

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2012年2月9日(09.02.2012)

PCT

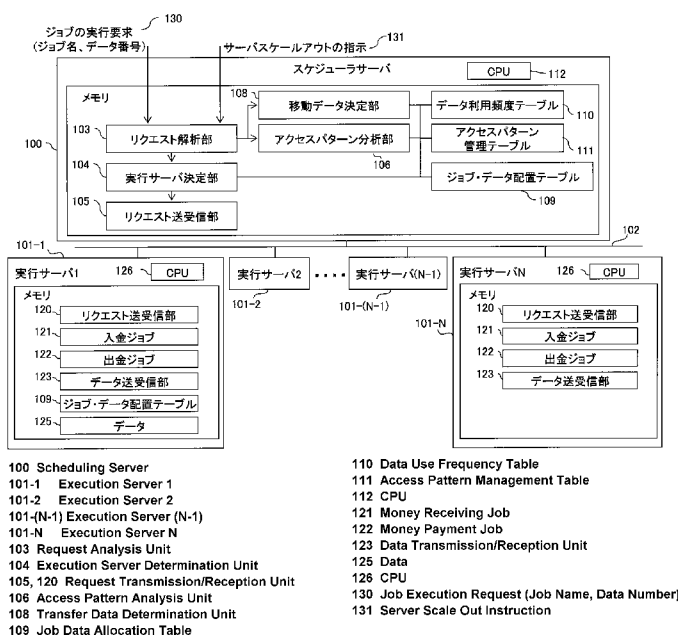
(10) 国際公開番号
WO 2012/017699 A1

- (51) 国際特許分類:
G06F 12/00 (2006.01) G06F 9/50 (2006.01)
- (21) 国際出願番号: PCT/JP2011/051347
- (22) 国際出願日: 2011年1月25日(25.01.2011)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2010-176992 2010年8月6日(06.08.2010) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社日立製作所 (HITACHI, LTD.) [JP/JP]; 〒1008280 東京都千代田区丸の内一丁目6番6号 Tokyo (JP).
- (72) 発明者: および
- (75) 発明者/出願人 (米国についてのみ): 渡辺 聡 (WATANABE, Satoru) [JP/JP]; 〒1858601 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内 Tokyo (JP).
- (74) 代理人: 後藤 政喜 (GOTO, Masaki); 〒1000013 東京都千代田区霞が関三丁目3番1号尚友会館 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): 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, ID, IL, IN, IS, 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.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (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), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- 添付公開書類:
— 国際調査報告 (条約第21条(3))
— 補正された請求の範囲 (条約第19条(1))

(54) Title: COMPUTER SYSTEM AND DATA MANAGEMENT METHOD

(54) 発明の名称: 計算機システム、及び、データ管理方法

[図1]



(57) Abstract: Provided is a computer system provided with: a plurality of execution servers that execute a plurality of jobs; and a scheduling server that is connected to the execution servers. The execution servers hold a plurality of data that are processed by the jobs, and the scheduling server generates a plurality of first sets that each contain a plurality of data processed by the plurality of jobs that are continuously executed, extracts, from the plurality of first sets, second sets containing data held by one of the execution servers and data held by another execution server, and determines data contained in the extracted second sets to be data to be transferred to a newly added execution server.

(57) 要約: 複数のジョブを実行する複数の実行サーバと、前記実行サーバに接続されるスケジューラサーバとを備える計算機システムであって、前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、前記スケジューラサーバは、連続して実行される前記複数のジョブによって処理される複数のデータを各々含む複数の第1の組を生成し、一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記複数の第1の組から抽出し、前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定する。

明 細 書

発明の名称： 計算機システム、及び、データ管理方法

参照による取り込み

- [0001] 本出願は、平成２２年（２０１０年）８月６日に出願された日本特許出願特願２０１０－１７６９９２の優先権を主張し、その内容を参照することにより、本出願に取り込む。

技術分野

- [0002] 本発明は、計算機システムに関し、特に、計算機システムに新たにサーバを追加する場合に、移動するデータを決定する計算機システムに関する。

背景技術

- [0003] 現在、スケジューラとジョブを実行するためのサーバとを用いてジョブを実行するシステムが、広く用いられている。このシステムは、ジョブを実行する要求を受信し、受信した要求を、ジョブを実行するためのサーバ（以降、実行サーバと記載）に送信するスケジューラと、実行サーバとを備える。
- [0004] このシステムのスケジューラは、ジョブリクエスト、すなわち、ジョブに関する要求を解析し、要求されたジョブを実行する実行サーバを決定し、ジョブを実行する要求を、決定された実行サーバに送信する（例えば、特許文献１参照）。
- [0005] また、ジョブの実行サーバのメモリにデータを格納する技術が提案されている（例えば、特許文献２参照）。特許文献２には、実行サーバのメモリ（主記憶装置）に、ジョブで使用するデータを格納するシステムが提案されている。
- [0006] さらに、スケジューラと実行サーバとを備えるシステムにおいて、システムの処理容量が不足する場合、実行サーバを追加することによって、システムの処理容量を大きくするシステムが提案されている。実行サーバを追加することを、実行サーバのスケールアウトと呼ぶ。このように、スケールアウトによって処理容量を増大させるシステムが提案されている（例えば、特許

文献 3 参照)。

先行技術文献

特許文献

[0007] 特許文献1：特開 2 0 0 8 － 1 5 2 6 1 8 号公報

特許文献2：特開平 7 － 3 3 4 4 0 2 号公報

特許文献3：特開 2 0 0 9 － 1 1 0 1 2 9 号公報

発明の概要

発明が解決しようとする課題

[0008] 実行サーバのメモリにデータを保持するシステムにおいて、実行サーバをスケールアウトする場合、既存の実行サーバに保持されるデータの一部は、追加する実行サーバに移される必要がある。

[0009] この際、実行サーバにかけられる負荷に偏りが生じると、負荷が偏った実行サーバは、ネック、すなわち処理の遅延の原因になり、全体的な処理が滞る可能性がある。このため、実行サーバをスケールアウトする場合、実行サーバにかかる負荷が平準化されるように、データが移動される必要がある。

[0010] また、システムにおいて実行されるジョブには、一つのデータを引数として処理するジョブだけでなく、複数のデータを処理するジョブが存在する。

[0011] 例えば、銀行口座の入出金システムは、口座番号と口座残高とのデータが、相互に対応して管理される。そして、このような入出金システムでは、一方の口座から出金し、他の口座に入金するといったジョブが実行される。このようなジョブは、出金された口座の口座番号及び口座残高のデータと、入金された口座の口座番号及び口座残高のデータとの複数のデータを、処理する。

[0012] 複数のデータを処理するジョブを、以降、連携ジョブと記載する。

[0013] 連携ジョブによって処理されるデータが異なる実行サーバに格納される場合、一つのサーバは、先行する処理（例えば、出金処理）を終了させた後、先行する処理が終了した旨を他の実行サーバに送信する。そして、他の実行

サーバは、その旨を受信した後、後続の処理（例えば、入金処理）を実行する。

[0014] このように、連携ジョブによって処理されるデータが異なる実行サーバに格納される場合、一方の実行サーバから他方の実行サーバに通信が必要である。この通信のオーバーヘッドによって、システムの性能が低下するため、連携ジョブによって処理されるデータは、同じ実行サーバに配置されることが望ましい。

[0015] 本発明の一つ目の課題は、実行サーバをスケールアウトする場合、実行サーバ間の負荷の平準化、連携ジョブの通信におけるオーバーヘッドの削減、又はその両方を行うため、追加する実行サーバに移動するデータを、適切に決定することである。

[0016] また、連携ジョブではないジョブも、特定のパターンによってデータにアクセスする場合がある。例えば、あるデータ（データ 1）を処理した後に、別のデータ（データ 2）を処理する可能性が高い、といった特定のパターンによって、データにアクセスするジョブが存在する。

[0017] 前述の例において、データ 1 とデータ 2 とが異なる実行サーバに格納される場合、データを処理するための順序を保証するため、スケジューラは、データ 1 の処理が終了するのを待ち、データ 1 の処理が終了後、実行サーバにデータ 2 の処理を指示する。

[0018] 一方、前述の例において、データ 1 とデータ 2 とが同じ実行サーバに格納される場合、実行サーバが指示を受けた順にジョブを実行するため、スケジューラは、データ 1 の処理を待たずに、データ 2 の処理を指示できる。そのため、スケジューラによる処理を低減するためには、データ 1 とデータ 2 とが同じ実行サーバに配置されることが望ましい。

[0019] 本発明の二つ目の課題は、実行サーバをスケールアウトする場合、実行サーバの負荷の平準化、データへのアクセスパターンの相違によるスケジューラの処理負担の低減、又は、その両方を行うため、追加する実行サーバに移動するデータを、適切に決定することである。

課題を解決するための手段

- [0020] 本願において開示される発明の代表的な一例を示せば以下の通りである。
- すなわち、複数のジョブを実行する複数の実行サーバと、前記実行サーバに接続されるスケジューラサーバとを備える計算機システムであって、前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、前記スケジューラサーバは、連続して実行される前記複数のジョブによって処理される複数のデータを各々含む複数の第1の組を生成し、一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記複数の第1の組から抽出し、前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定する。

