

(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
2016年7月28日(28.07.2016)



(10) 国際公開番号

WO 2016/117032 A1

- (51) 国際特許分類:
G06F 12/00 (2006.01)
- (21) 国際出願番号: PCT/JP2015/051414
- (22) 国際出願日: 2015年1月20日(20.01.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 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: 磯田 有哉(ISODA, Yuya); 〒1008280 東京
都千代田区丸の内一丁目6番6号 株式会社日
立製作所内 Tokyo (JP). 友田 敦(TOMODA, At-
sushi); 〒1008280 東京都千代田区丸の内一丁目6
番6号 株式会社日立製作所内 Tokyo (JP). 花井
知広(HANAI, Tomohiro); 〒1008280 東京都千代田
区丸の内一丁目6番6号 株式会社日立製作所内
Tokyo (JP). 牛嶋 一智(USHIJIMA, Kazutomo);
〒1008280 東京都千代田区丸の内一丁目6番6
号 株式会社日立製作所内 Tokyo (JP). 田中 剛
(TANAKA, Tsuyoshi); 〒1008280 東京都千代田区丸
の内一丁目6番6号 株式会社日立製作所内
Tokyo (JP).
- (74) 代理人: 特許業務法人ウィルフォート国際特許
事務所(WILLFORT INTERNATIONAL PATENT
FIRM); 〒1030016 東京都中央区日本橋小網町1
9-7 日本橋TCビル 1階 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, BN, 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,
IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR,
LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SA, 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, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユー
ラシア (AM, AZ, BY, KG, KZ, 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, KM,
ML, MR, NE, SN, TD, TG).

[続葉有]

(54) Title: DATABASE SYSTEM, COMPUTER SYSTEM, AND DATABASE MANAGEMENT METHOD

(54) 発明の名称: データベースシステム、計算機システム、及び、データベース管理方法

FIG. 1

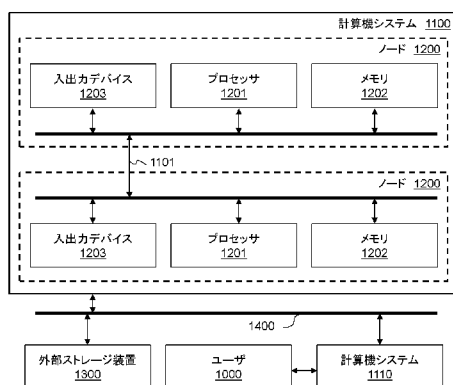


FIG. 1:
1000 User
1100, 1110 Computer system
1200 Node
1201 Processor
1202 Memory
1203 I/O device
1300 External storage device

(57) Abstract: Provided is a computer system (e.g., a database system), which, when executing a transaction which updates a record among a plurality of records in a storage unit, preserves in the storage unit pre-update records with which timestamps are correlated, and stores in the storage unit a post-update record with a correlated post-update timestamp. On the basis of a plurality of records at each of a plurality of timestamps, the computer system specifies a record which is not queried from other transactions being executed, among the records which correspond to the timestamps between a first timestamp which is a timestamp which is associated with a first transaction being executed and a second timestamp which is a timestamp which is associated with a second transaction being executed, and deletes the specified record from the storage unit.

(57) 要約: 計算機システム(例えばデータベースシステム)が、トランザクションを実行することにより記憶部内の複数のレコードのうちのいずれかのレコードを更新する場合、タイムスタンプが関連付けられている更新前レコードを記憶部に維持し、更新後のタイムスタンプを関連付けた更新後レコードを記憶部に格納する。計算機システムは、実行中の第1のトランザクションに対応付けられたタイムスタンプである第1のタイムスタンプと、実行中の第2のトランザクションに対応付けられたタイムスタンプである第2のタイムスタンプとの間のタイムスタンプに対応したレコードのうち、実行中の他のトランザクションから参照されないレコードを、複数のタイムスタンプの各々での複数のレコードを基に特定し、特定したレコードを記憶部から削除する。

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

データベースシステム、計算機システム、及び、データベース管理方法
技術分野

[0001] 本発明は、概して、トランザクション処理に関し、例えば、データベースの一貫性制御等のトランザクション処理におけるデータを管理する技術に関する。

背景技術

[0002] 一般に、計算機システムにおいてトランザクションの処理が行われている。トランザクションを処理するシステムとして、例えば、DBMS（データベース管理システム）が知られている。DBMSは、入出力するデータであるレコードの一貫性を制御するいわゆるトランザクション処理（Tx処理）を実行する。これにより、DBMSにおいては、ユーザは常に意図したレコードを入出力することが可能となる。Tx処理は、複数の参照、更新命令をアトミックに実現するとき用いられる。Tx処理では、アトミックに実行したい命令区間の始めと終わり（Tx処理区間）を宣言することにより、指定した区間の複数の命令をアトミックに処理する。

[0003] Tx処理を実現するためには、ACID特性を満たす必要がある。ACID特性とは、原子性（Atomicity）、一貫性（Consistency）、独立性（Isolation）及び永続性（Durability）を示す。原子性とは、Tx処理区間の複数の命令を全て実行するか、全て実行させないことである。一貫性とは、データベースに与えられたルール以外の状態に遷移させないことである。独立性とは、Tx処理の結果は、必ずTx処理を逐次的に実行したときと同じになることをいう。永続性とは、Tx処理がコミットしていれば、DBMSに故障が発生していても必ずTx処理の結果を復元できることをいう。このように、DBMSでは、ACID特性を満たすことによって、Tx処理の一貫性を維持している。

[0004] しかし、高度な一貫性制御により、DBMSの処理能力が低下する課題がある。例えば、Tx処理の一貫性制御を維持しつつ、不要なレコードを削除するガベージコレクション（GC）が処理能力低下の原因となる。

[0005] この課題に対し、特許文献1又は2に開示の技術がある。特許文献1には、Tx処理やレコードがタイムスタンプで管理されるシステムにおいて、DBMSの永続化処理であるスナップショットより以前のタイムスタンプを持つレコードを削除する技術が開示されている。特許文献2には、動作中のTx処理で最も古いタイムスタンプより以前のタイムスタンプを持つレコードを削除する技術が開示されている。

先行技術文献

特許文献

[0006] 特許文献1：米国特許公開2014／149353号明細書

特許文献2：米国特許6125371号

発明の概要

発明が解決しようとする課題

[0007] DBMSはデータベースの少なくとも一部であるレコード群（複数のレコード）をメモリで管理することが可能である。これにより、Tx処理におけるストレージアクセス回数を軽減でき、Tx処理を高速に実行することが可能である。

