスパッチしない。なお、分類 B, C のタスクは、既に別の分類 B, C のタスクがディスパッチされている場合には、タスクキュー 204, 214 に入れたままにする。

【0044】

また、タスクディスパッチ部 207, 217 は、分類 A, D, E のタスクがキューイングされた場合は、すぐにディスパッチする。これは、分類 A, D, E のタスクはアクセス競合による影響が小さいためである。分類 A のタスクは、全帯域を使用しないためであり、分類 D, E のタスクは、処理時間が非常に短いためである。

10

【0045】

また、タスクディスパッチ部 207, 217 は、分類 F のタスクがストレージアクセススケジューリングされており、各 CPU 101, 102 のタスクキュー 204, 214 に分類 B, C のタスクが存在せず各 CPU 101, 102 で分類 B, C のタスクがディスパッチされていない場合には、この分類 F のタスクをディスパッチする。

【0046】

また、タスクディスパッチ部 207, 217 は、分類 F のタスクがディスパッチ中に分類 B, C のタスクがキューイングされた場合には、ただちに分類 F のタスクをタスクキュー 204, 214 に戻す。これは、書き戻しによる性能低下を防ぐためである。

20

【0047】

図 6 - 1 ~ 図 6 - 5 は、ストレージアクセスとタスクアクセスを連動させたスケジューリングの例を示す図である。図 5 - 1 ~ 図 5 - 5 と同様に、分類 A, B の対象データは、メインストレージ 131 に存在し、いずれも利用帯域は 10、B のデータ量は 30、C1 のデータ量は 20、C2 のデータ量は 30 であるとする。また、メインストレージ 131、サブストレージ 132 とともにストレージ帯域は 50 であるとする。また、タスクの発生順序分類 A C1 C2 B D と、各タスクのストレージスケジューリングについては、図 5 - 1 ~ 図 5 - 5 を用いて説明した結果を利用するものとする。

【0048】

はじめに、図 6 - 1 に示すように、発生した分類 A のタスクは、ストレージスケジューラ 205 によりメインストレージ 131 にアクセススケジューリングされる（図 5 - 1 参照）。また、この分類 A のタスクは、いずれの CPU 101, 102 でも割り当て可能である。図示の例では、タスクスケジューラ 206 は、分類 A のタスクを CPU 101 に割り当てたものとする。

30

【0049】

図 6 - 2 に示すように、つぎに発生した分類 C1 のタスクは、ストレージスケジューラ 205 によりサブストレージ 132 にアクセススケジューリングされる（図 5 - 2 参照）。そして、タスクスケジューラ 206 は、この分類 C1 のタスクを分類 A のタスクとはアクセスストレージが異なるので、異なる CPU 102 に割り当てる。

【0050】

図 6 - 3 に示すように、つぎに発生した分類 C2 のタスクは、ストレージスケジューラ 205 によりメインストレージ 131 にアクセススケジューリングされる（図 5 - 3 参照）。そして、タスクスケジューラ 206 は、この分類 C2 のタスクについて、アクセス先が同一であるタスク A と同一の CPU 101 に割り当てる。上述したように、分類 A のタスクと分類 C (C2) のタスクは並列にアクセス可能であるため、CPU 101 はこれら分類 A と分類 C2 のタスクをマルチタスク処理する。

40

【0051】

図 6 - 4 に示すように、つぎに発生した分類 B のタスクは、ストレージスケジューラ 205 によりメインストレージ 131 にアクセススケジューリングされる（図 5 - 4 参照）。そして、タスクスケジューラ 206 は、この分類 B のタスクについて、アクセス先が同

50

一であるタスク A , C 2 と同一の CPU 1 0 1 に割り当てる。但し、上述したように、分類 B のタスクは、分類 C (C 2) のタスクと並列アクセス不可であるため、CPU 1 0 1 は、分類 B のタスクをタスクキュー 2 0 4 にキューイングしておく。

【 0 0 5 2 】

図 6 - 5 に示すように、つぎに発生した分類 D のタスクは、ストレージスケジューラ 2 0 5 によりサブストレージ 1 3 2 にアクセススケジューリングされる (図 5 - 5 参照) 。そして、タスクスケジューラ 2 0 6 は、この分類 D のタスクについて、アクセス先が同一であるタスク C 1 と同一の CPU 1 0 2 に割り当てる。但し、この分類 D のタスクは、アクセス時間が短いので、タスクディスパッチ部 2 1 7 は、分類 C 1 のタスクと、この分類 D のタスクのディスパッチ順を入れ替えて分類 D のタスクを先に処理させる。

10

【 0 0 5 3 】

(データの書き戻しについて)

ライトアクセスのタスクを対象ストレージではなく、別ストレージにスケジューリングした場合、F タスクを生成し、対象ストレージへの書き戻しをおこなう。このため、書き戻し終了までの期間、一時領域を確保する必要がある。

【 0 0 5 4 】

図 7 は、書き戻し前のデータへのアクセスの状態を示す図である。たとえば、分類 B のタスクが対象ストレージであるメインストレージ 1 3 1 ではなく、異なるサブストレージ 1 3 2 に一時保存された状態が生じたとする。このような書き戻しが終了する前の状態のデータ 7 0 1 に対するアクセスが発生した場合には、データ 7 0 1 の本来の保存先であるメインストレージ 1 3 1 ではなく、この時点でデータ 7 0 1 が一時保存されているサブストレージ 1 3 2 へアクセスする制御が必要となる。

20

【 0 0 5 5 】

上記データの書き戻しに対応するために、タスク毎に固有のタスク情報をあらかじめ付加しておき、タスクスケジューラ 2 0 6 と、ストレージスケジューラ 2 0 5 とを連携させる。この連携に用いる情報が、上述した図 3 - 1 に示したタスクテーブル 3 0 1 である。ストレージスケジューラ 2 0 5 は、このタスクテーブル 3 0 1 を参照して、アクセス先を特定する。また、ライトタスクテーブル 3 1 1 の一時保存先 (上記例ではサブストレージ 1 3 2) にアクセスしているリードタスクが存在する間は、この一時保存されているデータ 7 0 1 を保護しておく。そしてストレージスケジューラ 2 0 5 は、スケジューリング時に、このタスクテーブル 3 0 1 を参照して F タスクに対するスケジューリングを決定する。

30

【 0 0 5 6 】

(F タスクのディスパッチ例)

図 8 - 1 ~ 図 8 - 3 は、データの書き戻しにかかる処理を示す図である。図 8 - 1 に示すように、CPU 1 0 1 , 1 0 2 のキャッシュ (L 1 キャッシュ) 8 0 1 , 8 0 2 間には、スヌープコントローラ 8 0 3 が設けられる。このスヌープコントローラ 8 0 3 を介してキャッシュ 8 0 1 , 8 0 2 の更新データを CPU 1 0 1 , 1 0 2 間でやりとりし、キャッシュ 8 0 1 , 8 0 2 のコヒーレンシをとっている。

【 0 0 5 7 】

また、キャッシュ 8 0 1 には、CPU 1 0 1 , 1 0 2 の動作フラグ (C フラグと称す) 8 1 1 , 8 1 2 が設けられ、キャッシュ 8 0 2 にも CPU 1 0 1 , 1 0 2 の C フラグ 8 2 1 , 8 2 2 が設けられている。