発明の効果

- [0021] 本発明の代表的な実施形態によると、実行サーバ間の負荷を平準化し、通信におけるオーバーヘッドを低減できる。

図面の簡単な説明

- [0022] [図1]本発明の実施形態の計算機システムの構成を示すブロック図である。
- [図2]本発明の実施形態のデータを示す説明図である。
- [図3]本発明の実施形態のジョブ・データ配置テーブルに格納される情報を示す説明図である。
- [図4]本発明の実施形態のデータ利用頻度テーブルを示す説明図である。
- [図5]本発明の実施形態のアクセスパターン管理テーブルを示す説明図である。
- [図6]、本発明の実施形態の実行サーバに移動するデータを移動データ決定部が決定する手順を示すフローチャートである。
- [図7]本発明の実施形態の実行サーバ負荷リストを示す説明図である。
- [図8]本発明の実施形態の実行サーバに移動するデータを、通信頻度を優先して決定する手順を示すフローチャートである。

発明を実施するための形態

- [0023] 図1は、本発明の実施形態の計算機システムの構成を示すブロック図である。
- [0024] 図1に示す計算機システムは、スケジューラサーバ100と、1台以上の実行サーバ101（実行サーバ101-1、101-2、・・・、101-（N-1））とを備える。また、101-Nは、本実施形態における計算機システムに追加される実行サーバ101である。スケジューラサーバ100と実行サーバ101とは、ネットワーク102によって接続される。
- [0025] スケジューラサーバ100は、CPU112及びメモリを備える計算機である。スケジューラサーバ100は、リクエスト解析部103、実行サーバ決定部104、リクエスト送受信部105、アクセスパターン分析部106、移動データ決定部108、ジョブ・データ配置テーブル109、データ利用頻度テーブル110、及び、アクセスパターン管理テーブル111を備える。
- [0026] スケジューラサーバ100に備わるリクエスト解析部103、実行サーバ決定部104、リクエスト送受信部105、アクセスパターン分析部106、及び、移動データ決定部108は、スケジューラサーバ100のメモリに格納され、CPU112によって実行されるプログラムによって実装される。また、ジョブ・データ配置テーブル109、データ利用頻度テーブル110、及び、アクセスパターン管理テーブル111は、スケジューラサーバ100のメモリに格納されるテーブルである。
- [0027] スケジューラサーバ100のメモリは、プログラム及びデータ等を格納することができる。また、一時的に、データ等を格納することができる。
- [0028] リクエスト解析部103は、管理者若しくはユーザ、又は、別のスケジューラサーバ100等によって送られる要求等を解析し、要求等に含まれる情報を各プログラムに送るためのプログラムによって実装される。アクセスパターン分析部106は、各実行サーバ101に格納されたデータ125の利用頻度を測定するためのプログラムによって実装される。

- [0029] 実行サーバ決定部 104 は、ジョブの実行要求 130 によって要求されたジョブを、実行する実行サーバ 101 を決定するプログラムによって実装される。リクエスト送受信部 105 は、実行サーバ決定部 104 によって決定された実行サーバ 101 に、実行を要求されたジョブの情報等を送るプログラムによって実装される。また、リクエスト送受信部 105 は、実行サーバ 101 からジョブの実行結果を受信する。
- [0030] 移動データ決定部 108 は、実行サーバ 101 が追加される際に、既存の実行サーバ 101 から追加された実行サーバ 101 に移動するデータ 125 を、決定するプログラムによって実装される。
- [0031] 実行サーバ 101 は、CPU 126 及びメモリを備える計算機である。実行サーバ 101-1 ~ 101-(N-1) は、リクエスト送受信部 120、入金ジョブ 121、出金ジョブ 122、データ送受信部 123、ジョブ・データ配置テーブル 109、及び、データ 125 を備える。また、実行サーバ 101-N は、リクエスト送受信部 120、入金ジョブ 121、出金ジョブ 122、及び、データ送受信部 123 を備える。
- [0032] 実行サーバ 101 に備わるリクエスト送受信部 120 及びデータ送受信部 123 は、実行サーバ 101 のメモリに格納され、CPU 126 によって実行されるプログラムによって実装される。また、入金ジョブ 121、及び、出金ジョブ 122 は、CPU 126 によって実行されるプログラムを含むジョブである。
- [0033] また、実行サーバ 101-1 ~ 101-(N-1) に備わるジョブ・データ配置テーブル 109 及びデータ 125 (データ 125-1 ~ 125-(N-1)) は、実行サーバ 101 のメモリに格納されるテーブル及びデータである。
- [0034] リクエスト送受信部 120 は、スケジューラサーバ 100 から送信される実行を要求されたジョブの情報等を受信し、ジョブの実行結果をスケジューラサーバ 100 に送信するプログラムによって実装される。データ送受信部 123 は、実行サーバ 101 間で、データ 125 を送受信するためのプログ

ラムによって実装される。

[0035] 入金ジョブ 121 は、データ 125 に格納される値に、引数によって与えられた数を加算するジョブである。出金ジョブ 122 は、データ 125 に格納される値から、引数によって与えられた数を減算するジョブである。

[0036] 実行サーバ 101-N は、本実施形態の計算機システムに新たに追加される実行サーバ 101、すなわちスケールアウトされる実行サーバ 101 である。実行サーバ 101-N は、本実施形態の計算機システムに追加される前の時点において、ジョブ・データ配置テーブル 109 及びデータ 125 を備えていない。本実施形態のスケジューラサーバ 100 は、実行サーバ 101-N を追加するために、実行サーバ 101-N に格納されるデータ 125 を決定する。

[0037] なお、図 1 に示すスケジューラサーバ 100 と実行サーバ 101 とは、各々別の計算機によって実装されるが、本実施形態のスケジューラサーバ 100 及び実行サーバ 101 は、仮想的な計算機によって実装されてもよい。仮想的な計算機は、少なくとも一つの物理的な CPU と、少なくとも一つの物理的なメモリとによって実装される。

[0038] また、図 1 に示す実行サーバ 101 は、入金ジョブ 121 及び出金ジョブ 122 を保持するが、本実施形態の実行サーバ 101 は、入金ジョブ 121 及び出金ジョブ 122 に限らず、プログラムによって実行されるジョブであれば、いかなるジョブを保持してもよい。

[0039] 図 2 は、本発明の実施形態のデータ 125 を示す説明図である。

[0040] データ 125 は、実行サーバ 101 がジョブによって処理するデータである。

[0041] データ 125 は、口座番号 1251 及び口座残高 1252 を含む。口座番号 1251 は、各データ 125 を一意に示す識別子を示し、口座残高 1252 は各データ 125 の値を示す。

[0042] 実行サーバ 101-1 ~ 101-(N-1) には、異なるデータ 125-1 ~ 125-(N-1) が格納される。例えば、実行サーバ 101-1 には

、口座番号 1251 が 1 から 1000 を示すデータ 125 が格納され、実行サーバ 101-2 には、口座番号 1251 が 1001 から 2000 を示すデータ 125 が格納される。さらに、実行サーバ 101- (N-1) には、口座番号 1251 が $1000 \times (N-1) + 1$ から $1000 \times N$ のデータが格納される。

[0043] 図 3 は、本発明の実施形態のジョブ・データ配置テーブル 109 を示す説明図である。

[0044] ジョブ・データ配置テーブル 109 は、スケジューラサーバ 100 及び実行サーバ 101 によって保持される。ジョブ・データ配置テーブル 109 は、実行サーバ 101 と実行サーバ 101 が実行するジョブの種類と実行サーバ 101 が処理するデータとの対応関係を示す。ジョブ・データ配置テーブル 109 は、実行サーバ 101 がスケールアウトする際に更新される。更新されたジョブ・データ配置テーブル 109 は、スケジューラサーバ 100 によって、各実行サーバ 101 に送信される。

[0045] ジョブ・データ配置テーブル 109 は、実行サーバ 1091、ジョブ 1092 及び口座番号 1093 を含む。実行サーバ 1091 は、実行サーバ 101 を一意に示す識別子である。ジョブ・データ配置テーブル 109 は、実行サーバ 101 ごとにエントリを含む。

[0046] ジョブ 1092 は、実行サーバ 1091 が示す実行サーバ 101 において実行できるジョブの種類を含む。また、口座番号 1093 は、データ 125 の識別子、すなわち図 2 に示す口座番号 1251 を含む。口座番号 1093 は、複数の識別子を含んでよい。

[0047] 例えば、図 3 に示すとおり、実行サーバ 1091 が実行サーバ 101-1 ~ 101- (N-1) であり、ジョブ 1092 が“入金ジョブ、出金ジョブ”である場合、ジョブ・データ配置テーブル 109 は、実行サーバ 101-1 ~ 101- (N-1) が、入金ジョブ及び出金ジョブを実行可能であることを示す。

[0048] ここで、ジョブ・データ配置テーブル 109 に示される入金ジョブとは、

入力された引数をデータ 125 の口座残高 1252 に加算するジョブであり、入金ジョブ 121 によって実行されるジョブである。また、ジョブ・データ配置テーブル 109 に示される出金ジョブとは、入力された引数をデータ 125 の口座残高 1252 から減算するジョブであり、出金ジョブ 122 によって実行されるジョブである。

[0049] また、実行サーバ 1091 が実行サーバ 101-1 を示し、口座番号 1093 が 1~1000 を示す場合、実行サーバ 101-1 には、図 2 に示す口座番号 1251 が 1~1000 であるデータ 125 が格納される。また、実行サーバ 1091 がサーバ実行サーバ 101-2 を示し、口座番号 1093 が 1001~2000 を示す場合、実行サーバ 101-2 には、図 2 に示す口座番号 1251 が 1001~2000 であるデータ 125 が格納される。

[0050] また、実行サーバ 1091 が実行サーバ 101-(N-1) を示し、口座番号 1093 が $1000 \times (N-1) + 1 \sim 1000 \times N$ を示す場合、実行サーバ 101-(N-1) には、口座番号 1251 が $1000 \times (N-1) + 1 \sim 1000 \times N$ であるデータが格納される。