[0008] 特定の一貫性制御において、Tx処理期間中にTx処理が同一レコードを複数回参照するのであれば、必ず同じ結果が得られることが保証されなければならない。このとき、Tx処理が参照しているレコードがGCによってメモリから削除された場合、このTx処理が繰り返し同じレコードを参照するには、このレコードが永続化されているストレージ装置からそのレコードをメモリへと読み出す必要がある。

[0009] また、近年、マルチバージョンングを用いてレコードを管理ことが行われている。マルチバージョンングでは、同一のレコードに関する複数のバージ

ョン（同じレコードに関する複数の時点の状態）を管理する。このようなマルチバージョンングを用いてレコードを管理する場合には、同一のレコードについての複数のバージョンのレコードがメモリ上に配置される。

[0010] 例えば、大規模なレコード参照があると、その処理に長い時間（例えば数時間）かかることがある。また、たくさんのレコード更新があると、参照処理において参照済みのレコードを削除することはできても、最新の更新から参照処理間のレコードを削除することはできない。

[0011] メモリ容量には制限があり、複数のバージョンのレコードがメモリに配置されると、メモリ容量の不足が生じやすくなり、結果として、レコードの追加、更新、及び削除といった操作ができなくなる懸念がある。一方、無闇にGCを実施すると、レコードを記憶するストレージ装置へのアクセス回数が増加し、T×処理時間が長くなる懸念がある。

[0012] 上述のような課題は、DBMSを実行する計算機システムに限らず、レコード（データ単位）をトランザクション処理において更新する他の計算機システムにも有り得る。

課題を解決するための手段

[0013] 計算機システム（例えばデータベースシステム）が、トランザクションを実行することにより記憶部内の複数のレコードのうちのいずれかのレコードを更新する場合、タイムスタンプが関連付けられている更新前レコードを記憶部に維持し、更新後のタイムスタンプを関連付けた更新後レコードを記憶部に格納する。計算機システムは、実行中の第1のトランザクションに対応付けられたタイムスタンプである第1のタイムスタンプと、実行中の第2のトランザクションに対応付けられたタイムスタンプである第2のタイムスタンプとの間のタイムスタンプに対応したレコードのうち、実行中の他のトランザクションから参照されないレコードを、複数のタイムスタンプの各々の複数のレコードを基に特定し、特定したレコードを記憶部から削除する。

発明の効果

[0014] 実行中のトランザクション処理に関与しないレコードを特定してそのよう

なレコードを記憶部から削除できる。

図面の簡単な説明

- [0015] [図1]実施形態に係るシステム全体の概要のブロック図である。
- [図2]DBMSの構成図である。
- [図3]第1の時点でのデータ遷移管理情報を示す。
- [図4]第1の時点でのTxIDリストを示す。
- [図5A]Record Aの遷移を示す。
- [図5B]Record Bの遷移を示す。
- [図5C]Record Cの遷移を示す。
- [図6]第2の時点でのTxIDリストを示す。
- [図7]第2の時点でのGC（ガベージコレクション）リストである。
- [図8]GC設定情報を示す。
- [図9]GC処理のフローチャートである。
- [図10]マルチGC処理のフローチャートである。
- [図11]GCリスト回収処理のフローチャートである。
- [図12]Tx（トランザクション）処理のフローチャートである。

発明を実施するための形態

- [0016] 以下、図面を参照して、本発明の一実施形態が適用されたDBMSを実行する計算機システムを説明する。
- [0017] 以下の説明では、「×××リスト」の表現にて情報を説明することがあるが、情報は、どのようなデータ構造で表現されていてもよい。すなわち、情報がデータ構造に依存しないことを示すために、「×××リスト」を「×××情報」と呼ぶことができる。
- [0018] また、以下の説明では、「PDEV」は、物理的な記憶デバイスを示し、典型的には、不揮発性の記憶デバイス（例えば補助記憶デバイス）でよい。PDEVは、例えば、HDD（Hard Disk Drive）又はSSD（Solid State Drive）でよい。
- [0019] また、以下の説明では、機能部（例えば、クエリ受付部、クエリプラン生

成部、クエリ実行部)を主語として処理を説明する場合があるが、機能部は、プログラムがプロセッサ(例えばCPU(Central Processing Unit))によって実行されることで、定められた処理を、適宜に記憶部(例えばメモリ)及び／又はインターフェースデバイス(例えば通信ポート)等を用いながら行うため、処理の主語がプロセッサとされてもよい。機能部を主語として説明された処理は、プロセッサあるいはそのプロセッサを有する装置又はシステムが行う処理としてもよい。また、プロセッサは、処理の一部または全部を行うハードウェア回路を含んでもよい。複数の機能部のうちの少なくとも一部がハードウェア回路で実現されてもよい。プログラムは、プログラムソースから計算機のような装置にインストールされてもよい。プログラムソースは、例えば、プログラム配布サーバまたは計算機が読み取り可能な記憶メディアであってもよい。プログラムソースがプログラム配布サーバの場合、プログラム配布サーバはプロセッサ(例えばCPU)と記憶部を含み、記憶部はさらに配布プログラムと配布対象であるプログラムとを記憶してよい。そして、プログラム配布サーバのプロセッサが配布プログラムを実行することで、プログラム配布サーバのプロセッサは配布対象のプログラムを他の計算機に配布してよい。また、以下の説明において、2以上の機能部が1つの機能部として実現されてもよいし、1つの機能部が2以上の機能部として実現されてもよい。

[0020] また、以下の説明では、「トランザクション」を「Tx」と表記し、ガベージコレクションを「GC」と表記する。

[0021] また、以下の説明では、同一の要素を区別しないで説明する場合、参照符号の共通符号又は名称のみを使用するが(例えば、TxID、データ遷移管理情報2000)、同一の要素を区別して説明する場合、要素の参照符号全体、ID又は値を使用する(例えば、TxID「200」、レコード管理情報2200A)ことがある。

[0022] 図1は、実施形態に係るシステム全体の概要のブロック図である。

[0023] ユーザ1000は、計算機システム(例えばパーソナルコンピュータ)1

110からLAN等のネットワーク1400を通じて、計算機システム（例えば複数のノード1200により実現される大規模計算機システム）1100にアクセスすることによって、計算機システム1100に命令を送ることができる。ユーザ1000は、計算機システム1100を直接操作することも可能である。

[0024] 計算機システム1100は、1以上のノード1200を有する。ノード1200は、計算機の一例である。複数のノード1200間は、システムバス1101によって接続されていてもよい。