40

【 0 0 5 8 】

CPU 1 0 1 に関する C フラグ 8 1 1 , 8 2 1 は、CPU 1 0 1 で分類 B , C , D , E のタスクを実行時に 1、F タスクを実行時に 2 の値をとる。CPU 1 0 2 に関する C フラグ 8 1 2 , 8 2 2 についても同様に、CPU 1 0 2 で分類 B , C , D , E のタスクを実行時に 1、F タスクを実行時に 2 の値をとる。それ以外 (分類 A のタスク実行時を含む) の値は 0 (O F F) である。

【 0 0 5 9 】

50

そして、図 8 - 1 に示すように、CPU 101 は、分類 A のタスク A 1 と分類 C のタスクをマルチタスク処理しており、タスクキュー 204 に分類 F のタスクがキューイングされている。また、CPU 102 は、分類 D のタスクを処理し、タスクキュー 214 に分類 A 2 のタスクがキューイングされているとする。この場合、CPU 101 の C フラグ 811, 821 は、分類 C のタスク実行時に 1、分類 A のタスク実行時に 0 となる。CPU 102 の C フラグ 812, 822 は、分類 D のタスク実行により 1 となる。

【0060】

つぎに、図 8 - 2 に示すように、CPU 101 における分類 C のタスクが終了すると（分類 A のタスク（A 1）は実行継続）、このとき CPU 101 の C フラグ 811, 821 は、値が 0 になる。また、CPU 102 における分類 D のタスクが終了すると、つぎに、分類 A のタスク（A 2）が実行されるが、このとき CPU 102 の C フラグ 812, 822 は、値が 0 になる。これにより、CPU 101 は、すべての C フラグ 811, 812 の値が 0 になったことにより、タスクキュー 204 の F タスクをディスパッチする。

10

【0061】

これにより、図 8 - 3 に示すように、CPU 101 で分類 F の F タスクが実行され、F タスクの実行中は、CPU 101 の C フラグ 811, 821 は、値が 2 となる。このように、分類 F の F タスクは、複数の CPU 101, 102 のうち一つの CPU でのみ実行される。

【0062】

（システムの処理手順）

20

上述したように、実施の形態では、ストレージスケジューラ 205 と、タスクスケジューラ 206 が連動して動作する。ストレージスケジューラ 205 のアクセススケジューリング結果を基に、タスクスケジューラ 206 が CPU 101, 102 へのタスクの配布をおこなう。

【0063】

（ストレージスケジューラの処理手順）

図 9 は、ストレージスケジューラがおこなう処理手順を示すフローチャートである。マスタスケジューラ 201 に設けられるストレージスケジューラ 205 は、タスクの発生を待ち（ステップ S 901：No、およびステップ S 902：No のループ）、タスクが発生すると（ステップ S 901：Yes）、タスクの種類を判別する（ステップ S 903）。タスクが発生せず（ステップ S 901：No）、すべてのタスクが終了すると（ステップ S 902：Yes）、終了する。

30

【0064】

ステップ S 903 のタスクの種別の判別では、あらかじめ定めた分類 A ~ F のタスクであれば（ステップ S 903：Yes）、つぎに、分類 A, B, E のリードタスクであるか判別し（ステップ S 904）、分類 A ~ F のタスクでなければ（ステップ S 903：No）、ストレージスケジューリングを実施せず、タスクスケジューラ 206 の処理（後述する図 10）へ移行する。

【0065】

ステップ S 904 において、分類 A, B, E のリードタスクであれば（ステップ S 904：Yes）、アクセスデータが存在するストレージをチェックし、アクセスデータが存在するストレージにスケジューリングする。はじめに、タスクテーブル 301 を参照し、ライトタスクテーブル 311 の（2）指定書き込み領域と、リードタスクテーブル 312 の（2）指定アクセス領域を比較し、一致するライトタスクを検索する（ステップ S 905）。

40

【0066】

つぎに、該当するライトタスクの（4）書き戻し判定フラグをチェックし（ステップ S 906）、書き戻しが完了しているか判定する（ステップ S 907）。書き戻しが完了していれば（ステップ S 907：Yes）、リードタスクテーブル 312 の（3）読み出し指定領域をライトタスクテーブル 311 の（2）指定書き込み領域に更新し（ステップ S

50

908)、ステップS911に移行する。

【0067】

一方、ステップS907において、書き戻しが完了していなければ(ステップS907:No)、リードタスクテーブル312の(3)読み出し指定領域をライトタスクテーブル311の(3)書き込み指定領域に更新し(ステップS909)、また、ライトタスクテーブル311の(5)一時保存先アクセスリードタスクにリードタスクテーブル312の(1)リードタスクIDを付加し(ステップS910)、ステップS911に移行する。

【0068】

ステップS911では、ライトタスクテーブル311の(6)データサイズに、読み出すデータサイズを書き込み(ステップS911)、該当するタスクを対象データのあるストレージに割り当てる(ステップS912)。この後、ストレージアクセススケジューリングテーブル321を更新し(ステップS913)、タスクスケジューラ206の処理(図10)へ移行する。

10

【0069】

ステップS904において、分類A, B, Eのリードタスク以外のときには(ステップS904:No)、分類Fのタスクであるか判定する(ステップS914)。分類Fのタスクでなければ(ステップS914:No)、タスクの種別は、残る分類C, Dのいずれかのライトタスクであるため、ストレージアクセス終了時間を予測し、アクセスするストレージを決定し、割り当てる(ステップS915)。この後、ライトタスクテーブル311の(2)指定書き込み領域と、(3)書き込み指定領域が等しいか(すなわち、書き戻し不要か)判定する(ステップS916)。

20

【0070】

ステップS916の判定結果、ライトタスクテーブル311の(2)指定書き込み領域と、(3)書き込み指定領域が等しければ(ステップS916:Yes)、ライトタスクテーブル311を更新する(ステップS917)。この際、ライトタスクテーブル311の(2)指定書き込み領域~(4)書き戻し判定フラグ、(6)データサイズを更新する。ステップS916の判定結果、ライトタスクテーブル311の(2)指定書き込み領域と、(3)書き込み指定領域が等しくなければ(ステップS916:No)、書き戻しが必要であり、Fタスクをタスクキュー204に挿入してから(ステップS918)、ステップS917に移行する。ステップS917の実行後は、ストレージアクセススケジューリングテーブル321を更新し(ステップS913)、タスクスケジューラ206の処理(図10)へ移行する。

30

【0071】

また、ステップS914の判定結果、分類Fのタスクであれば(ステップS914:Yes)、このFタスクを対象データのある(書き戻しをおこなうデータが存在する)ストレージに割り当て(ステップS919)、ストレージアクセススケジューリングテーブル321を更新し(ステップS913)、タスクスケジューラ206の処理(図10)へ移行する。

40

【0072】

(タスクスケジューラの処理手順)

図10は、タスクスケジューラがおこなう処理手順を示すフローチャートである。マスタースケジューラ201に設けられるタスクスケジューラ206は、図9に示したストレージスケジューラ205によるストレージアクセススケジューリングの処理後に実行される。

【0073】

はじめに、該当するタスクが分類A~Fのタスクであるか判別し(ステップS1001)、分類A~Fのタスクであれば(ステップS1001:Yes)、同一ストレージに割り当てられているタスクがいずれかのCPU101, 102のタスクキュー204, 214に存在するか判断する(ステップS1002)。判断結果、同一ストレージに割り当て

50

られているタスクがいずれかのCPU101, 102のタスクキュー204, 214に存在すれば(ステップS1002: Yes)、このタスクを同一のタスクキュー204, 214に割り当て(ステップS1003)、処理を終了してストレージスケジューラ205の処理(図9)に戻る。