[0051] なお、図 3 のジョブ 1092 の各エントリには、入金ジョブ 121 と出金ジョブ 122 が一つずつ格納されているが、複数の入金ジョブ 121 又は出金ジョブ 122 が格納されてもよい。

[0052] 図 4 は、本発明の実施形態のデータ利用頻度テーブル 110 を示す説明図である。

[0053] データ利用頻度テーブル 110 は、スケジューラサーバ 100 によって保持される。データ利用頻度テーブル 110 は、データ 125 の利用頻度、すなわち、データ 125 の各エントリがジョブによって処理される頻度を示す。

[0054] データ利用頻度テーブル 110 は、口座番号 1101 及び利用頻度 1102 を含む。口座番号 1101 は、データ 125 の識別子を示し、図 2 に示す口座番号 1251 と、図 3 に示す口座番号 1093 の値とに相当する。利用頻度 1102 は、例えば、口座番号 1101 が示すデータ 125 が、ジョブ

によって1分あたりに処理される頻度を示す。

[0055] なお、図4に示すデータ利用頻度テーブル110は、口座番号1101ごとに利用頻度1102を格納するテーブルであるが、データ利用頻度テーブル110のエントリの個数を減らすため、データ125を所定の数ごとに区切ることによって、複数のデータ利用頻度テーブル110が生成されてもよい。すなわち、例えば、口座番号1101が1～100、口座番号1101が実行サーバ101～200であるエントリを各々まとめることによって、複数のデータ利用頻度テーブル110に利用頻度1102が格納されてもよい。

[0056] 図5は、本発明の実施形態のアクセスパターン管理テーブル111を示す説明図である。

[0057] アクセスパターン管理テーブル111は、口座番号の組1111及び利用頻度1112を含む。口座番号の組1111には、一つのデータ125が処理された後に、別のデータ125が処理される場合のデータ125の組を示す。口座番号の組1111に示される値は、図2に示す口座番号1251と、図3に示す口座番号1093の値と、図4に示す口座番号1101に相当する。ここで、口座番号の組1111に含まれるデータ125は、連携ジョブによって連続して処理されるデータ125の組だけでなく、通常のジョブが実行される際に、ある一定期間内に処理される可能性が高いデータ125の組であってもよい。

[0058] 利用頻度1112は、口座番号の組1111に示されるデータ125の組が、処理される頻度を示す。利用頻度1112は、例えば、口座番号の組1111が示すデータ125の組が、ジョブによって1分あたりに処理される頻度を示す。

[0059] なお、図5に示すアクセスパターン管理テーブル111は、口座番号の組1111ごとに利用頻度1112を格納するテーブルであるが、アクセスパターン管理テーブル111のエントリの個数を減らすため、データを所定の数ごとに区切り、複数のアクセスパターン管理テーブル111が作成されて

もよい。

[0060] すなわち、例えば、口座番号の組 1 1 1 1 に含まれるデータ 1 2 5 の値のうち一つ目の値が、1 ~ 1 0 0 であり、口座番号の組 1 1 1 1 に含まれるデータ 1 2 5 の値のうち二つ目の値が、1 0 0 0 ~ 1 1 0 0 である組といったように、エントリを各々まとめることによって、複数のアクセスパターン管理テーブル 1 1 1 に利用頻度 1 1 1 2 が格納されてもよい。

[0061] なお、図 5 の口座番号の組 1 1 1 1 には、二つの口座番号 1 1 0 1 の値のみが格納されているが、三つ以上の口座番号 1 1 0 1 の値が格納されてもよい。

[0062] また、本実施形態のデータ 1 2 5 は、口座番号 1 2 5 1 及び口座残高 1 2 5 2 に限られず、各データ 1 2 5 を一意に識別する識別子と、値とを含めば、いかなる種類のデータでもよい。このため、ジョブ・データ配置テーブル 1 0 9、データ利用頻度テーブル 1 1 0、及び、アクセスパターン管理テーブル 1 1 1 の列名は、データ 1 2 5 に格納されるデータの種類の従い、異なる名称でもよい。

[0063] (ジョブのスケジュール)

スケジューラサーバ 1 0 0 は、管理者若しくはユーザ、又は、別のスケジューラサーバ 1 0 0 等によって、ジョブの実行要求 1 3 0 を送られた後、ジョブの実行要求 1 3 0 によって要求されたジョブを実行サーバ 1 0 1 に実行させる。ジョブの実行要求 1 3 0 には、実行されるジョブを一意に示す情報と、実行されるジョブが処理するデータ 1 2 5 を一意に示す情報とが含まれる。また、ジョブの実行要求 1 3 0 には、必要に応じて、ジョブに入力される引数等が含まれる。

[0064] 例えば、本実施形態におけるジョブの実行要求 1 3 0 には、「口座番号が 1 番の口座から、1 円を出金する出金ジョブの実行」を要求する情報が含まれる。この場合、“1 番”が、データ 1 2 5 を一意に示す情報を示し、“出金ジョブ”が実行されるジョブを一意に示す情報を示し、“1 円”が引数を示す。なお、ジョブの実行要求 1 3 0 に含まれる値は、データ 1 2 5 及びジョブを一

意に示せばよく、識別子でも名称でもよい。

[0065] また、ジョブの実行要求 130 が、連携ジョブの実行の要求を示す情報を含む場合、ジョブの実行要求 130 には、ジョブとデータ 125 を示す情報が、それぞれ、二つ以上含まれる。例えば、ジョブの実行要求 130 には、「口座番号が 1 番である口座から、1 円を出金する出金ジョブの実行と、口座番号が 2 番である口座に、1 円を入金する入金ジョブの実行」を要求する情報が含まれる。

[0066] ジョブの実行要求 130 は、前述のような情報を含めば、いかなる方法によって指定されてもよい。ジョブの実行要求 130 は、例えば、CSV (Comma Separated Values) 形式によって、記載されてもよい。

[0067] スケジューラサーバ 100 がジョブの実行要求 130 を受信した後、リクエスト解析部 103 は、ジョブの実行要求 130 に含まれる、ジョブ名とデータ 125 を示す識別子等とを抽出し、アクセスパターン分析部 106 及び実行サーバ決定部 104 に送る。

[0068] アクセスパターン分析部 106 は、例えば 1 分等の一定時間ごとに、すべての実行サーバ 101 に格納されたデータ 125 の利用頻度を測定する。そして、データ利用頻度テーブル 110 の利用頻度 1102 及びアクセスパターン管理テーブル 111 の利用頻度 1112 を最新の値に更新する。

[0069] 実行サーバ 101 は、ジョブを実行する際に、処理したデータ 125 とジョブとを示す情報をスケジューラサーバ 100 に送信する。このため、スケジューラサーバ 100 は、実行サーバ 101 から情報を送信されることによって、データ 125 の利用頻度を測定する。また、アクセスパターン分析部 106 は、一つのデータ 125 が処理された後に、他のデータ 125 が処理される頻度を、一定時間測定する。

[0070] リクエスト解析部 103 によってジョブとデータ 125 を示す情報とを送られた後、アクセスパターン分析部 106 は、リクエスト解析部 103 から受信したデータ 125 を示す情報に基づいて、一定時間、各実行サーバ 10

1に格納されたデータ125の利用頻度を測定する。そして、測定された利用頻度によって、データ利用頻度テーブル110の利用頻度1102、及び、アクセスパターン管理テーブル111の利用頻度1112を更新する。

[0071] 具体的に、アクセスパターン分析部106は、「口座番号が1番の口座から、1円を出金する出金ジョブの実行」を要求する情報を受信した場合、口座番号1251が1であるデータ125の利用頻度を、例えば1分等の一定時間測定する。そして、測定された利用頻度によって、データ利用頻度テーブル110を更新する。

[0072] また、アクセスパターン分析部106は、「口座番号が1番の口座から1円を出金する出金ジョブの実行と、口座番号が2番の口座に1円を入金する入金ジョブの実行」を要求する情報を受信した場合、口座番号1251が1であるデータ125と口座番号1251が2であるデータ125との組について、一定時間利用頻度を測定し、アクセスパターン管理テーブル111を更新する。

[0073] アクセスパターン分析部106は、連携ジョブの実行要求を受信した場合以外にも、アクセスパターン管理テーブル111を更新してよい。例えば、アクセスパターン分析部106は、口座番号1251が1であるデータ125の処理に続いて、口座番号1251が2であるデータ125が処理されることを、実行サーバ101から送信される頻度を測定する。

[0074] そして、アクセスパターン分析部106は、実行サーバ101から送信された頻度が、予め決められた頻度を上回った場合、口座番号1251が1であるデータ125と、口座番号1251が2であるデータ125の組とを、アクセスパターン管理テーブル111の口座番号の組1111に格納し、実行サーバ101から送信された頻度を利用頻度1112に格納する。

[0075] なお、アクセスパターン分析部106は、データ利用頻度テーブル110及びアクセスパターン管理テーブル111を更新する場合、過去に測定された利用頻度1102又は利用頻度1112と、新たに測定された利用頻度1102又は利用頻度1112との平均を算出し、算出された平均値を利用頻

度 1 1 0 2 又は利用頻度 1 1 1 2 に格納してもよい。また、過去に測定された利用頻度 1 1 0 2 又は利用頻度 1 1 1 2 を削除し、新たに測定された利用頻度 1 1 0 2 又は利用頻度 1 1 1 2 を格納してもよい。