[0025] ノード1200は、メモリ1202、入出力デバイス1203、及びそれらに接続されたプロセッサ1201を有する。メモリ1202は、主メモリ（例えば、DRAM（Dynamic Random Access Memory）のような揮発メモリ）及び補助メモリ（例えばフラッシュメモリのような不揮発メモリ）のうちの少なくとも主メモリを含む。入出力デバイス1203は、通信インターフェイスデバイスでもよいし、ヒューマンマシンインターフェイスデバイス（例えば、キーボード及びポインティングデバイスのような入力デバイスと、表示デバイスのような出力デバイス）でもよい。また、計算機システム1100の外部に、ネットワーク1400経由で通信可能な外部ストレージ装置1300があってもよい。

[0026] 図2は、DBMSの構成図である。

[0027] DBMS（Database Management System）1301は、1以上のノード1200（具体的には、1以上のプロセッサ1201）で実行されることにより構成される。DBMS1301は、クエリソース（例えば計算機システム1110）からクエリを受け付けるクエリ受付部1311と、そのクエリの実行プランであるクエリプランを生成するクエリプラン生成部1312と、そのクエリプランに従いクエリを実行するクエリ実行部1313とを有する。Tx処理で実行されるTxは、クエリプランに従う1つのクエリのTxに対応する。つまり、本実施形態では、1つのクエリに対して1つのTxIDがクエリ実行部1313により割り振られる。クエリ実行部1313は、

GCに関する処理を実行する。GCに関する処理については、後述する。これらの機能部1311、1312、及び1313は、DBMS1301が1以上のプロセッサ1201に実行されることにより実現されてよい。クエリは、例えば、SQL (Structured Query Language) で記述されている。

[0028] DBMS1301によって管理（参照又は更新等）されるデータベースは、1以上のメモリ1202に格納されており、いわゆるインメモリデータベースが実現されている。ここで、外部ストレージ装置1300からよりも、メモリ1202からの方が、高速にデータを取得することができる。データベースの全部又は一部が、外部ストレージ装置1300に格納されてもよい。

[0029] 図3は、第1の時点でのデータ遷移管理情報を示す。以下、レコード群（複数のレコード）として、主に、Record A～Record Cを例に取る。

[0030] データ遷移管理情報2000は、DBMS1301がマルチバージョンングを用いてテーブルの複数のレコードを管理するための情報である。ここで、マルチバージョンングとは、レコードの更新履歴を管理する方法の一つであり、レコードが上書きされる（更新後のレコードが残るが更新前のレコードが残らない）シングルバージョンングとは異なり、1つのレコードについて更新前のレコードと更新後のレコードの両方を管理する方法である。データ遷移管理情報2000は、メモリ1202又は外部ストレージ装置1300の少なくとも1つに保持される。

[0031] データ遷移管理情報2000は、タイムスタンプ毎にカラムを有する。各カラムが記憶する情報は、例えば、タイムスタンプ301及びレコード値302A～302Cである。タイムスタンプ301は、タイムスタンプの値である。レコード値302は、対応するレコードに格納されている値である。

[0032] カラムは、レコード群の状態を遷移させるTxが発生する都度（タイムスタンプが追加される都度）に追加される。タイムスタンプの値は、時刻やカウンタでもよく（例えば更新のTxのコミットの際に1インクリメント

されるカウンタでよく)、要は、レコード更新の順序を特定することのできる情報であればよい。本実施形態では、タイムスタンプの値は、新しいほど(将来であるほど)大きく古いほど(過去であるほど)小さい。参照のみの場合、タイムスタンプは更新されないでよい。タイムスタンプは、レコード群のバージョンと言い換えられてもよい。図3の例によれば、タイムスタンプ「0」(バージョン「0」)に対応するレコード群の状態は、Record A「100」、Record B「200」、Record C「300」である。タイムスタンプ「1」では、Record Aが「100」から「101」に更新され、結果、タイムスタンプ「1」に対応するレコード群の状態は、Record A「101」、Record B「200」、及びRecord C「300」である。

[0033] 図4は、第1の時点でのTxIDリストを示す。つまり、図4と図3は互いに対応している。

[0034] TxIDリスト2100は、実行が開始されたTxごとのタイムスタンプを管理するためのリストであり、メモリ1202又は外部ストレージ装置1300の少なくとも1つに保持される。

[0035] TxIDリスト2100は、実行が開始されたTx毎にカラムを有する。各カラムが記憶する情報は、例えば、TxID401、開始時刻402、参照タイムスタンプ403、更新タイムスタンプ404、及び、状態405である。TxIDリスト2100(及び後述のTxIDリスト2500)では、参照タイムスタンプ403の小さい順にカラムが並ぶ。

[0036] TxID401は、TxのIDである。開始時刻402は、Txの開始時刻を示す。

[0037] 参照タイムスタンプ403は、Txの開始時のタイムスタンプである。参照タイムスタンプ403は、レコードを参照可能か否かの判断のために用いられる。例えば、TxID「1」に対応する参照タイムスタンプ403は「1」である。データ遷移管理情報2000を参照することにより、タイムスタンプ「1」に対応したレコード群の状態は、Record A「101」、

Record B「201」、Record C「300」であることを把握できる。

[0038] 更新タイムスタンプ404は、Txでレコードを更新した際の更新後のタイムスタンプである。更新タイムスタンプ404の値は、レコードを更新したTxのみにより取得される。例えば、Tx「0」（TxID「0」のTx）は、Record Aの更新を行った場合には、データ遷移管理情報2000に新たなカラムを追加し、Record Aの値を「101」に更新し、タイムスタンプを「1」に更新し、TxID「0」に対応したカラム（TxIDリスト2100のカラム）に、更新タイムスタンプ「1」を格納する。このようにして、Record A「101」と、タイムスタンプ「1」とが関連付けられる。

[0039] 状態405は、Txの種類（処理内容）を格納する。具体的には、例えば、参照を行うTxについては「R」（Read）が格納され、更新を行うTxについては「W」（Write）が格納され、処理内容が確定したときに「C」（Commit）が格納される。従って、例えば、SELECT、UPDATE、COMMITの順に処理が進む場合、SELECTのときに状態405が「R」となり、UPDATEのときに状態405が「RW」となり、COMMITのときに状態405が「RWC」となる。なお、参照のみに限定したTxについては、「OR」（Only Read）のような特定の値が格納されてよい。また、処理内容が事前にクエリソース（例えば計算機システム1110）から宣言された場合、SELECT、UPDATE及びCOMMITのいずれのときも状態405は「RWC」となる。