【0074】

ステップS1001の判断結果が分類A~Fのタスクではないとき(ステップS1001: No)、およびステップS1002の判断結果が同一ストレージに割り当てられているタスクがいずれかのCPU101, 102のタスクキュー204, 214に存在しないとき(ステップS1002: No)、通常スケジューリングを実施し(ステップS1004)、処理を終了してストレージスケジューラ205の処理(図9)に戻る。

10

【0075】

(メインスケジューラ、スレーブスケジューラの処理手順)

図11は、メインスケジューラ、スレーブスケジューラがおこなう処理手順を示すフローチャートである。はじめに、メインスケジューラ201、スレーブスケジューラ211は、タスクキュー204, 214にタスクが存在するか、もしくはいずれかのCPUのフラグ値が変化したかを判定し(ステップS1101)、タスクが存在すれば(ステップS1101:(1))、分類B~Fのタスクであるか判定する(ステップS1102)。いずれかのCPUのフラグ値が変化した場合は(ステップS1101:(2))、Fタスクを実行中であるかを判定し(ステップS1111)、Fタスク実行中であれば(ステップS1111: Yes)、Fタスクをタスクキューに戻し(ステップS1112)、ステップS1101に戻り、Fタスク実行中でなければ(ステップS1111: No)、処理を終え、ステップS1101に戻る。いずれの判定にも当てはまらない場合(ステップS1101: No)は、上記処理を終え、タスクの発生、もしくは、いずれかのCPUフラグ値の変更を待つ。

20

【0076】

ステップS1102において、分類B~Fのタスクであれば(ステップS1102: Yes)、分類Fのタスクであるか判断し(ステップS1103)、分類B~Fのタスクでなければ(ステップS1102: No)、ステップS1110へ移行する。分類Fのタスクであれば(ステップS1103: Yes)、いずれかのCPU101, 102で分類B~Fのタスクがスレッド化されているか判断する(ステップS1104)。判断結果、分類B~Fのタスクがスレッド化されていれば(ステップS1104: Yes)、処理を終了し、分類B~Fのタスクがスレッド化されていなければ(ステップS1104: No)、Cフラグを2に設定し(ステップS1105)、このFタスクをスレッド化して処理を開始し(ステップS1106)、処理終了により上記処理を終える。

30

【0077】

ステップS1103において、分類Fのタスクでなければ(ステップS1103: No)、同一CPU101, 102において(1)分類B~Eのタスクがスレッド化されているか、あるいは(2)Fタスクがスレッド化されているか判定する(ステップS1107)。判定結果、分類B~Eのタスクがスレッド化されていれば(ステップS1107: 結果(1))、このタスクをスレッド化せずに処理終了する。また、(2)Fタスクがスレッド化されていれば(ステップS1107: 結果(2))、Fタスクをタスクキュー214に戻し(ステップS1108)、Cフラグを1に設定し(ステップS1109)、このタスクをスレッド化して処理を開始し(ステップS1110)、処理終了により、上記一連の処理を終える。

40

【0078】

また、ステップS1107において、上記(1)分類B~Eのタスクがスレッド化されておらず、さらに、(2)Fタスクがスレッド化されていない場合には(ステップS1107: No)、ステップS1110により該当タスクをスレッド化して処理を開始し(ステップS1110)、処理終了により、上記一連の処理を終える。

【0079】

50

(アクセス監視部の処理手順)

図12は、アクセス監視部がおこなう処理手順を示すフローチャートである。アクセス監視部202, 212は、それぞれストレージに対するリード、ライトの発生時に、実行中タスクの情報に基づき、タスクがアクセスするストレージ(メインストレージ131あるいはサブストレージ132)を指定する。

【0080】

はじめに、(1)タスクディスパッチ発生、あるいは(2)ディスパッチ中のタスク処理終了が発生したか判定する(ステップS1201)。(1)タスクディスパッチが発生したときには(ステップS1201:結果(1))、このタスクが分類A~Fのタスクか判定する(ステップS1202)。分類A~Fのタスクであれば(ステップS1202:Yes)、つぎに、分類A, B, Eのリードタスクであるか判定する(ステップS1203)。ステップS1201, およびステップS1202のいずれにおいても該当しないタスクであれば(ステップS1201:No, およびステップS1202:No)、今回のタスクに関する特別なアクセス制御は実施せずに処理を終える。

10

【0081】

ステップS1203において、分類A, B, Eのリードタスクであれば(ステップS1203:Yes)、リードタスクテーブル312から情報(3)読み出し指定領域を取得し(ステップS1204)、この読み出し指定領域からデータの読み出しを開始する(ステップS1205)。この後、このタスクの(1)タスク終了、あるいは(2)タスクスイッチを判定し(ステップS1206)、(1)タスク終了時には(ステップS1206:結果(1))、ステップS1207に移行し、(2)タスクスイッチ時には(ステップS1206:結果(2))、処理を終了する。また、(1)タスク終了、(2)タスクスイッチのいずれでもなければ、タスク終了、あるいはタスクスイッチの発生を待つ(ステップS1206:Noのループ)。

20

【0082】

ステップS1207では、リードタスクテーブル312の(2)指定アクセス領域と(3)読み出し指定領域とが一致しているか判定する。判定の結果、(2)指定アクセス領域と(3)読み出し指定領域とが一致していれば(ステップS1207:Yes)、処理を終了し、(2)指定アクセス領域と(3)読み出し指定領域とが一致していなければ(ステップS1207:No)、一時保存先からデータを読み込んでいたため、ライトタスクテーブル311の(2)指定書き込み領域と、リードタスクテーブル312の(2)指定アクセス領域が一致するライトタスクを検索し、該当ライトタスクについてライトタスクテーブル311の(5)一時保存先アクセスリードタスクから終了したリードタスクのIDを除去し(ステップS1208)、処理を終了する。

30

【0083】

また、ステップS1201において、(2)ディスパッチ中のタスク処理終了が発生したときには(ステップS1201:結果(2))、分類B~Fのタスクであるか判定し(ステップS1209)、分類B~Fのタスクであれば(ステップS1209:Yes)、Cフラグを初期値(0)に設定し(ステップS1210)、処理を終了する。一方、分類B~Fのタスクでなければ(分類Aのタスクを含む)(ステップS1209:No)、処理をおこなわず、終了する。

40

【0084】

また、ステップS1203において、分類A, B, Eのリードタスク以外であれば(ステップS1203:No)、このタスクがライトタスクであるかFタスクであるか判定する(ステップS1211)。判定の結果、Fタスクであれば(ステップS1211:Yes)、Fタスクテーブル313の情報(2)書き戻しライトタスクIDに該当するライトタスクをライトタスクテーブル311から検索し、該当タスクの情報(2)指定書き込み領域と、(3)書き込み指定領域を取得し(ステップS1212)、対象領域の書き戻しを開始する(ステップS1213)。

【0085】

50

この後、このタスクの(1)タスク終了、あるいは(2)タスクスイッチを判定し(ステップS1214)、(1)タスク終了時には(ステップS1214:結果(1))、書き戻したライトタスクの情報(4)書き戻し判定を更新(終了)し(ステップS1215)、ライトタスクテーブル311の情報(2)指定書き込み領域の保護を解除し(ステップS1216)、処理を終了する。また、ステップS1214において、(2)タスクスイッチ時には(ステップS1214:結果(2))、処理をおこなわず、終了する。