[0076] リクエスト解析部 1 0 3 によってジョブとデータ 1 2 5 を示す情報とを送られた後、実行サーバ決定部 1 0 4 は、ジョブ・データ配置テーブル 1 0 9 を参照し、リクエスト解析部 1 0 3 によって送られた情報が示すジョブ 1 0 9 2 と口座番号 1 0 9 3 とを抽出する。そして、抽出されたジョブ 1 0 9 2 と口座番号 1 0 9 3 とを含む実行サーバ 1 0 9 1 を、ジョブの実行要求 1 3 0 によって要求されたジョブを実行する実行サーバ 1 0 1 に決定する。そして、決定された実行サーバ 1 0 1 と実行するジョブとを、リクエスト送受信部 1 0 5 に送信する。

[0077] リクエスト送受信部 1 0 5 は、実行サーバ決定部 1 0 4 によって決定された実行サーバ 1 0 1 のリクエスト送受信部 1 2 0 に、ジョブを実行する要求を送信する。この際、リクエスト送受信部 1 2 0 に送られる要求には、ジョブの実行要求 1 3 0 に含まれる情報が含まれる。

[0078] リクエスト送受信部 1 2 0 が、リクエスト送受信部 1 0 5 によってジョブを実行する要求を受信した後、実行サーバ 1 0 1 の CPU 1 2 6 は、要求に含まれる情報によって、ジョブを実行する。CPU 1 2 6 が、ジョブを実行した後、リクエスト送受信部 1 2 0 は、ジョブの実行結果をリクエスト送受信部 1 0 5 に送信する。

[0079] 前述のように、本実施形態の計算機システムにジョブの実行要求 1 3 0 が入力された後、実行サーバ 1 0 1 によって、要求されたジョブが実行される。ここで、入力されたジョブの実行要求 1 3 0 が連携ジョブの実行を要求する場合、複数の実行サーバ 1 0 1 によって実行される場合がある。

[0080] 例えば、ジョブの実行要求 1 3 0 が、「口座番号が 1 番である口座から 1 円を出金する出金ジョブの実行と、口座番号が 1 0 0 1 番である口座に 1 円を入金する入金ジョブの実行」を要求する場合、スケジューラサーバ 1 0 0 は、まず、前述の手順によって、一つ目のジョブを実行する実行サーバ 1 0

1を決定する。ここで、実行サーバ決定部104が一つ目のジョブを実行する実行サーバ101を実行サーバ101-1に決定した場合、リクエスト送受信部105は、実行サーバ101-1に出金ジョブ122を実行する要求を送信する。

[0081] その後、実行サーバ101-1のCPU126が出金ジョブ122を終了した後、実行サーバ101-1のCPU126は、ジョブ・データ配置テーブル109を参照し、二つ目のジョブを実行する実行サーバ101を決定する。ここで、CPU126が二つ目のジョブ実行サーバ101に実行サーバ101-2を決定した場合、実行サーバ101-1のCPU126は、実行サーバ101-2にジョブの実行の要求を送信する。この際、実行サーバ101-1のリクエスト送受信部120、及び、実行サーバ101-2のリクエスト送受信部120は、ジョブの実行の要求を送受信する。

[0082] (スケールアウトによって移動するデータ125の決定)

本実施形態の計算機システムに、実行サーバ101-Nが追加される場合、管理者等によってスケジューラサーバ100にサーバスケールアウトのコマンド131が入力される。

[0083] サーバスケールアウトのコマンド131には、通信頻度の閾値が含まれる。通信頻度の閾値とは、管理者等が、連携ジョブを実行することによって発生する実行サーバ101間の通信の頻度を、指定する場合に用いる値である。本実施形態の通信頻度は、既存の実行サーバ101間、すなわち、実行サーバ101-1~101-(N-1)間において発生する通信頻度である。なお、通信頻度の閾値は、サーバスケールアウトのコマンド131によって、管理者等から指定されてもよいし、スケジューラサーバ100にあらかじめ格納されていてもよい。

[0084] スケジューラサーバ100の移動データ決定部108は、通信頻度の閾値を指定されることによって、指定された通信頻度の閾値以下になるように、実行サーバ101-Nへ移動するデータ125を決定する。通信頻度の閾値は、例えば、1分あたりに発生する通信の回数によって指定される。

- [0085] また、サーバスケールアウトのコマンド 131 には、平準化優先、通信削減優先、又は、その両方のいずれかを示す情報が含まれる。平準化優先とは、実行サーバ 101-N が追加される際、データ 125 の利用頻度に基づいて、すべての実行サーバ 101 の負荷が平準化されるように、追加された実行サーバ 101-N にデータ 125 を移動する指示である。通信削減優先とは、連携ジョブによって処理されるデータ 125 が、同じ実行サーバ 101 に格納されるようにデータ 125 を移動する指示である。
- [0086] 本実施形態のサーバスケールアウトのコマンド 131 は、平準化優先又は通信削減優先のいずれかを、通信頻度の閾値によって示す。すなわち、通信頻度の閾値が 0 回／分に指定された場合、サーバスケールアウトのコマンド 131 は、通信削減優先を示し、通信頻度の閾値が無限大に指定された場合、サーバスケールアウトのコマンド 131 は、平準化優先を示す。
- [0087] また、通信頻度の閾値が、0 回／分と無限大との間の値である場合、サーバスケールアウトのコマンド 131 は、平準化優先及び通信削減優先の両方を示す。サーバスケールアウトのコマンド 131 は、通信頻度の閾値の大小によって、平準化優先又は通信削減優先のいずれを優先させるかを示す。
- [0088] 移動データ決定部 108 は、実行サーバ 101-N が追加される際、実行サーバ 101-N へ移動するデータ 125 を決定するためのプログラムによって実装される。
- [0089] サーバスケールアウトのコマンド 131 がスケジューラサーバ 100 に入力された場合、リクエスト解析部 103 は、サーバスケールアウトのコマンド 131 が示す情報を移動データ決定部 108 に送る。移動データ決定部 108 は、リクエスト解析部 103 からサーバスケールアウトのコマンド 131 が示す情報を受信した場合、サーバスケールアウトのコマンド 131 が示す情報に従って、新たに追加される実行サーバ 101-N に移動するデータ 125 を決定する。
- [0090] 以下に、100 が、追加された実行サーバ 101 に移動するデータ 125 を、決定する処理を示す。

- [0091] 図6は、本発明の実施形態の実行サーバ101-Nに移動するデータ125を移動データ決定部108が決定する手順を示すフローチャートである。
- [0092] 以下、図6のフローチャートによって、移動データ決定部108の処理を説明する。図6に示す移動データ決定部108の処理は、実行サーバ101間の通信頻度を削減するように移動するデータ125を決定するステップと、すべての実行サーバ101の負荷量を平準化するように移動するデータ125を決定するステップを含む。
- [0093] ステップ601、及び、ステップ603において、移動データ決定部108は、通信頻度を削減するように移動するデータ125を決定する。
- [0094] まず、移動データ決定部108は、サーバスケールアウトのコマンド131によって指定された通信頻度の閾値より、実行サーバ101の通信頻度が大きいかな否かを判定する(601)。具体的には、移動データ決定部108は、ジョブ・データ配置テーブル109を参照することによって、アクセスパターン管理テーブル111のエントリのうち、異なる実行サーバ101に格納されるデータ125の組を示す口座番号の組1111を含むエントリを抽出する。そして、抽出されたすべてのエントリの利用頻度1112を合計する。これによって、移動データ決定部108は、既存の実行サーバ101間、すなわち、実行サーバ101-1~101-(N-1)間の通信頻度を算出する。
- [0095] さらに移動データ決定部108は、ステップ601において、サーバスケールアウトのコマンド131によって指定された通信頻度の閾値と、算出された実行サーバ101-1~101-(N-1)間の通信頻度とを比較する。そして、移動データ決定部108は、ステップ601において、実行サーバ101-1~101-(N-1)間の通信頻度が、サーバスケールアウトのコマンド131によって指定された通信頻度の閾値以上であるかな否かを判定する。
- [0096] ステップ601において、実行サーバ101-1~101-(N-1)間の通信頻度が、サーバスケールアウトのコマンド131によって指定された

通信頻度の閾値以上であると判定された場合、実行サーバ１０１間の通信頻度を減らすため、移動データ決定部１０８は、ステップ６０３に移行する。

[0097] ステップ６０１の後、移動データ決定部１０８は、ステップ６０１において抽出された、異なる実行サーバ１０１に格納されるデータ１２５の組を示す口座番号の組１１１１のうち、最も利用頻度１１１２が大きいエントリを選択する。そして、移動データ決定部１０８は、選択されたエントリに基づいて、最も利用頻度１１１２が高く、異なる実行サーバ１０１に格納されるデータ１２５の組を、実行サーバ１０１－Ｎに移動することを決定する（６０３）。

[0098] また、移動データ決定部１０８は、ステップ６０３において、実行サーバ１０１－Ｎに移動することを決定されたデータ１２５の組と実行サーバ１０１－Ｎとが対応するように、ジョブ・データ配置テーブル１０９を更新する。

[0099] ステップ６０３の後、移動データ決定部１０８は、実行サーバ１０１－１～１０１－（Ｎ－１）間の通信頻度が、サーバスケールアウトのコマンド１３１によって指定された通信頻度の閾値以下になるまで、ステップ６０１及びステップ６０３を繰り返す。

[0100] ステップ６０１において、実行サーバ１０１－１～１０１－（Ｎ－１）間の通信頻度が、サーバスケールアウトのコマンド１３１によって指定された通信頻度の閾値より小さいと判定された場合、実行サーバ１０１間の通信頻度は問題ない頻度であり、実行サーバ１０１間の通信頻度を変更する必要はないため、移動データ決定部１０８は、ステップ６０４に移行する。

[0101] ステップ６０４からステップ６１３までにおいて、移動データ決定部１０８は、実行サーバ１０１の負荷を平準化するために、移動するデータ１２５を決定する。

[0102] ステップ６０１の後、移動データ決定部１０８は、実行サーバ１０１－１～１０１－（Ｎ－１）の実行サーバ１０１ごとの負荷量を算出し、実行サーバ負荷リスト１１３を生成する（６０４）。具体的には、移動データ決定部