[0040] 図5A～図5Cは、Record A～Cの遷移を示す。図6は、第2の時点でのTxIDリストを示す。図5A～図5Cと図6は互いに対応している。第1の時点でのTxIDリストとの混同を避けるために、第2の時点でのTxIDリストには、異なる参照符号（2500）を付し、TxIDリスト2500の各カラムが記憶する情報にも、異なる参照符号（601）～（605）を付す。

- [0041] Record Aを例に取り、レコードの構成を説明する。図5 Aの各カラムが、そのカラムに対応したバージョンのRecord Aである。Record Aは、タイムスタンプ501 A、Next 502 A、Prev 503 A、GC 504 A、及び、Data 505 Aを記憶する。
- [0042] タイムスタンプ501 Aは、Record Aのバージョンを示すタイムスタンプの値である。タイムスタンプ501 Aは、更新タイムスタンプ604と関連し、レコードが更新された際に取得されるレコードに対するタイムスタンプである。
- [0043] Next 502 Aは、更新後Record Aのアドレスである。本実施形態では、インメモリデータベースが採用されているので、更新後Record Aは、メモリ領域のアドレス（又はそれに関連付けられた値）でよい。更新後Record Aが無い場合、Next 502 Aは、不定値でよい。
- [0044] Prev 503 Aは、更新前Record Aのアドレスである。更新前Record Aのアドレスも、メモリ領域のアドレス（又はそれに関連付けられた値）でよい。更新前Record Aのアドレスが無い場合、Prev 503 Aには、不定値でよい。
- [0045] GC 504 Aは、GCにて用いられる情報である。GC 504 Aは、このバージョン（タイムスタンプ）のRecord A以外のレコードを特定する情報（例えばアドレス（又はID）である。
- [0046] Data 505 Aは、Record Aに格納された値である。その値が、Txにより、参照又は更新される。
- [0047] タイムスタンプ「124」を含んだRecord Aを説明する。Next 502 Aは、タイムスタンプ「155」を含んだRecord Aのアドレスである。Prev 503 Aは、タイムスタンプ「80」を含んだRecord Aのアドレスである。GC 504 Aは、後述するGC処理が開始されていない状態であるので、不定値である。Data 505 Aは、タイムスタンプ「124」のTxにより格納された値である。
- [0048] 図7は、GCリストである。なお、以下、タイムスタンプ「n」のレコー

ドを、「レコード# n」と表記する。値 y を格納したレコードは、「レコード「y」」と表記されるが、タイムスタンプ「n」のレコードは、異なる表記「レコード# n」となる。

[0049] GCリスト2600は、メモリ1202及び外部ストレージ装置1300の少なくとも1つに保持される。GCリスト2600は、GC処理で参照するタイムスタンプ毎のカラムを有する。各カラムが記憶する情報は、例えば、タイムスタンプ701及びレコード群702である。

[0050] タイムスタンプ701は、Txで参照可能なレコード群（複数のレコード）を示すタイムスタンプの値である。

[0051] レコード群702は、タイムスタンプ701に対応する参照可能なレコード群を特定する情報（例えばレコード毎のアドレス（又はID））である。タイムスタンプ「102」の時点で参照可能なレコード群は、例えば、Record A # 80及びRecord C # 86であるとする、タイムスタンプ「102」に対応したレコード群702は、最も小さいタイムスタンプ「80」を含んだRecord A # 80のアドレス（例えば、メモリアドレス）である。そして、Record A # 80内のGC504Aが、Record C # 86のアドレスである。このように、タイムスタンプ「102」に対応したレコード群702は、タイムスタンプ「102」の時点で参照可能なレコード群をリスト形式（リンクドリスト）で管理するときの起点となるレコードのアドレスである。このような構成により、Txの参照タイムスタンプ603から、参照可能なレコード群を特定できる。

[0052] 図8は、GC設定情報である。

[0053] GC設定情報6000は、GC処理の各種設定を格納する。GC設定情報6000は、ユーザ1000又は計算機システム1100が定義する情報であり、メモリ1202及び外部ストレージ装置1300の少なくとも1つに保持される。

[0054] GC設定情報6000は、実行有無801、方式802と、間隔閾値803及び開始閾値804を含む。

- [0055] 実行有無 801 は、GC 処理を実行するか否かのフラグである（「ON」が実行を意味）。
- [0056] 方式 802 は、GC 処理の方式を表す。GC 処理の方式を表す値としては、本願特有の GC 方式を意味する「MULTI」と、レガシーな GC 方式を意味する「OLD」がある。
- [0057] 間隔閾値 803 は、2 つのタイムスタンプ間の差の閾値である。間隔閾値 803 は、GC リスト 2600 に Tx のタイムスタンプを登録するときの判断に用いられる。連続した 2 つの実行中 Tx の参照タイムスタンプ間の差が間隔閾値 803 以上離れていれば、その 2 つの参照タイムスタンプ間のタイムスタンプのレコードが、GC 処理の対象となり得る。「連続した 2 つの実行中 Tx」とは、任意の 1 つの実行中 Tx と、その実行中 Tx の参照タイムスタンプの次に新しい参照タイムスタンプに対応した実行中 Tx である。間隔閾値 803 の値を、Tx の性質等に応じて変えることにより、GC 処理対象のレコードを適切に決定できる。
- [0058] 開始閾値 804 は、時間（時間長）の閾値である。開始閾値 804 は、GC リスト 2600 に Tx のタイムスタンプを登録するときの判断に用いられる。現在時刻よりも開始閾値 804 前以上過去における時刻が開始時刻である Tx のタイムスタンプのレコードが、GC 処理の対象となり得る。開始閾値 804 の値を、Tx の性質等に応じて変えることにより、GC 処理対象のレコードを適切に決定できる。
- [0059] 本実施形態では、クエリ実行部 1313 が、GC 処理対象の候補を例えば次のように決定できる。すなわち、現在時刻よりも開始閾値 804 前以上過去における時刻が開始時刻である Tx のタイムスタンプのレコードを、GC 処理対象の候補とする。また、クエリ実行部 1313 は、現在時刻と現在時刻よりも開始閾値 804 前との間を開始時刻とする Tx のタイムスタンプのレコードについてのみ、間隔閾値 803 以上離れている差を有する 2 つの参照タイムスタンプ（連続した 2 つの実行中 Tx の参照タイムスタンプ）間のタイムスタンプに対応したレコードを特定し、GC 処理対象の候補としてよ

い。これにより、現在時刻と現在時刻から開始閾値 804 前との間にたくさんの更新が発生することによりたくさんのレコードがメモリに格納されることになっても、そのたくさんのレコードから GC 処理対象を適切に選択し削除することができる。勿論、クエリ実行部 1313 は、現在時刻よりも開始閾値 804 前以上過去における時刻が開始時刻である T_x のタイムスタンプのレコードについても、間隔閾値 803 を用いて GC 対象候補のレコードを選定してもよい。間隔閾値 803 及び開始閾値 804 の少なくとも 1 つは、ユーザ 1000 により設定可能でよい。間隔閾値 803 及び開始閾値 804 の一方が設定されていなくてもよい。