【0086】

また、ステップS1211の判定の結果、Fタスクでなければ(ステップS1211:No)、分類C、Dのライトタスクであり、ライトタスクテーブル311から情報(2)指定書き込み領域と、(3)書き込み指定領域を取得する(ステップS1217)。この後、(2)指定書き込み領域と(3)書き込み指定領域が一致しているか判定する(ステップS1218)。判定の結果、一致していなければ(ステップS1218:No)、書き込み先のストレージ領域を保護し(ステップS1219)、指定領域へのデータの書き込みを開始し(ステップS1220)、処理終了により、上記一連の処理を終える。ステップS1218にて、(2)指定書き込み領域と(3)書き込み指定領域が一致していれば(ステップS1218:Yes)、ステップS1219をスキップして指定領域へのデータの書き込みを開始し(ステップS1220)、処理終了により、上記一連の処理を終える。

10

【0087】

(実施の形態による競合回避および全アクセスの処理時間について)

20

図13は、実施の形態による競合の回避および全アクセスの処理時間を説明する図である。上述したストレージアクセススケジューリングの結果、メインストレージ131に対して分類CのライトタスクC1、C3を割り当て、サブストレージ132に対して分類CのライトタスクC2、C4を割り当てたと仮定する。ここで、各タスクC1~C4の書き込みデータ量はいずれも50、各ストレージの帯域は50であるとする。

【0088】

上述の実施の形態の処理(図中右下)によれば、同一アクセス先のタスクが同一CPU101、102に割り当てられるので、メインストレージ131とサブストレージ132のいずれにもアクセス競合は発生しない。また、各CPU101、102でシーケンシャルにアクセスがおこなわれると仮定すると、全アクセスの処理時間は、 $50/50 + 50/50 = 2$ となる。すなわち、メインストレージ131とサブストレージ132の全帯域を有効に使用することができる。

30

【0089】

一方、システムの基本構成は実施の形態と同一構成であるが、従来の未対策技術、すなわち、ストレージアクセススケジュールのみ実行し、ストレージアクセススケジューリングとタスクスケジューリングとが連携していない構成を対比のために図中左下に示した。この場合、図示のように、タスクC1とC2がCPU101に割り当てられ、タスクC3とC4がCPU102に割り当てられる場合がある。このようにタスクスケジューリングされ、メインストレージ131に割り当てられているタスクC1、C3に対しCPU101のタスクC1と、CPU102のタスクC3が実行されると、メインストレージ131に対してCPU101、102が同時にアクセスするアクセス競合が発生する。そして、各CPU101、102でシーケンシャルにアクセスがおこなわれると仮定すると、全アクセスの処理時間は、タスクC1とC3の実行によるアクセス競合、およびタスクC2とC4の時刻によるアクセス競合の発生により、 $50/25 + 50/25 = 4$ となり、実施の形態に比して倍の処理時間がかかる。この例からみても、実施の形態の処理によれば、高いアクセス性能を達成できることがわかる。

40

【0090】

(システムの適用例)

図14は、図1に示したコンピュータを用いたシステムの適用例を示す図である。図14において、ネットワークNWは、サーバ1401、1402とクライアント1431~

50

1434とが通信可能なネットワークであり、たとえば、LAN(Local Area Network)、WAN(Wide Area Network)、インターネット、携帯電話網などで構成される。

【0091】

サーバ1402は、クラウド1420を構成するサーバ群(サーバ1421~1425)の管理サーバである。クライアント1431~1434のうち、クライアント1431はノート型パソコン、クライアント1432はデスクトップ型パソコン、クライアント1433は携帯電話機(スマートフォン、PHS(Personal Handyphone System)でもよい)、クライアント1434はタブレット型端末である。図14のサーバ1401, 1402, 1421~1425、クライアント1431~1434は、たとえば、図1に示したコンピュータにより実現される。

10

【0092】

また、本実施の形態は、図1に示した各CPU101, 102、およびストレージ131, 132がそれぞれ異なるコンピュータ(たとえば図14の携帯電話機やサーバ)に搭載され、複数のコンピュータがネットワークNWを介して分散型の並列処理をおこなう構成に適用することもできる。

【0093】

以上説明した開示技術によれば、同一のストレージにアクセスするタスクを同じCPUに割り当てることにより、複数のCPUが一つのストレージにアクセスするアクセス競合を回避することができ、効果的なスケジューリングをおこなえるようになる。

20

【0094】

また、タスクの種別を判定して、あるタスクによってアクセスしているストレージに対してさらに他のタスクが生じることによるアクセス競合を避けるようにしている。タスクの種別については、データサイズによる処理時間等に基づき、複数の種別のタスクについて、それぞれ分類しておけばよく、瞬間的にアクセス競合が発生するが全体のアクセス時間に影響を与えない場合には、アクセス競合を許容することも判断している。たとえば、全帯域ではなく一定帯域のみ利用するリードタスクであれば、他のライトタスク等は、並行して同じストレージへのアクセスを許容してもよい。

【0095】

また、ライトタスクのデータ書き込み対象のストレージは、複数のタスクがそれぞれのストレージにアクセスするときの時間に基づき決定するため、アクセス時間を短時間とすることができるようになる。この決定には、各ストレージの帯域や、分類されたタスクのデータ量などの複数の値を用いており、適したストレージを決定できるようになる。

30

【0096】

また、データの読み出し先のストレージが定まっているリードタスクを一つのCPUで実行中に、異なるCPUでライトタスクを実行する際には、ライトタスクのデータを一旦、他のストレージに書き込む。他のストレージに一旦書き込まれたデータは、ストレージへのアクセス終了後に本来のストレージに移動させる。これにより、システム全体のストレージに対するアクセス処理を効率化できるようになる。

【0097】

上記の実施形態において説明したストレージは、たとえばディスク装置であるが、これに限らず、複数のCPUが共有使用することによりアクセス競合が生じるメモリであれば、他の各種データメモリ装置であっても同様に適用することができる。

40

【符号の説明】

【0098】

100 システム

101, 102 CPU

103 バス

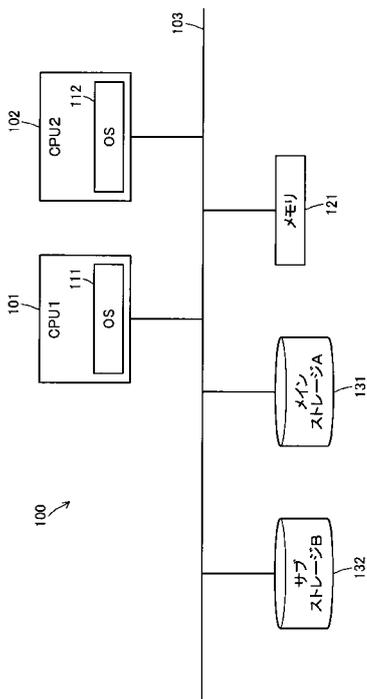
111, 112 オペレーティングシステム(OS)