108は、データ利用頻度テーブル110とジョブ・データ配置テーブル109とを参照することによって、データ利用頻度テーブル110の利用頻度1102を実行サーバ101ごとに加算する。これによって、実行サーバ101ごとの負荷量を算出する。そして、ステップ604において、移動データ決定部108は、算出された負荷量が多い順に、実行サーバ101を示す情報がソートされた、実行サーバ負荷リスト113を生成する。

[0103] 図7は、本発明の実施形態の実行サーバ負荷リスト113を示す説明図である。

[0104] 実行サーバ負荷リスト113は、実行サーバ1131及び負荷量1132を含む。実行サーバ1131は、実行サーバ101を一意に示す識別子、又は、名称である。負荷量1132は、ステップ604において算出された負荷量である。

[0105] 実行サーバ負荷リスト113には、負荷量1132が多い順に、エントリが格納される。図7に示す実行サーバ負荷リスト113は、負荷量1132を降順にソートしているが、移動データ決定部108が負荷量1132の大小を識別できれば、昇順でもよい。

[0106] ステップ604の後、移動データ決定部108は、初期設定として、パラメータKに1を格納し、パラメータJに1を格納する(605)。パラメータKは、負荷量が多い順番を示すパラメータである。パラメータJは、データ125の利用頻度が多い順番を示すパラメータである。

[0107] ステップ605の後、移動データ決定部108は、実行サーバ負荷リスト113を参照し、K番目に負荷量1132が多い実行サーバ1131(実行サーバ101に相当)を抽出する。そして、ジョブ・データ配置テーブル109及びデータ利用頻度テーブル110を参照し、抽出された実行サーバ101(実行サーバ1091に相当)に保持されるデータ125(口座番号1093に相当)のうち、J番目に利用頻度1102が多いデータ125(口座番号1101に相当)を抽出する(606)。

[0108] さらに、移動データ決定部108は、ステップ606において、抽出され

たK番目に負荷量1132が大きい実行サーバ101が保持するデータ125のうち、J番目に大きいデータ125を、実行サーバ101-Nに移動させた場合のジョブ・データ配置テーブル109を一時的に生成する。なお、ここで一時的に生成されたジョブ・データ配置テーブル109は、スケジューラサーバ100に備わるメモリの一時的な記憶領域に格納される。

[0109] そして、移動データ決定部108は、ステップ606において、アクセスパターン管理テーブル111と、一時的に生成されたジョブ・データ配置テーブル109とを参照し、ステップ601と同様な手順によって、実行サーバ101-1~101-N間の通信頻度を算出する。すなわち、移動データ決定部108は、一時的に生成されたジョブ・データ配置テーブル109を参照することによって、アクセスパターン管理テーブル111のエントリのうち、異なる実行サーバ101に格納されるデータ125の組を示す口座番号の組1111を含むエントリを抽出する。そして、抽出されたエントリの利用頻度1112をすべて合計することによって実行サーバ101-1~101-N間の通信頻度を算出する。

[0110] ステップ606の後、移動データ決定部108は、算出された実行サーバ101-1~101-N間の通信頻度が、サーバスケールアウトのコマンド131によって指定された通信頻度の閾値以上であるか否かを判定する（607）。すなわち移動データ決定部108は、ステップ607において、K番目に負荷の大きい実行サーバ101のデータ125のうち、J番目に利用頻度の大きいデータ125を、実行サーバ101-Nに移動した場合、通信頻度が閾値以上であるか否かを判定する。

[0111] 本実施形態の移動データ決定部108は、利用頻度の大きいデータ125を、実行サーバ101-Nに移動することによって、実行サーバ101の負荷を軽減する。しかし、利用頻度の大きいデータ125が、連携ジョブによって参照されるデータ125である場合、その利用頻度の大きいデータ125を実行サーバ101-Nに移動することによって、通信頻度が高くなる場合がある。このため、移動データ決定部108は、ステップ607において

、実行サーバ１０１間の通信頻度が、サーバスケールアウトのコマンド１３１によって指定された通信頻度の閾値以上であるか否かを判定する。

[0112] ステップ６０７において、算出された実行サーバ１０１－１～１０１－Ｎ間の通信頻度が、サーバスケールアウトのコマンド１３１によって指定された通信頻度の閾値以上である場合、実行サーバ１０１間の通信頻度が、管理者等が指定する通信頻度以上になるため、移動データ決定部１０８は、Ｊ番目に利用頻度の大きいデータ１２５を実行サーバ１０１－Ｎに移動せず、一時的に生成されたジョブ・データ配置テーブル１０９を削除する。そして、ステップ６１０に移行する。

[0113] ステップ６０７において、算出された実行サーバ１０１－１～１０１－Ｎ間の通信頻度が、サーバスケールアウトのコマンド１３１によって指定された通信頻度の閾値より小さい場合、実行サーバ１０１間の通信頻度は、管理者等が指定する通信頻度より小さいため、移動データ決定部１０８は、ステップ６０８に移行する。

[0114] ステップ６０７の後、移動データ決定部１０８は、ステップ６０８において、実行サーバ１０１－Ｎの負荷量と、Ｋ番目に負荷量大きい実行サーバ１０１の負荷量との差が、Ｊ番目に利用頻度の大きいデータ１２５を実行サーバ１０１－Ｎに移動した後において、移動する前より小さくなるか否かを判定する。

[0115] 具体的には、移動データ決定部１０８は、Ｊ番目に利用頻度１１０２の大きいデータ１２５を実行サーバ１０１－Ｎに移動する前における、実行サーバ１０１－Ｎの負荷量１１３２と、Ｋ番目に負荷量１１３２が大きい実行サーバ１０１の負荷量１１３２との差を算出し、さらに、Ｊ番目に利用頻度１１０２の大きいデータ１２５を実行サーバ１０１－Ｎに移動した後における、実行サーバ１０１－Ｎの負荷量１１３２と、Ｋ番目に負荷量１１３２が大きい実行サーバ１０１の負荷量１１３２との差を算出する。そして、算出された二つの差を比較して、Ｊ番目に利用頻度１１０２の大きいデータ１２５を実行サーバ１０１－Ｎに移動した後における差が小さいか否かを判定する。

(608)。

- [0116] なお、ステップ608における処理は、実行サーバ101-Nへデータ125を移動することによって、実行サーバ101-Nの負荷量1132が、データ125の移動元の実行サーバ101の負荷量1132よりも大きく、かつ、実行サーバ101-Nの負荷量1132とデータ125の移動元の実行サーバ101の負荷量1132との差が増加するか否かを判定できれば、いかなる算出方法を用いてもよい。
- [0117] 例えば、移動データ決定部108は、J番目に利用頻度の大きいデータ125の利用頻度1102を、実行サーバ101-Nの負荷量に加算することによって、J番目に利用頻度の大きいデータ125を実行サーバ101-Nに移動した場合の実行サーバ101-Nの、新たな負荷量を算出してもよい。また、K番目に負荷量1132が大きい実行サーバ1131の負荷量1132から、J番目に利用頻度の大きいデータ125の利用頻度1102を減算することによって、J番目に利用頻度の大きいデータ125を実行サーバ101-Nに移動した場合のK番目に負荷量1132が大きい実行サーバ101の新たな負荷量を算出してもよい。そして、移動データ決定部108は、算出された、実行サーバ101-Nの新たな負荷量と、K番目に1132が大きい実行サーバ101の新たな負荷量を減算し、差を算出してもよい。
- [0118] ステップ608において、実行サーバ101-Nの負荷量1132と、K番目に負荷量1132が大きい実行サーバ101の負荷量1132との差が、データ125を移動することによって大きくなる場合、実行サーバ101-Nの負荷量1132が過多となり、実行サーバ101間の負荷が平準化されないため、移動データ決定部108は、J番目に利用頻度の大きいデータ125を実行サーバ101-Nに移動せず、ステップ606において一時的に生成されたジョブ・データ配置テーブル109を削除する。そして、ステップ610に移行する。
- [0119] ステップ608において、実行サーバ101-Nの負荷量1132と、K番目に負荷量1132が大きい実行サーバ101の負荷量1132との差が

、データ 125 を移動することによって小さくなる場合、すべての実行サーバ 101 間の負荷が平準化されるため、移動データ決定部 108 は、K 番目に 1132 が大きい実行サーバ 101 が保持するデータ 125 のうち、J 番目に大きいデータ 125 を、実行サーバ 101-N へ移動することを決定する（609）。そして、ステップ 606 において一時的に生成されたジョブ・データ配置テーブル 109 によって、スケジューラサーバ 100 のジョブ・データ配置テーブル 109 を更新する。

[0120] ステップ 609 の後、移動データ決定部 108 は、ステップ 604 に戻る。そして、ステップ 604 において、ステップ 609 において決定されたデータ 125 を、実行サーバ 101-N へ移動した後の、実行サーバ負荷リスト 113 を生成する。

[0121] そして、移動データ決定部 108 は、ステップ 605 に移行し、パラメータ K 及びパラメータ J に、各々 1 を格納する。本実施形態のスケジューラサーバ 100 の移動データ決定部 108 は、ステップ 605 によって、実行サーバ 101-N に移動するデータ 125 を決定した後、最も負荷量 1132 が大きい実行サーバ 101 を抽出しなおすことができる。すなわち、本実施形態のスケジューラサーバ 100 の移動データ決定部 108 は、最も負荷量 1132 を減らすべき実行サーバ 101 を抽出することができる。