[0060] なお、タイムスタンプ値と時刻との間には、相関関係が成り立つので、間隔閾値 803 及び開始閾値 804 の少なくとも 1 つの値は、タイムスタンプ値又は時間情報（値が時刻で表現された情報）でもよい。

[0061] 以下、本実施形態で行われる処理を説明する。

[0062] 図 9 は、GC 処理のフローチャートである。

[0063] GC 処理は、クエリ実行部 1313 により実行される。クエリ実行部 1313 は、GC 処理を開始すると (S3000)、GC 設定情報 6000 の実行有無 801 の値に基づいて処理を分ける (S3100)。例えば、実行有無 801 が「OFF」であれば (S3100:OFF)、クエリ実行部 1313 は、GC 処理を終了する (S3900)。一方、実行有無 801 が「ON」であれば (S3100:ON)、クエリ実行部 1313 は、 T_x ID リスト 2500 を生成する (S3200)。具体的には、クエリ実行部 1313 は、実行中の T_x の情報が登録された T_x ID リスト 2500 を生成する。 T_x ID リスト 2500 のカラムは、参照タイムスタンプ 603 の小さい順に並んでいることが好ましい。参照タイムスタンプ 603 が小さい順に GC 処理が進むためである。

[0064] 次に、クエリ実行部 1313 は、GC 設定情報 6000 の方式 802 に基づいて、実行する処理を分ける (S3300)。具体的には、クエリ実行部 1313 は、方式 802 が「MULTI」であれば、マルチ GC 処理を実行

し（S3600：図10参照）、方式802が「OLD」であれば、旧GC処理（レガシーのGC処理）を実行する（S3700）。方式802の値がTxの性質に応じた値とすることで、適切なGC処理を行うことができる。なお、旧GC処理では、クエリ実行部1313は、TxIDリスト2500に登録されている最先の参照タイムスタンプ102よりも前のタイムスタンプを含むレコードを削除する。

[0065] S3600又はS3700の処理を実行した後、クエリ実行部1313は、GC処理を終了する（S3900）。

[0066] 図10は、マルチGC処理（S3600）のフローチャートである。

[0067] クエリ実行部1313は、マルチGC処理を開始すると（S3610）、GC設定情報6000の間隔閾値803と開始閾値804とを参照する（S3620）。なお、S3620は、S3630にマージされてもよく、その場合、S3620が無しにS5000が行われてよい。

[0068] 次に、クエリ実行部1313は、GCリスト2600に登録されているタイムスタンプに対応するレコード群のGCが可能か否かの判断結果に基づきGCを実行するGCリスト回収処理（S5000：図11参照）を実行する。ここでは、TxIDリスト2500の参照タイムスタンプ603に該当しないタイムスタンプ（すなわち、終了したTxのタイムスタンプ）がGCリスト2600に残っている場合に、そのタイムスタンプに対応するバージョンのレコードが、GCリスト回収処理（S5000）において削除される。

[0069] 次に、クエリ実行部1313は、S3620で参照した間隔閾値803及び開始閾値804に基づいて、GCリスト2600に登録するタイムスタンプを選択し、選択したタイムスタンプをGCリスト2600に登録する（S3630）。具体的には、クエリ実行部1313は、現在時刻より開始閾値804前以上過去における時刻が開始時刻であるTxのタイムスタンプをGCリスト2600に登録することと、間隔閾値803以上離れている差を有する2つの参照タイムスタンプをGCリスト2600に登録することとの少なくとも一方を実行する。

[0070] より具体的には、例えば、現在時刻が「12:10」であり、GC設定情報6000が図8に示す状態であり、TxIDリスト2500が図6に示す状態であり、RecordA~RecordCが、図5A~図5Cに示す状態である場合、現在時刻「12:10」から開始閾値「01:00」前は、時刻「11:10」である。時刻「11:10」以上過去に実行が開始されたTxは、開始時刻602が「10:23」であるTx「1」であり、Tx「1」の参照タイムスタンプは「102」である。開始閾値804を用いてGC対象のTxを選択することの処理負荷は、開始閾値を用いないでGC処理対象のTxを選択することの処理負荷に比べて低い。Txの数が少なく済むからである。

[0071] また、例えば、間隔閾値「30」以上離れている差を有する2つの参照タイムスタンプ（連続した2つの実行中Txの参照タイムスタンプ）は、「102」と「203」の組、及び、「203」と「234」の組である。従って、GCリスト2600に登録されるタイムスタンプは、「102」、「203」及び「234」であり、それらに対応したTxは、Tx「1」、「0」及び「4」である。間隔閾値803を用いてGCリスト2600に登録されるタイムスタンプを選択する理由は、現在時刻と現在時刻から開始閾値804前との間にたくさんの更新が発生すると、メモリに多くのレコードが残っているためである。

[0072] 以上の結果、S3630では、タイムスタンプ701として、「102」、「203」及び「234」が、GCリスト2600に登録される。

[0073] S3630の後、クエリ実行部1313は、GCリスト2600へのレコード群の登録、及び、GCを実行する（S3650）。

[0074] 具体的には、クエリ実行部1313は、登録されたタイムスタンプ701に対応するレコード群702についてのリンクドリストを生成し、リンクドリストの先頭のレコードのアドレスをレコード群702としてGCリスト2600に登録する。ここで、クエリ実行部1313は、タイムスタンプの大きい順にリンクドリストを生成する（リンクドリストにおけるレコードの並

びは、タイムスタンプの大きい順に限らず、他の基準に沿った順序でもよいし、不規則な順序でもよい）。なお、タイムスタンプ701に対応する参照可能なレコード群702（リンクドリストの先頭のレコード）は、タイムスタンプ701より小さなタイムスタンプ501を持つ最新のレコードである。レコードは更新（UPDATE）される度に、既存のレコードは古いレコードとなり、更新されたレコードが「最新のレコード」となる。例えば、図5AのRecord Aで言うと、Record A # 264（タイムスタンプ「264」に対応したRecord A）が最新のRecord Aであり、Record A # 80が、最古のRecord Aである。このとき、タイムスタンプ「130」について参照可能なRecord Aは、Record A # 124である。また、リンクドリストを生成するときに、或るレコードのGC504が既に記載済みであれば、クエリ実行部1313は、そのレコードを含んだレコード群に対応するタイムスタンプ701をGCリスト2600から削除し、且つ、リンクドリストの生成において更新したレコードの504を初期化する。クエリ実行部1313は、リンクドリストの生成において、リンクドリストの最後のレコードのGC504には、終端を意味する値（例えば「Null」）を設定する。このようにレコード群をリンク構造とすることにより、データベースの全てのレコードを順次参照してレコードを探すいわゆるデータベーススキャン処理に比べて、低負荷でGC可能なレコードを見つけることができる。