121 メモリ

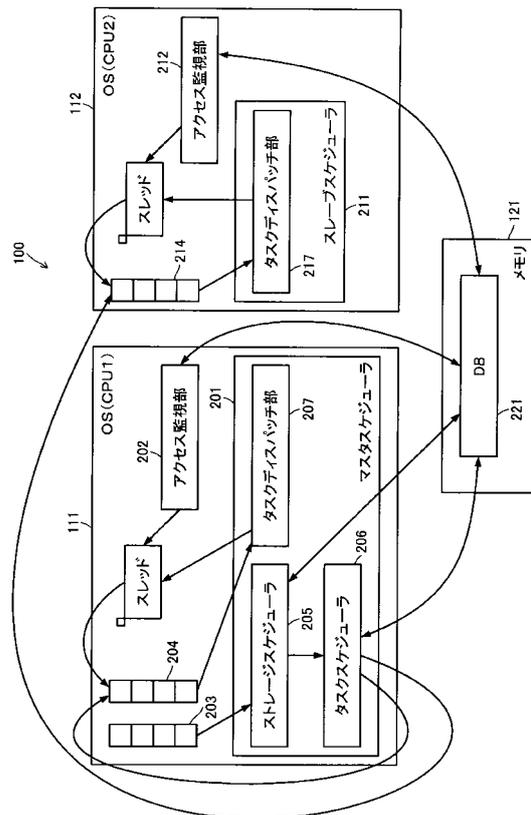
50

- 1 3 1 メインストレージ
- 1 3 2 サブストレージ
- 2 0 1 マスタスケジューラ
- 2 0 2 , 2 1 2 アクセス監視部
- 2 0 3 ウェイトキュー
- 2 0 4 , 2 1 4 タスクキュー
- 2 0 5 ストレージスケジューラ
- 2 0 6 タスクスケジューラ
- 2 0 7 , 2 1 7 タスクディスパッチ部
- 2 1 1 スレーブスケジューラ
- 2 2 1 データベース (D B)
- 3 0 1 タスクテーブル
- 3 1 1 ライトタスクテーブル
- 3 1 2 リードタスクテーブル
- 3 1 3 Fタスクテーブル
- 3 2 1 ストレージアクセススケジューリングテーブル
- 3 2 2 タスクスケジューリングテーブル

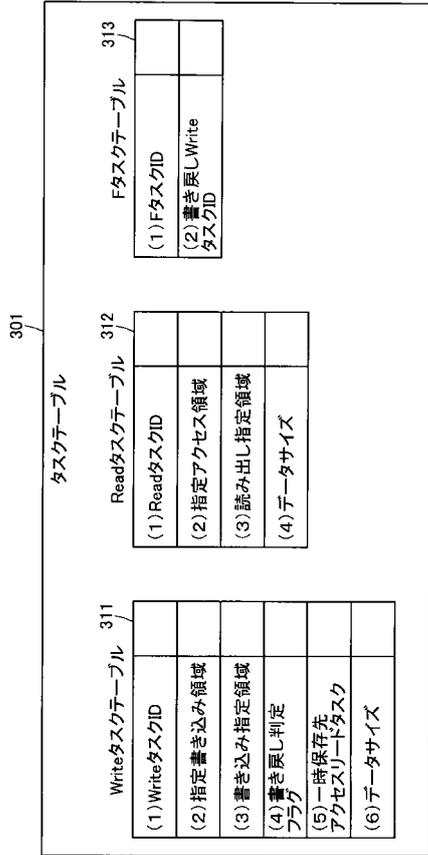
【 図 1 】



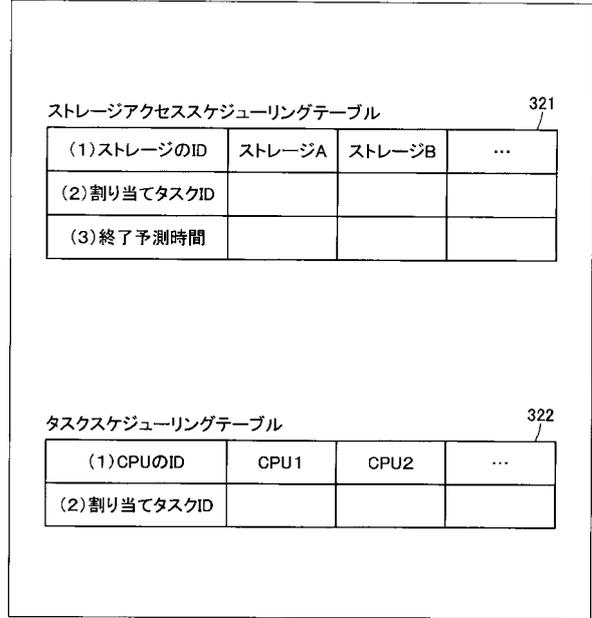
【 図 2 】



【 図 3 - 1 】



【 図 3 - 2 】

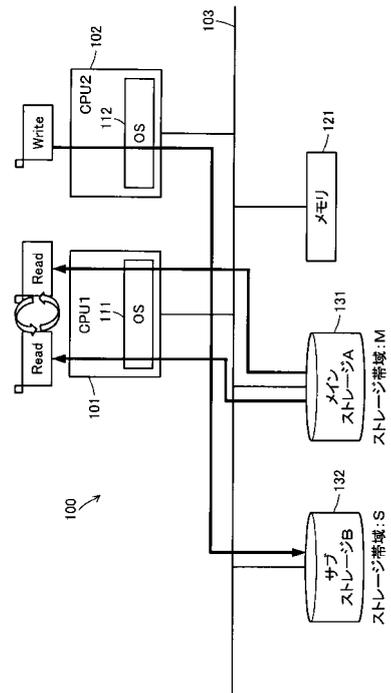


【 図 3 - 3 】

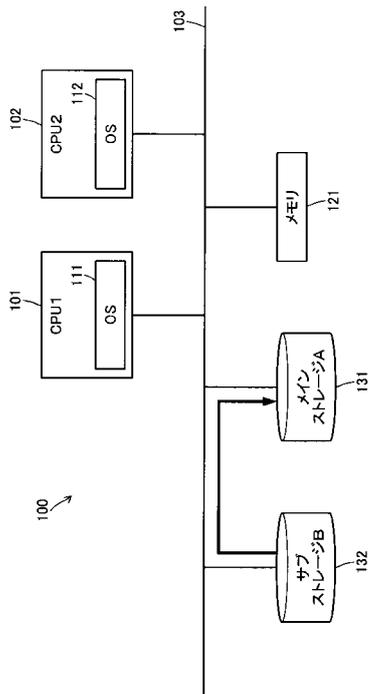
331

分類	アクセスタイプ	帯域利用	アプリ例	処理時間
A	Read	一定帯域	ビデオ、ナビ	データサイズによる
B	Read	全帯域 (利用可能な限り)	アップロード	データサイズによる
C	Write	全帯域 (利用可能な限り)	ダウンロード	データサイズによる
D	Write	最優先	スワップアウト	一瞬
E	Read	最優先	スワップイン	一瞬
F	Read and Write	全帯域 (利用可能な限り)	サブディスクから メインディスクへの書き戻し	データサイズによる