[0122] ステップ 607 又はステップ 608 の後、移動データ決定部 108 は、J 番目の次に大きいデータ 125 を選択するため、パラメータ J に 1 を加算する。

[0123] そして移動データ決定部 108 は、パラメータ J の値が、K 番目に負荷量 1132 が大きい実行サーバ 101 が保持するすべてのデータ 125 の個数より、大きいかな否かを判定する（611）。パラメータ J の値が、K 番目に負荷量 1132 が大きい実行サーバ 101 が保持するすべてのデータ 125 の個数以下である場合、K 番目に負荷量 1132 が大きい実行サーバ 101 が保持するすべてのデータ 125 を、実行サーバ 101-N に移動できるかな否かを判定していないため、移動データ決定部 108 は、ステップ 606 に

戻る。

- [0124] パラメータ J の値が、K 番目に負荷量 1 1 3 2 が大きい実行サーバ 1 0 1 が保持するすべてのデータ 1 2 5 の個数よりも大きい場合、K 番目の次に負荷量 1 1 3 2 が大きい実行サーバ 1 0 1 が保持するデータ 1 2 5 を実行サーバ 1 0 1 - N へ移動するか否かを決定するため、移動データ決定部 1 0 8 は、ステップ 6 1 2 に移行する。移動データ決定部 1 0 8 は、パラメータ K に 1 を加算し、パラメータ J に 1 を格納する（6 1 2）。
- [0125] ステップ 6 1 2 の後、移動データ決定部 1 0 8 は、パラメータ K と N とが等しいか否かを判定する（6 1 3）。パラメータ K と N とが異なる場合、すべての既存の実行サーバ 1 0 1 - 1 ~ 1 0 1 - (N - 1) が保持するデータ 1 2 5 を、実行サーバ 1 0 1 - N に移動するか否かを判定していないため、移動データ決定部 1 0 8 は、ステップ 6 0 6 に移行する。
- [0126] ステップ 6 1 2 において、パラメータ K と N とが等しい場合、既存の実行サーバ 1 0 1 - 1 ~ 1 0 1 - (N - 1) が保持するデータ 1 2 5 を、実行サーバ 1 0 1 - N に移動するか否かを判定したため、移動データ決定部 1 0 8 は、図 6 に示す処理を終了する。
- [0127] 移動データ決定部 1 0 8 は、図 6 に示す処理によって移動するデータ 1 2 5 を決定すると、データ 1 2 5 の移動を、データ 1 2 5 の移動元となる実行サーバ 1 0 1 のデータ送受信部 1 2 3 に指示する。また、移動データ決定部 1 0 8 は、図 6 の処理によって更新されたジョブ・データ配置テーブル 1 0 9 を実行サーバ 1 0 1 に送信する。
- [0128] これによって、実行サーバ 1 0 1 - N には、ジョブ・データ配置テーブル 1 0 9 及びデータ 1 2 5 が格納される。
- [0129] 以上によって、本実施形態の計算機システムは、実行サーバ 1 0 1 - N がシステムに追加された際に、実行サーバ 1 0 1 - N に移動するデータ 1 2 5 を決定することができる。また、移動するデータ 1 2 5 を決定する際に、実行サーバ 1 0 1 間の負荷量 1 1 3 2 の平準化を重視する判定と、実行サーバ 1 0 1 間の通信頻度の削減を重視する判定との両方の処理を行うことによっ

て、より管理者等にとって最適なジョブの実行環境を生成することができる。

[0130] なお、前述の説明は、Nが3以上である場合、すなわち、既存の実行サーバ101が2台以上である場合の計算機システムについて説明したが、Nが2である場合、すなわち、既存の実行サーバ101が1台のみである場合も、本実施形態の計算機システムは有効である。なお、Nが2である場合、図6のステップ604からの処理が実行される。

[0131] 図8は、本発明の実施形態の実行サーバ101-Nに移動するデータ125を、通信頻度を優先して決定する手順を示すフローチャートである。

[0132] 移動データ決定部108は、サーバスケールアウトのコマンド131を受信した後、アクセスパターン管理テーブル111のエントリのうち、異なる実行サーバ101に格納されているデータ125の組を実行サーバ101-Nに移動することを決定する(801)。

[0133] 図8に示す処理は、特にNが3などの少ない値である場合、実行サーバ101間の負荷の平準化と、通信頻度の低減とを図ることができる。このため、図6に示す処理と、図8に示す処理とは、実行サーバ101の数に従って実行されてもよい。

[0134] 本実施形態によれば、管理者等の要求に従って、実行サーバ101間の負荷の平準化、通信頻度の削減、又は、その両方を行うことができる。また、通信頻度の削減に際し、連携ジョブによって処理される頻度が多いデータ125の組、又は、連続して処理される頻度が多いデータ125の組を抽出し、実行サーバ101-Nに追加することを決定するため、より効率的に通信頻度を低減することができる。

[0135] 以上、本発明を添付の図面を参照して詳細に説明したが、本発明はこのような具体的構成に限定されるものではなく、添付した請求の範囲の趣旨内における様々な変更及び同等の構成を含むものである。

請求の範囲

[請求項1]

複数のジョブを実行する複数の実行サーバと、前記実行サーバに接続されるスケジューラサーバとを備える計算機システムであって、

前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、

前記スケジューラサーバは、

連続して実行される前記複数のジョブによって処理される複数のデータを各々含む複数の第1の組を生成し、

一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記複数の第1の組から抽出し、

前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする計算機システム。

[請求項2]

前記計算機システムは、前記第2の組に含まれるデータを移動する前の第1の状態をとることが可能であり、

前記スケジューラサーバは、

前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記連続して実行される複数のジョブが前記第2の組に含まれる複数のデータを処理する第2の頻度とを算出し、

前記第2の頻度と、所定の閾値とを比較する第1の比較を行い、

前記第1の比較によって、前記第1の状態における前記第2の頻度が大きいと判定された場合、前記抽出された第2の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第1の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバ

に移動するデータに決定することを特徴とする請求項 1 に記載の計算機システム。

[請求項3]

前記計算機システムは、前記算出された複数の負荷量に従って抽出されたデータを移動する前の第 2 の状態と、移動した後の第 3 の状態とをとることが可能であり、

前記スケジューラサーバは、

前記第 1 の比較によって、前記第 1 の状態における前記第 2 の頻度が小さいと判定された場合、前記第 3 の状態における前記第 2 の頻度を算出し、前記第 3 の状態における第 2 の頻度と、前記所定の閾値とを比較する第 2 の比較を行い、

前記第 2 の比較によって、前記第 3 の状態における第 2 の頻度が小さいと判定された場合、前記第 2 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 1 の差、及び、前記第 3 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 2 の差を算出し、前記第 1 の差と前記第 2 の差とを比較する第 3 の比較を行い、

前記第 3 の比較によって、前記第 2 の差が小さいと判定された場合、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする請求項 2 に記載の計算機システム。

[請求項4]

複数のジョブを実行する複数の実行サーバと、前記実行サーバに接続されるスケジューラサーバとを備える計算機システムであって、

前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、

前記スケジューラサーバは、

連続して処理される前記複数のデータを含む第 1 の組を取得し、

一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第 2 の組を、前記複数の第 1 の組から抽

出し、

前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする計算機システム。

[請求項5]

前記計算機システムは、前記第2の組に含まれるデータを移動する前の第1の状態をとることが可能であり、

前記スケジューラサーバは、

前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記第2の組が連続して処理される第2の頻度とを算出し、

前記第2の頻度と、所定の閾値とを比較する第1の比較を行い、

前記第1の比較によって、前記第1の状態における前記第2の頻度が大きいと判定された場合、前記抽出された第2の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第1の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定することを特徴とする請求項4に記載の計算機システム。

[請求項6]

前記計算機システムは、前記算出された複数の負荷量に従って抽出されたデータを移動する前の第2の状態と、移動した後の第3の状態とをとることが可能であり、

前記スケジューラサーバは、

前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第3の状態における前記第2の頻度を算出し、前記第3の状態における第2の頻度と、前記所定の閾値とを比較する第2の比較を行い、

前記第 2 の比較によって、前記第 3 の状態における第 2 の頻度が小さいと判定された場合、前記第 2 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 1 の差、及び、前記第 3 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 2 の差を算出し、前記第 1 の差と前記第 2 の差とを比較する、第 3 の比較を行い、

前記第 3 の比較によって、前記第 2 の差が小さいと判定された場合、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする請求項 5 に記載の計算機システム。

[請求項 7]

複数のジョブを実行する複数の実行サーバに接続されるスケジューラサーバにおける移動データ決定方法であって、

前記実行サーバは、前記ジョブによって処理される複数のデータをメモリに保持し、

前記方法は、

前記スケジューラサーバが、連続して実行される前記複数のジョブによって処理される複数のデータを各々含む複数の第 1 の組を生成し、前記生成された第 1 の組をメモリに格納し、

前記スケジューラサーバが、一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第 2 の組を、前記メモリに格納された複数の第 1 の組から抽出し、

前記スケジューラサーバが、前記抽出された第 2 の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とするデータ管理方法。

[請求項 8]

前記複数の実行サーバ及び前記スケジューラサーバは、計算機システムに含まれ、

前記計算機システムは、前記第 2 の組に含まれるデータを移動する前の第 1 の状態と、前記データを移動する前の第 2 の状態と、前記デ

ータを移動した後の第3の状態とをとることが可能であり、

前記方法は、

前記スケジューラサーバが、前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記連続して実行される複数のジョブが前記第2の組に含まれる複数のデータを処理する第2の頻度とを算出し、