[0075] レコード群702を登録した後、クエリ実行部1313は、GCを実行する。例えば、クエリ実行部1313は、タイムスタンプ「102」より大きくタイムスタンプ「203」より小さいタイムスタンプのレコードをGC（削除）する場合、以下の処理を実行する。すなわち、クエリ実行部1313は、まず、タイムスタンプ「102」と「203」のうち大きい方のタイムスタンプ「203」に対応するレコード群702が示すRecord A # 155を特定する。クエリ実行部1313は、Record A # 155のPrev503Aを参照して、その更新前Record A # 124を特定する

。Record A # 124 は、GC リスト 2600 のタイムスタンプ「203」の実行中Tx「0」で参照されるレコードでなく、また、タイムスタンプ「203」の1つ前のタイムスタンプ「102」の実行中Tx「1」で参照されるレコードでもないため、GC 可能であると判断する。同様に、クエリ実行部1313は、Record A # 124 のPrev 503 A が示す更新前Record A # 80 をGC 可能であるか否かを判断する。Record A # 80 は、タイムスタンプ「102」の実行中Tx「1」で参照可能なレコードであるため、クエリ実行部1313は、GC 不可能であると判断する。次に、GC 可能であると判断されたRecord A # 124 をGC するために、クエリ実行部1313は、Record A # 80 のNext 502 A を、Record A # 124 のアドレスからRecord A # 155 のアドレスに変更し、Record A # 155 のPrev 503 A を、Record A # 124 のアドレスからRecord A # 80 のアドレスに変更し、その後、Record A # 124 をメモリから削除する。次に、クエリ実行部1313は、Record A # 155 のGC 504 A が示すRecord B # 172（レコード群に登録されているタイムスタンプのうちタイムスタンプ「155」の次に大きいタイムスタンプ「172」を含むレコード）を特定する。クエリ実行部1313は、このような一連の処理と同様な処理を、参照したGC 504 が終端値（NULL）になるまで実行する。この処理により、タイムスタンプ「102」より大きくタイムスタンプ「203」より小さいタイムスタンプ（バージョン）に対応したレコードであって、実行中Txによって参照されないレコードを、メモリから削除することができる。

[0076] 同様に、GC リスト 2600 の他のタイムスタンプ間（例えば、タイムスタンプ「203」と「234」の間）のバージョンのレコードについても、実行中Txにより参照されないレコードをメモリから削除することができる。この処理では、レコード群のリンク（レコード群702が示すアドレス及びGC 504 が示すアドレス）を辿ることにより、データベーススキャン無

しに、GC対象になり得るレコードを特定でき、また、レコードのPrev 503を辿ることにより、データベーススキャン無しに、更新前レコードを特定できるので、処理の負荷を抑えることができる。

[0077] また、クエリ実行部1313は、GCリスト2600の最も小さいタイムスタンプ「102」より小さいタイムスタンプに対応したレコードであって、実行中のTxに参照されていないレコードもメモリから削除する。なお、最小タイムスタンプ「102」より小さいタイムスタンプに対応したレコードは、最小タイムスタンプ「102」に対応したレコード群702から辿ることによって特定される。これにより、最小タイムスタンプ「102」より小さいタイムスタンプ（古いバージョン）のレコードであって、参照されないレコードをメモリから削除できる。

[0078] S3650の後に、クエリ実行部1313は、マルチGC処理を終了する（S3690）。

[0079] 図11は、GCリスト回収処理（S5000）のフローチャートである。

[0080] GCリスト回収処理を開始すると（S5100）、クエリ実行部1313は、GCの実行を指定されたタイムスタンプ（以下、指定タイムスタンプ）があるか否かを判断する（S5200）。この結果、指定タイムスタンプがあれば（S5200：該当あり）、クエリ実行部1313は、処理をS5300に移行し、指定タイムスタンプがなければ（S5200：該当なし）、GCリスト回収処理を終了する（S5900）。

[0081] S5300では、クエリ実行部1313は、指定タイムスタンプのGCを実行する。ここで、指定タイムスタンプがGCリスト2600のタイムスタンプ「203」である場合を例に取る。指定タイムスタンプ「203」に対応するレコード群702は、Record A #155のアドレスを示す。クエリ実行部1313は、Record A #155を削除するために、Record A #155のNext 502 Aが示す更新後Record A #231と、Record A #155のPrev 503 Aが示す更新前Record A #124とを関連付ける。つまり、クエリ実行部1313は、Recor

d A # 2 3 1 の P r e v 5 0 3 A を R e c o r d A # 1 2 4 の アドレスに変更するとともに、R e c o r d A # 1 2 4 の N e x t 5 0 2 A を R e c o r d A # 2 3 1 のアドレスに変更する。これにより、指定タイムスタンプに対応したレコードが削除されてもレコードバージョンの前後関係が維持される。次に、クエリ実行部 1 3 1 3 は、R e c o r d A # 1 5 5 の G C 5 0 4 A が示す R e c o r d B # 1 7 2 を特定し、その R e c o r d # 1 7 2 について、R e c o r d # 1 5 5 と同様の処理を行う。クエリ実行部 1 3 1 3 は、このような処理を、G C 5 0 4 が終端値「N U L L」になるまで実行する。この後、クエリ実行部 1 3 1 3 は、R e c o r d A # 1 5 5 を先頭とするレコード群のレコードをメモリから削除する。

[0082] S 5 3 0 0 を実行した後、クエリ実行部 1 3 1 3 は、G C リスト回収処理を終了する (S 5 9 0 0)。なお、この G C リスト回収処理 (S 5 0 0 0) は、上述した契機に代えて又は加えて別の契機に実行されてもよい。