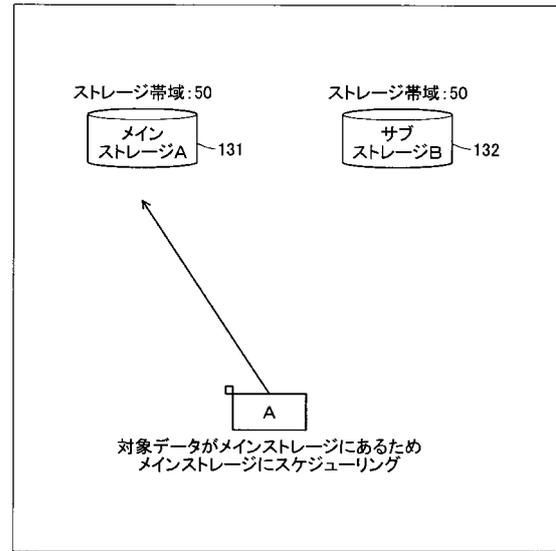
【 図 4 - 1 】



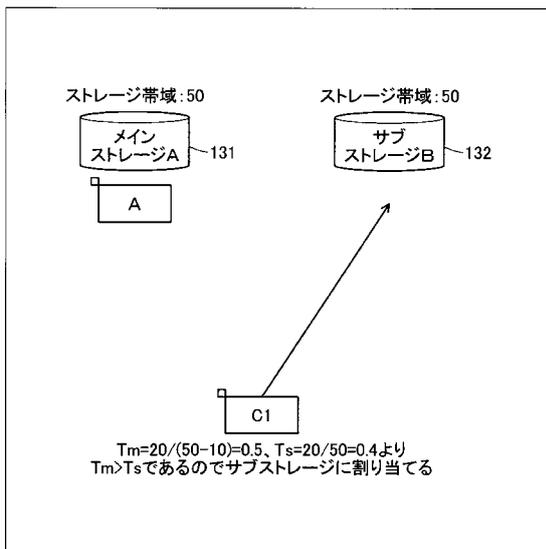
【 図 4 - 2 】



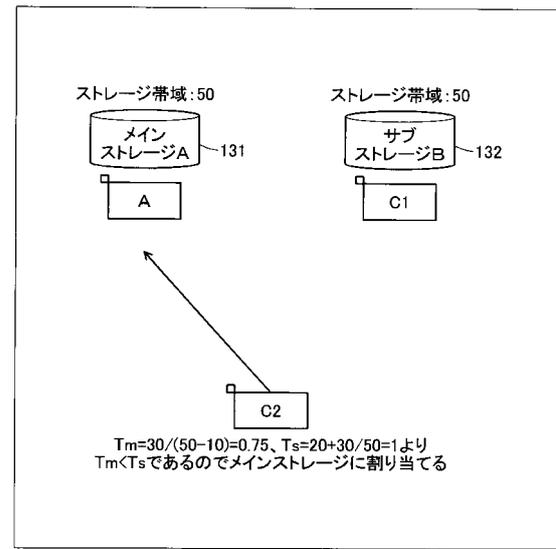
【 図 5 - 1 】



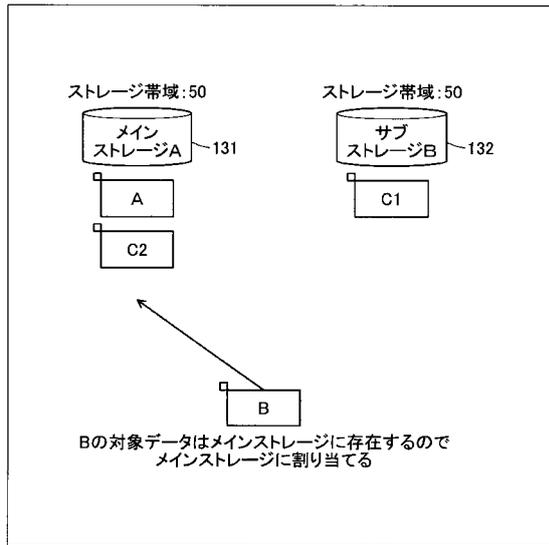
【 図 5 - 2 】



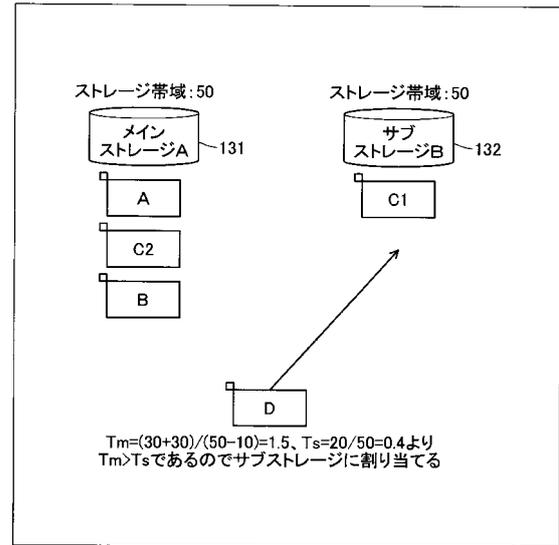
【 図 5 - 3 】



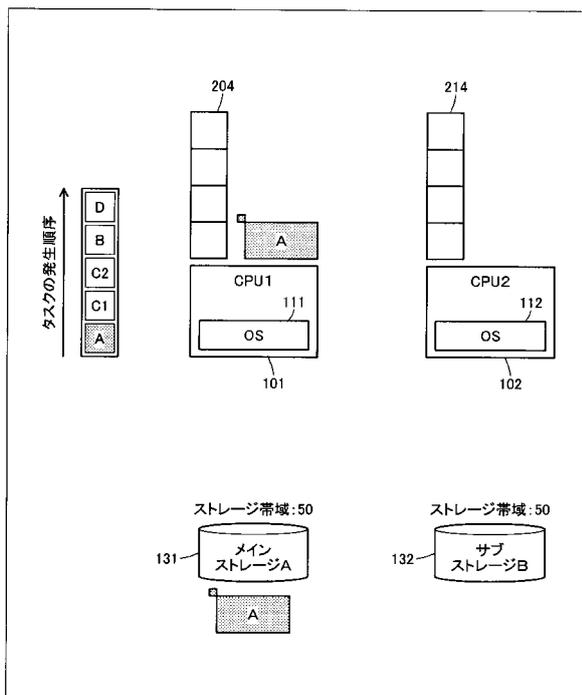
【 図 5 - 4 】



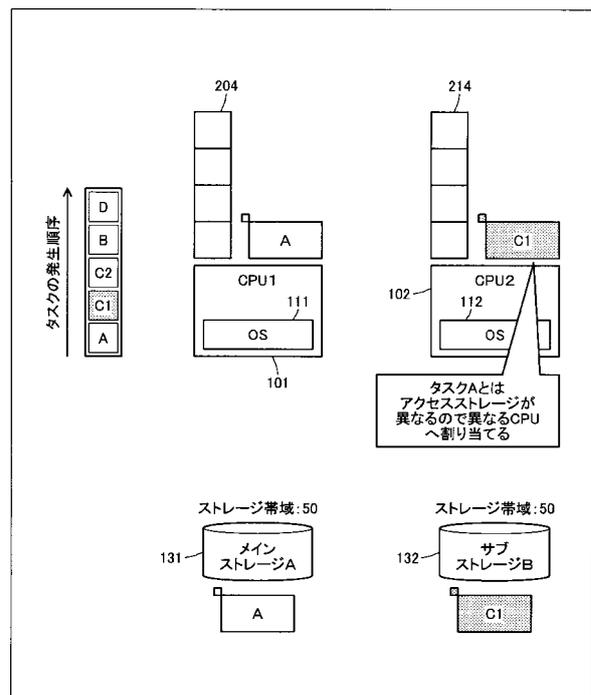
【 図 5 - 5 】



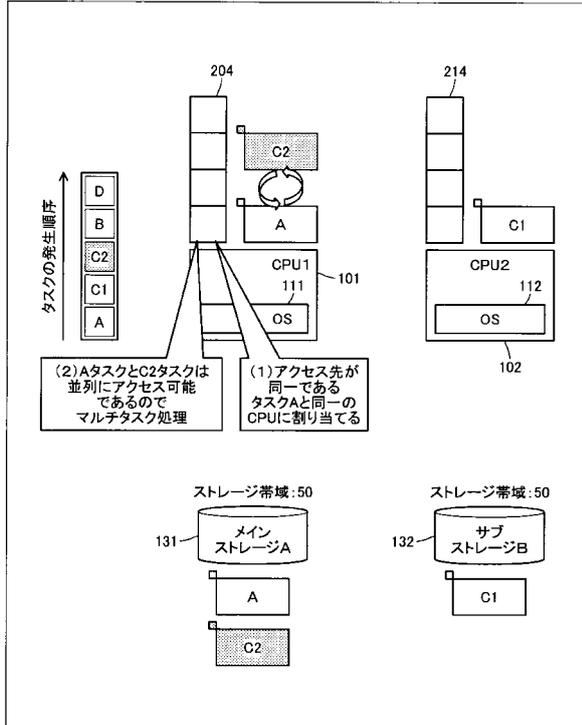
【 図 6 - 1 】



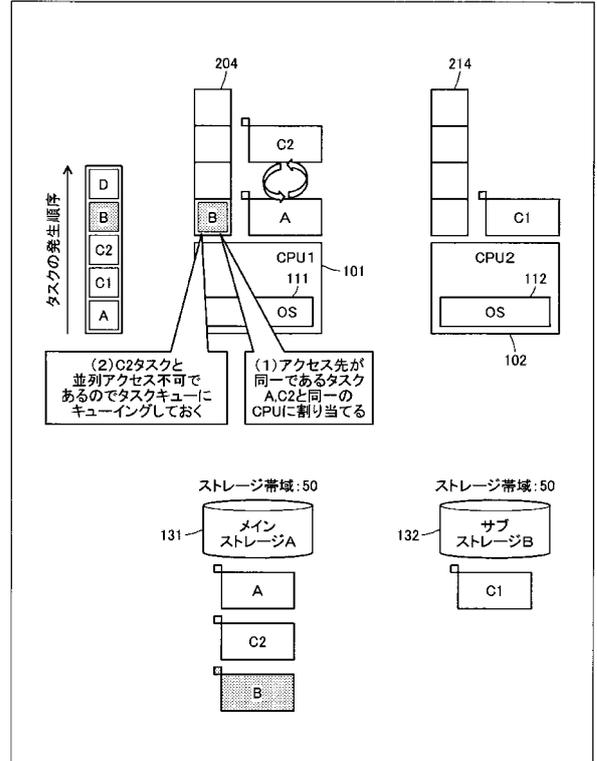
【 図 6 - 2 】



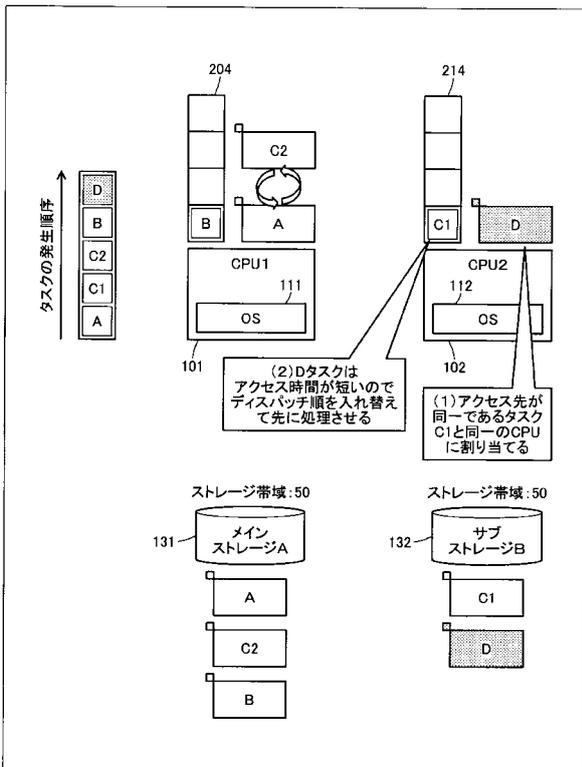
【 図 6 - 3 】



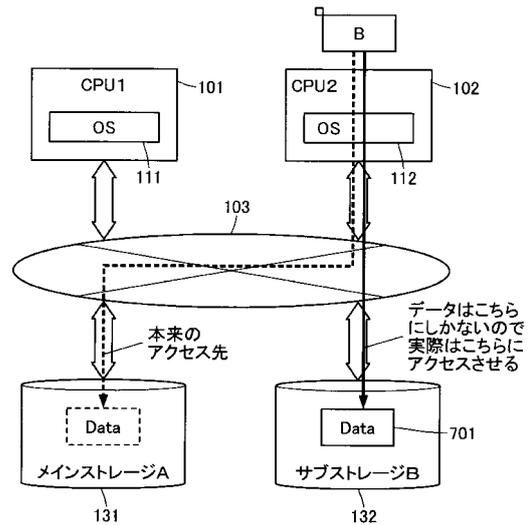
【 図 6 - 4 】



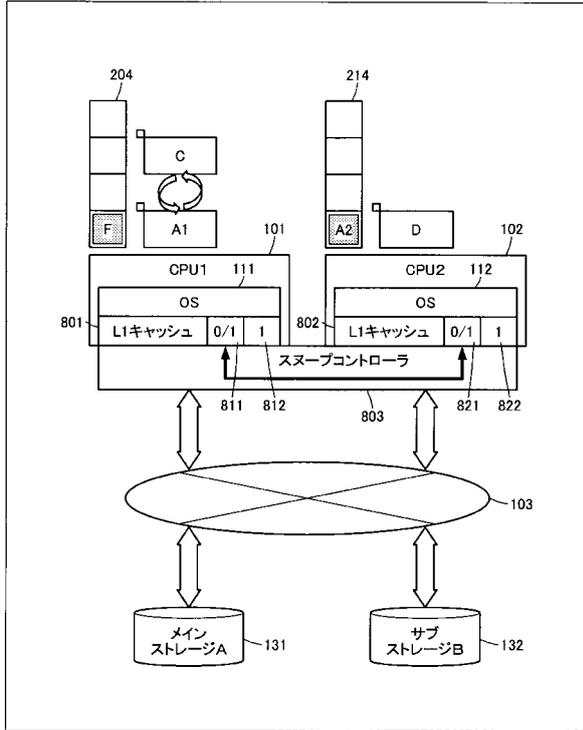
【 図 6 - 5 】



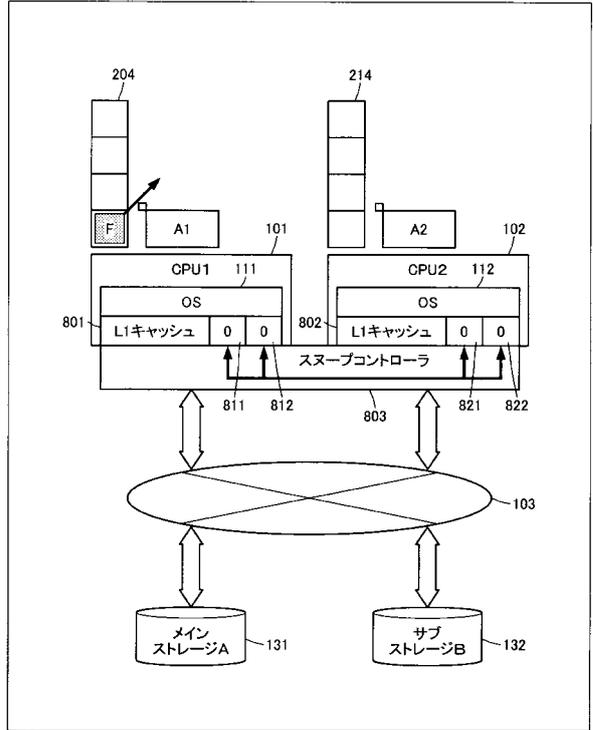
【 図 7 】



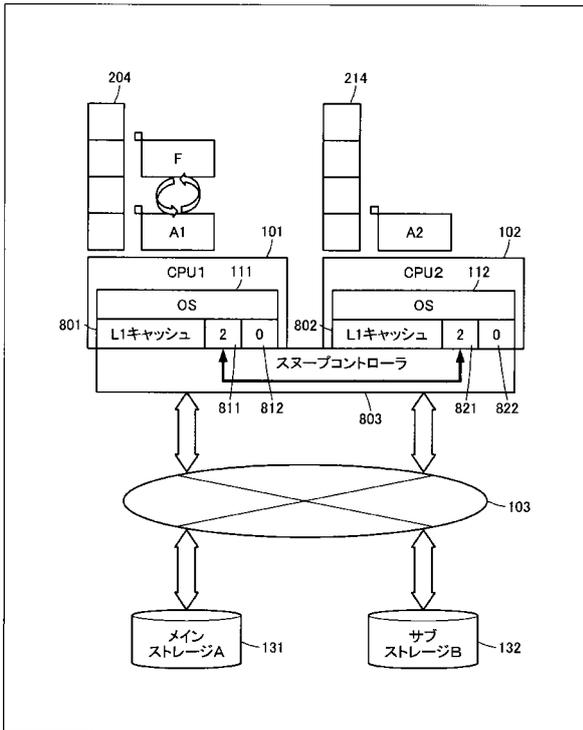
【図 8 - 1】



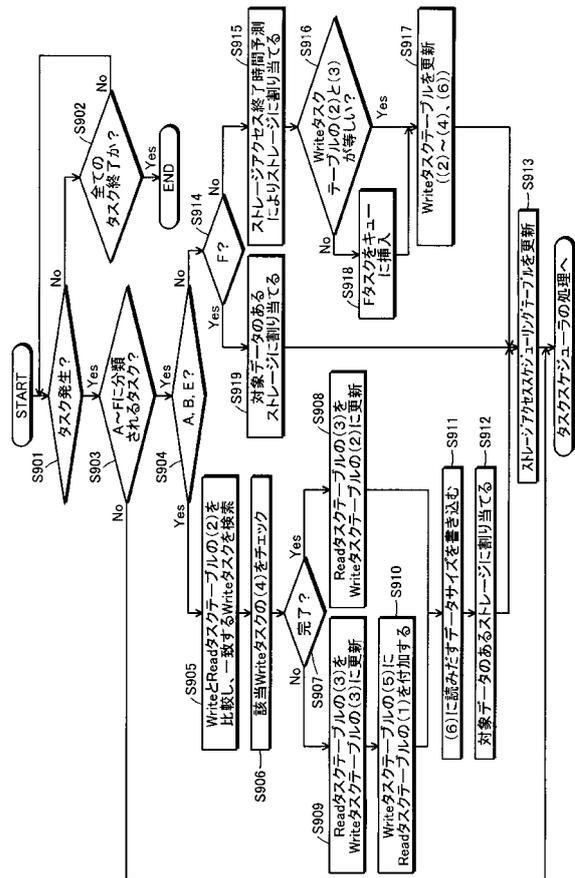
【図 8 - 2】



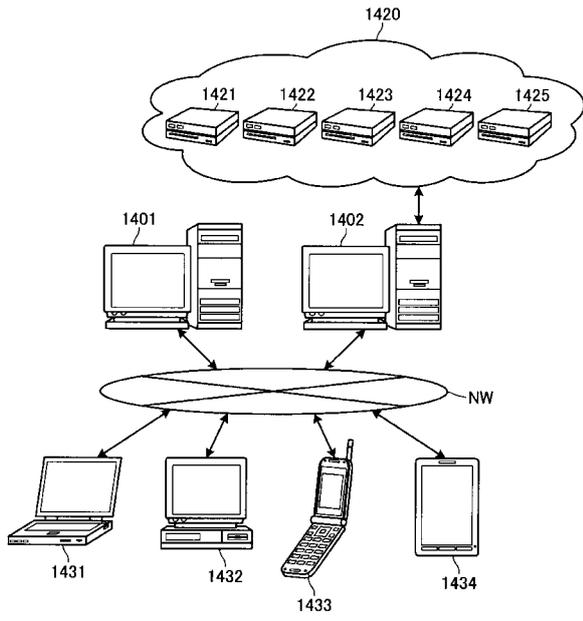
【図 8 - 3】



【図 9】



【 図 1 4 】



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/064841

A. CLASSIFICATION OF SUBJECT MATTER G06F3/06(2006.01)i, G06F12/00(2006.01)i, G06F13/10(2006.01)i, G06F15/17(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06F3/06, G06F12/00, G06F13/10, G06F15/17 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2011 Kokai Jitsuyo Shinan Koho 