前記スケジューラサーバが、前記第2の頻度と、所定の閾値とを比較する第1の比較を行い、

前記第1の比較によって、前記第1の状態における前記第2の頻度が大きいと判定された場合、前記スケジューラサーバが、前記抽出された第2の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第1の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定し、

前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第3の状態における前記第2の頻度を算出し、前記第3の状態における第2の頻度と、前記所定の閾値とを比較する第2の比較を行い、

前記第2の比較によって、前記第3の状態における第2の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第2の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第1の差、及び、前記第3の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第2の差を算出し、前記第1の差と前記第2の差とを比較する第3の比較を行い、

前記第 3 の比較によって、前記第 2 の差が小さいと判定された場合、前記スケジューラサーバが、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする請求項 7 に記載のデータ管理方法。

[請求項 9]

複数のジョブを実行する複数の実行サーバに接続されるスケジューラサーバにおける移動データ決定方法であって、

前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、

前記方法は、

前記スケジューラサーバが、連続して処理される前記複数のデータを含む複数の第 1 の組を、メモリに格納し、

前記スケジューラサーバが、一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第 2 の組を、前記メモリに格納された複数の第 1 の組から抽出し、

前記スケジューラサーバが、前記抽出された第 2 の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とするデータ管理方法。

[請求項 10]

前記複数の実行サーバ及び前記スケジューラサーバは、計算機システムに含まれ、

前記計算機システムは、前記第 2 の組に含まれるデータを移動する前の第 1 の状態と、前記データを移動する前の第 2 の状態と、前記データを移動した後の第 3 の状態とをとることが可能であり、

前記方法は、

前記スケジューラサーバが、前記第 1 の状態において、前記各データが前記ジョブによって処理される第 1 の頻度と、前記第 2 の組が連続して処理される第 2 の頻度とを算出し、

前記スケジューラサーバが、前記第 2 の頻度と、所定の閾値とを比較する第 1 の比較を行い、

前記第 1 の比較によって、前記第 1 の状態における前記第 2 の頻度が大きいと判定された場合、前記スケジューラサーバが、前記抽出された第 2 の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

前記第 1 の比較によって、前記第 1 の状態における前記第 2 の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第 1 の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定し、

前記スケジューラサーバが、前記第 1 の状態における前記第 2 の頻度が小さいと判定された場合、前記第 3 の状態における前記第 2 の頻度を算出し、前記第 3 の状態における第 2 の頻度と、前記所定の閾値とにおいて、第 2 の比較を行い、

前記第 2 の比較によって、前記第 3 の状態における第 2 の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第 2 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 1 の差、及び、前記第 3 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 2 の差を算出し、前記第 1 の差と前記第 2 の差とを比較する第 3 の比較を行い、

前記第 3 の比較によって、前記第 2 の差が小さいと判定された場合、前記スケジューラサーバが、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする請求項 9 に記載のデータ管理方法。

補正された請求の範囲
[2011年4月25日(25.04.2011)国際事務局受理]

[請求項1] (補正後)

複数のジョブを実行する複数の実行サーバと、前記実行サーバに接続されるスケジューラサーバとを備える計算機システムであって、

- 5 前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、
前記スケジューラサーバは、

連続して実行される前記複数のジョブによって処理される複数のデータを各々含む複数の第1の組を生成し、

- 10 一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記複数の第1の組から抽出し、

前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定し、

前記計算機システムは、前記第2の組に含まれるデータを移動する前の第1の状態をとることが可能であり、

- 15 前記スケジューラサーバは、

前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記連続して実行される複数のジョブが前記第2の組に含まれる複数のデータを処理する第2の頻度とを算出し、

前記第2の頻度と、所定の閾値とを比較する第1の比較を行い、

- 20 前記第1の比較によって、前記第1の状態における前記第2の頻度が大きいと判定された場合、前記抽出された第2の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第1の頻度を合計することによって、前記各実行サーバの
25 負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定することを特徴とする計算機システム。

[請求項2] (削除)

[請求項3]

- 30 前記計算機システムは、前記算出された複数の負荷量に従って抽出されたデー

タを移動する前の第2の状態と、移動した後の第3の状態とをとることが可能であり、

前記スケジューラサーバは、

- 5 前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第3の状態における前記第2の頻度を算出し、前記第3の状態における第2の頻度と、前記所定の閾値とを比較する第2の比較を行い、

- 前記第2の比較によって、前記第3の状態における第2の頻度が小さいと判定された場合、前記第2の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第1の差、及び、前記第3の状態における、
10 前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第2の差を算出し、前記第1の差と前記第2の差とを比較する第3の比較を行い、

前記第3の比較によって、前記第2の差が小さいと判定された場合、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする請求項2に記載の計算機システム。

- 15 [請求項4] (補正後)

複数のジョブを実行する複数の実行サーバと、前記実行サーバに接続されるスケジューラサーバとを備える計算機システムであって、

前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、

前記スケジューラサーバは、

- 20 連続して処理される前記複数のデータを含む第1の組を取得し、

一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記複数の第1の組から抽出し、

前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定し、

- 25 前記計算機システムは、前記第2の組に含まれるデータを移動する前の第1の状態をとることが可能であり、

前記スケジューラサーバは、

前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記第2の組が連続して処理される第2の頻度とを算出し、

- 30 前記第2の頻度と、所定の閾値とを比較する第1の比較を行い、

前記第1の比較によって、前記第1の状態における前記第2の頻度が大きいと判定された場合、前記抽出された第2の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

- 5 前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第1の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定することを特徴とすることを特徴とする計算機システム。

[請求項5] (削除)

- 10 [請求項6]

前記計算機システムは、前記算出された複数の負荷量に従って抽出されたデータを移動する前の第2の状態と、移動した後の第3の状態とをとることが可能であり、

前記スケジューラサーバは、

- 15 前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記第3の状態における前記第2の頻度を算出し、前記第3の状態における第2の頻度と、前記所定の閾値とを比較する第2の比較を行い、

- 前記第2の比較によって、前記第3の状態における第2の頻度が小さいと判定された場合、前記第2の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第1の差、及び、前記第3の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第2の差を算出し、前記第1の差と前記第2の差とを比較する、第3の比較を行い、

- 20 前記第3の比較によって、前記第2の差が小さいと判定された場合、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とする請求項5に記載の計算機システム。

[請求項7] (補正後)

複数のジョブを実行する複数の実行サーバに接続されるスケジューラサーバにおける移動データ決定方法であって、

- 30 前記実行サーバは、前記ジョブによって処理される複数のデータをメモリに保持し、

前記方法は、

前記スケジューラサーバが、連続して実行される前記複数のジョブによって処理される複数のデータを各々含む複数の第1の組を生成し、前記生成された第1の組をメモリに格納し、

- 5 前記スケジューラサーバが、一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記メモリに格納された複数の第1の組から抽出し、

前記スケジューラサーバが、前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定し、

- 10 前記スケジューラサーバは、前記第2の組に含まれるデータを移動する前の第1の状態と、前記データを移動する前の第2の状態と、前記データを移動した後の第3の状態とをとることが可能であり、

前記方法は、

- 15 前記スケジューラサーバが、前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記連続して実行される複数のジョブが前記第2の組に含まれる複数のデータを処理する第2の頻度とを算出し、

前記スケジューラサーバが、前記第2の頻度と、所定の閾値とを比較する第1の比較を行い、

- 20 前記第1の比較によって、前記第1の状態における前記第2の頻度が大きいと判定された場合、前記スケジューラサーバが、前記抽出された第2の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

- 前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第1の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定し、
- 25

- 前記第1の比較によって、前記第1の状態における前記第2の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第3の状態における前記第2の頻度を算出し、前記第3の状態における第2の頻度と、前記所定の閾値とを比較する第2の比較を行い、
- 30

前記第2の比較によって、前記第3の状態における第2の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第2の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第1の差、及び、前記第3の状態における、前記各実行サーバの負荷量と前記新たに追加された
5 実行サーバの負荷量との第2の差を算出し、前記第1の差と前記第2の差とを比較する第3の比較を行い、

前記第3の比較によって、前記第2の差が小さいと判定された場合、前記スケジューラサーバが、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とするデータ管理方法。

10 [請求項8] (削除)

[請求項9] (補正後)

複数のジョブを実行する複数の実行サーバに接続されるスケジューラサーバにおける移動データ決定方法であって、

前記実行サーバは、前記ジョブによって処理される複数のデータを保持し、

15 前記方法は、

前記スケジューラサーバが、連続して処理される前記複数のデータを含む複数の第1の組を、メモリに格納し、

前記スケジューラサーバが、一つの前記実行サーバに保持されるデータと、他の前記実行サーバに保持されるデータとを含む第2の組を、前記メモリに格納さ

20 れた複数の第1の組から抽出し、

前記スケジューラサーバが、前記抽出された第2の組に含まれるデータを、新たに追加された前記実行サーバに移動するデータに決定し、

前記スケジューラサーバは、前記第2の組に含まれるデータを移動する前の第1の状態と、前記データを移動する前の第2の状態と、前記データを移動した後

25 の第3の状態とをとることが可能であり、

前記方法は、

前記スケジューラサーバが、前記第1の状態において、前記各データが前記ジョブによって処理される第1の頻度と、前記第2の組が連続して処理される第2の頻度とを算出し、

30 前記スケジューラサーバが、前記第2の頻度と、所定の閾値とを比較する第1

の比較を行い、

前記第 1 の比較によって、前記第 1 の状態における前記第 2 の頻度が大きいと判定された場合、前記スケジューラサーバが、前記抽出された第 2 の組に含まれるデータを、前記新たに追加された実行サーバに移動するデータに決定し、

- 5 前記第 1 の比較によって、前記第 1 の状態における前記第 2 の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第 1 の頻度を合計することによって、前記各実行サーバの負荷量を算出し、前記算出された複数の負荷量に従って抽出された前記データを、前記新たに追加された実行サーバに移動するデータに決定し、