[0083] 図 1 2 は、T x 処理のフローチャートである。

[0084] クエリ実行部 1 3 1 3 は、T x 処理を開始し、T x I D リスト 2 5 0 0 に、その T x に対応した T x I D 6 0 1、開始時刻 6 0 2、参照タイムスタンプ 6 0 3、更新タイムスタンプ 6 0 4、及び、状態 6 0 5 を含んだカラムを追加する。参照タイムスタンプ 6 0 3 は、現在のタイムスタンプ (最新のタイムスタンプ) の値である。その後、クエリ実行部 1 3 1 3 は、データベースのレコードの参照及び更新のうちの少なくとも 1 つを含む S Q L 処理を実行する (S 4 1 0 0)。この際、クエリ実行部 1 3 1 3 は、S Q L 処理の内容に応じて T x I D リスト 2 5 0 0 の状態 6 0 5 を更新する。このため、S 4 0 0 0 において登録された状態 6 0 5 は、不定値でよい。なお、S 4 0 0 0 において登録された状態 6 0 5 は、クエリソース (例えば計算機システム 1 1 1 0) から宣言された値でもよい。また、S Q L 処理とは、受け付けたクエリに基づき生成されたクエリ実行プランに従う処理である。

[0085] 次に、クエリ実行部 1 3 1 3 は、T x の後処理 (C O M M I T 処理) として、終了する T x に対応する更新タイムスタンプを取得し、ログ出力を行う

(S4200)。次に、クエリ実行部1313は、Tx終了時のGCリスト回収処理を実行する(S5000)。このGCリスト回収処理では、クエリ実行部1313は、終了するTxのTxIDに対応する参照タイムスタンプ603と同じ値のタイムスタンプ701が、GCリスト2600にあれば、このタイムスタンプ701に対応したレコードのGCを実行する。GCリスト回収処理(S5000)の終了後、クエリ実行部1313は、Tx処理を終了する(S4900)。

[0086] このTx処理においては、Txが参照していたバージョンのレコードがGC対象であれば、このレコードのGCをTxの終了時に(COMMIT)後に行うことができる。

[0087] 以上、実施形態を説明したが、本発明は、この実施形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。例えば、メモリ1202は、記憶部の一例でよい。記憶部は、メモリ1202に代えて又は加えて、補助記憶デバイス(典型的には不揮発性の記憶デバイス)を含んでもよい。

符号の説明

[0088] 1100: 計算機システム 1200: ノード 1201: プロセッサ
1202: メモリ 1301: DBMS

請求の範囲

[請求項1] 記憶部に格納されており複数のレコードを有するデータベースに対するクエリを受け付ける受付部と、

前記クエリに従いトランザクションを実行することによりいずれかのレコードを更新する場合、タイムスタンプが関連付けられている更新前レコードを前記記憶部に維持し、レコードの更新に伴いタイムスタンプを更新し、更新後のタイムスタンプを関連付けた更新後レコードを前記記憶部に格納する実行部とを有し、

前記実行部は、

実行中の第1のトランザクションに対応付けられたタイムスタンプである第1のタイムスタンプと、実行中の第2のトランザクションに対応付けられたタイムスタンプである第2のタイムスタンプとの間のタイムスタンプに対応したレコードのうち、実行中の他のトランザクションから参照されないレコードを、複数のタイムスタンプの各々での複数のレコードを基に特定し、

特定したレコードを前記記憶部から削除する、データベースシステム。

[請求項2] 前記第1のトランザクションは、任意の1つの実行中のトランザクションであり、

前記第2のトランザクションは、実行中のトランザクションのうちの、前記第1のタイムスタンプの次に値が大きいタイムスタンプに対応したトランザクションであり、

前記第1及び第2のタイムスタンプは、それらの差分がタイムスタンプ間の差分の閾値である間隔閾値以上である、請求項1記載のデータベースシステム。

[請求項3] 前記実行部が、現在時刻よりも所定時間前以上過去において開始された実行中トランザクションに対応したタイムスタンプのレコードを

前記記憶部から削除し、

前記実行部は、現在時刻と現在時刻よりも所定時間前との間に開始された実行中のトランザクションから、前記間隔閾値を用いて、前記第1及び第2のトランザクションに該当するトランザクションを特定する、

請求項2記載のデータベースシステム。

[請求項4] 前記実行部が、現在時刻よりも所定時間前以上過去において開始された実行中トランザクションに対応したタイムスタンプのレコードを前記記憶部から削除し、

前記実行部は、現在時刻よりも所定時間前以上過去において開始された実行中トランザクションに対応したタイムスタンプと、前記第1及び第2のタイムスタンプとのうちの最も古いタイムスタンプよりも古いタイムスタンプのレコードを前記記憶部から削除する、

請求項3記載のデータベースシステム。

[請求項5] 前記間隔閾値は、ユーザから入出力デバイス経由で設定された値である、

請求項2記載のデータベースシステム。

[請求項6] 前記実行部は、タイムスタンプ毎に、そのタイムスタンプの時点でのレコード群を構成する2以上のレコードがシーケンシャルになるようレコード間を関連付ける、

前記実行部は、対象のタイムスタンプの時点でのレコード群をレコード間の関連付けを辿ることで特定する、

請求項1記載のデータベースシステム。

[請求項7] 前記実行部は、前記第1及び第2のタイムスタンプの各々について、そのタイムスタンプと、そのタイムスタンプの時点でのレコード群の先頭のレコードのアドレスとを、削除対象のレコードに関する情報が登録される削除管理情報に登録し、

前記実行部は、前記第1及び第2のタイムスタンプの各々について

、前記削除管理情報を参照し、そのタイムスタンプに対応したアドレスからレコード間の関連付けを辿ることで、そのタイムスタンプの時点でのレコード群を特定する、

請求項 6 記載のデータベースシステム。

[請求項8] 前記実行部は、更新前レコードにその更新前レコードの更新後レコードのアドレスを関連付け、その更新後レコードにその更新後レコードの更新前レコードのアドレスを関連付け、

前記実行部は、前記 2 のタイムスタンプに対応したレコードに関連付けられているアドレスに従いそのレコードの更新前レコードを特定し、特定した更新前レコードに対応するタイムスタンプを特定する、請求項 6 記載のデータベースシステム。