1971-2011 Toroku Jitsuyo Shinan Koho 1994-2011 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2007-249729 A (Hitachi, Ltd.), 27 September 2007 (27.09.2007), paragraphs [0011] to [0016], [0025], [0033] to [0106]; fig. 1 to 21 & US 2007/0220514 A1 & EP 1835389 A2	1-14
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 13 September, 2011 (13.09.11)		Date of mailing of the international search report 27 September, 2011 (27.09.11)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

国際調査報告		国際出願番号 PCT/JP2011/064841									
A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. G06F3/06(2006.01)i, G06F12/00(2006.01)i, G06F13/10(2006.01)i, G06F15/17(2006.01)i											
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G06F3/06, G06F12/00, G06F13/10, G06F15/17											
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2011年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2011年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2011年</td> </tr> </table>				日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2011年	日本国実用新案登録公報	1996-2011年	日本国登録実用新案公報	1994-2011年
日本国実用新案公報	1922-1996年										
日本国公開実用新案公報	1971-2011年										
日本国実用新案登録公報	1996-2011年										
日本国登録実用新案公報	1994-2011年										
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)											
C. 関連すると認められる文献											
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号									
X	JP 2007-249729 A (株式会社日立製作所) 2007.09.27, 段落 0011-00016, 0025, 0033-0106, 図 1-21 & US 2007/0220514 A1 & EP 1835389 A2	1-14									
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。											
* 引用文献のカテゴリー		の日の後に公表された文献									
「A」特に関連のある文献ではなく、一般的技術水準を示すもの		「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの									
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの		「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの									
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)		「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの									
「O」口頭による開示、使用、展示等に言及する文献		「&」同一パテントファミリー文献									
「P」国際出願日前で、かつ優先権の主張の基礎となる出願											
国際調査を完了した日 13.09.2011		国際調査報告の発送日 27.09.2011									
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 菅原 浩二	5 I 9460								
		電話番号 03-3581-1101 内線 3565									

フロントページの続き

(72)発明者 大舘 尚記

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 大友 俊也

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

Fターム(参考) 5B061 BA01 GG11

(注)この公表は、国際事務局(WIPO)により国際公開された公報を基に作成したものである。なおこの公表に係る日本語特許出願(日本語実用新案登録出願)の国際公開の効果は、特許法第184条の10第1項(実用新案法第48条の13第2項)により生ずるものであり、本掲載とは関係ありません。