- 10 前記スケジューラサーバが、前記第 1 の状態における前記第 2 の頻度が小さいと判定された場合、前記第 3 の状態における前記第 2 の頻度を算出し、前記第 3 の状態における第 2 の頻度と、前記所定の閾値とにおいて、第 2 の比較を行い、

前記第 2 の比較によって、前記第 3 の状態における第 2 の頻度が小さいと判定された場合、前記スケジューラサーバが、前記第 2 の状態における、前記各実行

- 15 サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 1 の差、及び、前記第 3 の状態における、前記各実行サーバの負荷量と前記新たに追加された実行サーバの負荷量との第 2 の差を算出し、前記第 1 の差と前記第 2 の差とを比較する第 3 の比較を行い、

前記第 3 の比較によって、前記第 2 の差が小さいと判定された場合、前記スケ
20 ジューラサーバが、前記負荷量に従って抽出されたデータを、新たに追加された前記実行サーバに移動するデータに決定することを特徴とするデータ管理方法。

[請求項 10] (削除)

[図2]

口座番号	口座残高
1	100,000,000
2	1,000,000
⋮	
1000	100,000

[図3]

実行サーバ	ジョブ	口座番号
実行サーバ1	出金ジョブ、入金ジョブ	1～1000
実行サーバ2	出金ジョブ、入金ジョブ	1001～2000
⋮		
実行サーバN-1	出金ジョブ、入金ジョブ	$1000 \times (N-1) + 1 \sim 1000 \times N$

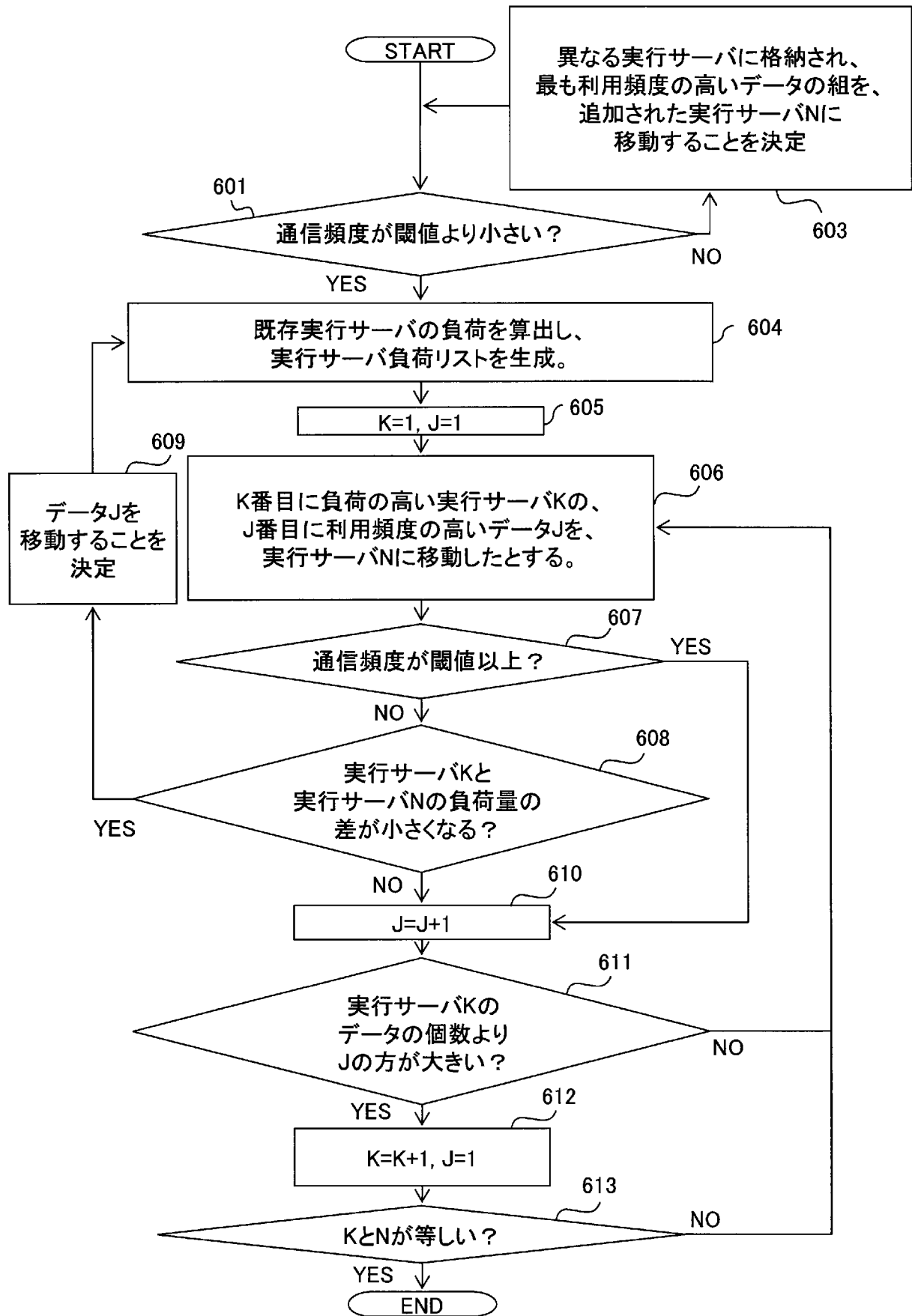
[図4]

口座番号	利用頻度(回/分)
1	10
2	0.5
⋮	
$1000 \times N$	20

[図5]

口座番号の組	利用頻度(回/分)
1,5000	3
3,4010	4.1
⋮	
2034,9012	10.4

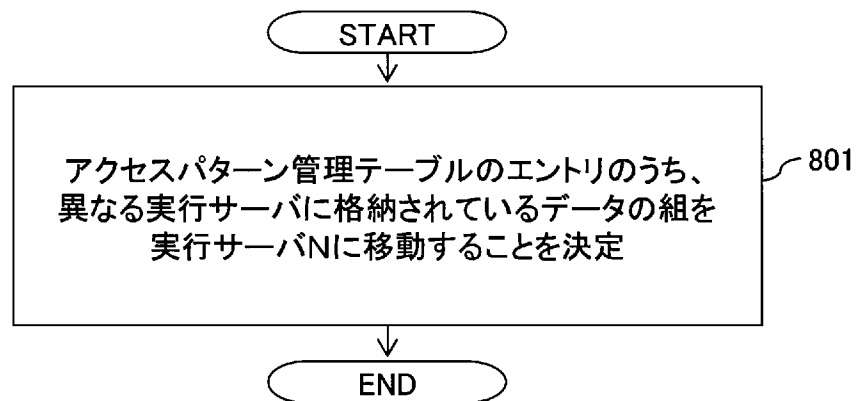
[図6]



[図7]

実行サーバ	負荷量(回/分)
実行サーバ5	100
実行サーバ7	84
⋮	
実行サーバ3	3

[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/051347

A. CLASSIFICATION OF SUBJECT MATTER

G06F12/00 (2006.01) i, G06F9/50 (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)

G06F12/00, G06F9/50

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)

JSTPlus (JDreamII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	JP 2002-007364 A (Fujitsu Ltd.), 11 January 2002 (11.01.2002), paragraphs [0017] to [0105] & US 2002/0002578 A1	1, 4, 7, 9 2-3, 5-6, 8, 10
A	JP 2007-102332 A (Toshiba Corp.), 19 April 2007 (19.04.2007), paragraphs [0001] to [0101] (Family: none)	1-10
A	Hirokazu UCHI, "Heiretsu Bunsan Program no Tameno Job Migration Scheduling no Jisso to Hyoka", IPSJ SIG Notes, 02 June 2005 (02.06. 2005), vol.2005, no.57, pages 27 to 32	1-10



Further documents are listed in the continuation of Box C.



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

21 February, 2011 (21.02.11)

Date of mailing of the international search report

01 March, 2011 (01.03.11)

Name and mailing address of the ISA/

Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/051347

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Tsuguhiro KONDO, "A Parallel Computing System on Data Grid Environment using Process Migration", IEICE Technical Report, 20 January 2005 (20.01.2005), vol.104, no.593, pages 31 to 36	1-10

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F12/00(2006.01)i, G06F9/50(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（I P C））

Int.Cl. G06F12/00, G06F9/50

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

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 1 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 1 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 1 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）

JSTPlus(JDreamII)

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 A	JP 2002-007364 A（富士通株式会社）2002.01.11, 段落 0017-0105 & US 2002/0002578 A1	1, 4, 7, 9 2-3, 5-6, 8, 10
A	JP 2007-102332 A（株式会社東芝）2007.04.19, 段落 0001-0101（ファミリーなし）	1-10
A	内 裕和, 並列分散プログラムのためのジョブマイグレーションスケジューリングの実装と評価, 情報処理学会研究報告, 2005.06.02, 第2005巻, 第57号, p.27-32	1-10

☒ C欄の続きにも文献が列挙されている。☐ 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

「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
「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
「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）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献
「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
「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
「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
「&」同一パテントファミリー文献

国際調査を完了した日

2 1 . 0 2 . 2 0 1 1

国際調査報告の発送日

0 1 . 0 3 . 2 0 1 1

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号 1 0 0 - 8 9 1 5

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）

田川 泰宏

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 9

5 M

4 2 3 6

C (続き) . 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
A	近藤 紹弘, プロセスマイグレーションを用いたデータグリッド環境における並列計算システム, 電子情報通信学会技術研究報告, 2005. 01. 20, 第 104 巻, 第 593 号, p. 31-36	1-10