[請求項9] 前記実行部は、

削除対象のレコードの更新後レコードに関連付けられている更新前レコードアドレスを、前記削除対象のレコードの更新前レコードアドレスに変更し、

前記削除対象のレコードの更新前レコードに関連付けられている更新後レコードアドレスを、前記削除対象のレコードの更新後レコードアドレスに変更する、

請求項 1 記載のデータベースシステム。

[請求項10] 前記実行部は、

トランザクションをコミットした後に、そのトランザクションが参照可能なレコードが削除対象か否かを判断し、

その判断結果が肯定の場合に、その削除対象のレコードを前記記憶部から削除する、

請求項 1 記載のデータベースシステム。

[請求項11] 前記実行部は、複数の方式のうちから選択されている方式に従ってレコード削除を実行し、

前記複数の方式の 1 つが、前記第 1 及び第 2 のタイムスタンプ間の

タイムスタンプに対応したレコードのうち実行中の他のトランザクションから参照されないレコードを削除する方式である、
請求項 1 記載のデータベースシステム。

- [請求項12] 記憶部内のデータベースが有する複数のレコードのうちのいずれかのレコードを、トランザクションを実行することにより更新する場合、タイムスタンプが関連付けられている更新前レコードを記憶部に維持し、レコードの更新に伴いタイムスタンプを更新し、
更新後のタイムスタンプを関連付けた更新後レコードを記憶部に格納し、
実行中の第 1 のトランザクションに対応付けられたタイムスタンプである第 1 のタイムスタンプと、実行中の第 2 のトランザクションに対応付けられたタイムスタンプである第 2 のタイムスタンプとの間のタイムスタンプに対応したレコードのうち、実行中の他のトランザクションから参照されないレコードを、複数のタイムスタンプの各々での複数のレコードを基に特定し、
特定したレコードを前記記憶部から削除する、
データベース管理方法。

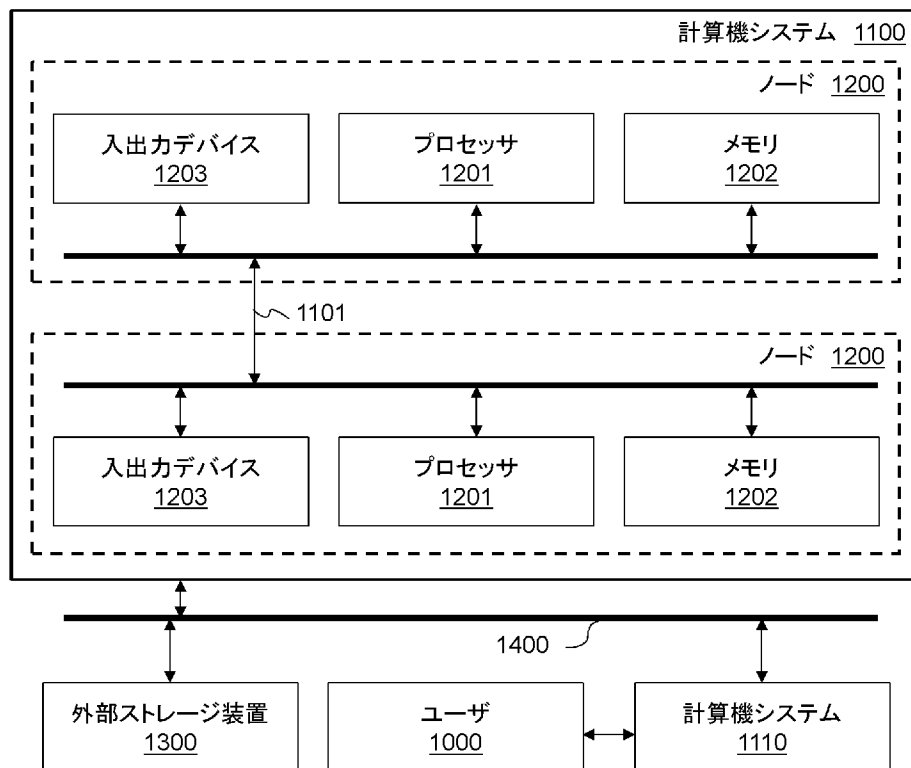
- [請求項13] 複数のレコードを記憶する記憶部と、
トランザクションを実行することによりいずれかのレコードを更新する場合、タイムスタンプが関連付けられている更新前レコードを前記記憶部に維持し、レコードの更新に伴いタイムスタンプを更新し、更新後のタイムスタンプを関連付けた更新後レコードを前記記憶部に格納するプロセッサと
を有し、
前記プロセッサは、
実行中の第 1 のトランザクションに対応付けられたタイムスタンプである第 1 のタイムスタンプと、実行中の第 2 のトランザクションに対応付けられたタイムスタンプである第 2 のタイムスタンプとの間

のタイムスタンプに対応したレコードのうち、実行中の他のトランザクションから参照されないレコードを、複数のタイムスタンプの各々での複数のレコードを基に特定し、

特定したレコードを前記記憶部から削除する、
計算機システム。

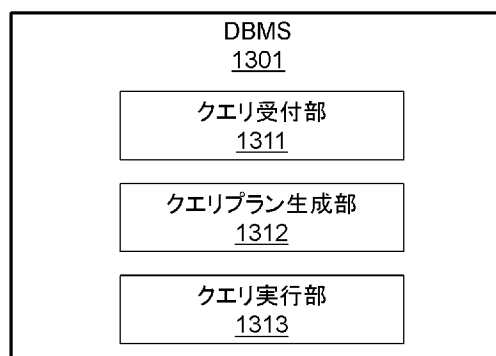
[図1]

FIG. 1



[図2]

FIG. 2



[図3]

FIG. 3

データ遷移管理情報 - 2000

301	タイムスタンプ	0	1	2	3	4
302A	RecordA	100	101		102	
302B	RecordB	200		201		
302C	RecordC	300			301	302

[図4]

FIG. 4

TxIDリスト - 2100

401	TxID	2	0	1	5	3
402	開始時刻	10:23	11:45	12:01	12:04	12:06
403	参照タイムスタンプ	0	0	1	1	0
404	更新タイムスタンプ	-	1	-	2	-
405	状態	R	RW	R	RW	R

[図5A]

FIG. 5A

RecordAの遷移

501A	タイムスタンプ	80	124	155	231	264
502A	Next	*	*	*	*	*
503A	Prev	*	*	*	*	*
504A	GC	*	*	*	*	*
505A	Data	...	...	...	...	...

[図5B]

FIG. 5B

RecordBの遷移						
501B	タイムスタンプ	141	172	219	232	263
502B	Next	*	*	*	*	*
503B	Prev	*	*	*	*	*
504B	GC	*	*	*	*	*
505B	Data	...	...	...	...	...

[図5C]

FIG. 5C

RecordCの遷移						
501C	タイムスタンプ	52	86	103	151	204
502C	Next	*	*	*	*	*
503C	Prev	*	*	*	*	*
504C	GC	*	*	*	*	*
505C	Data	...	...	...	...	...

[図6]

FIG. 6

TxIDリスト - 2500

601	TxID	1	0	4	3	2
602	開始時刻	10:23	11:45	12:01	12:04	12:06
603	参照タイムスタンプ	102	203	234	246	251
604	更新タイムスタンプ	-	-	249	-	-
605	状態	R	R	RW	R	RW

[図7]

FIG. 7

GCリスト - 2600

701	タイムスタンプ	102	203	234		
702	レコード群	*	*	*		

[図8]

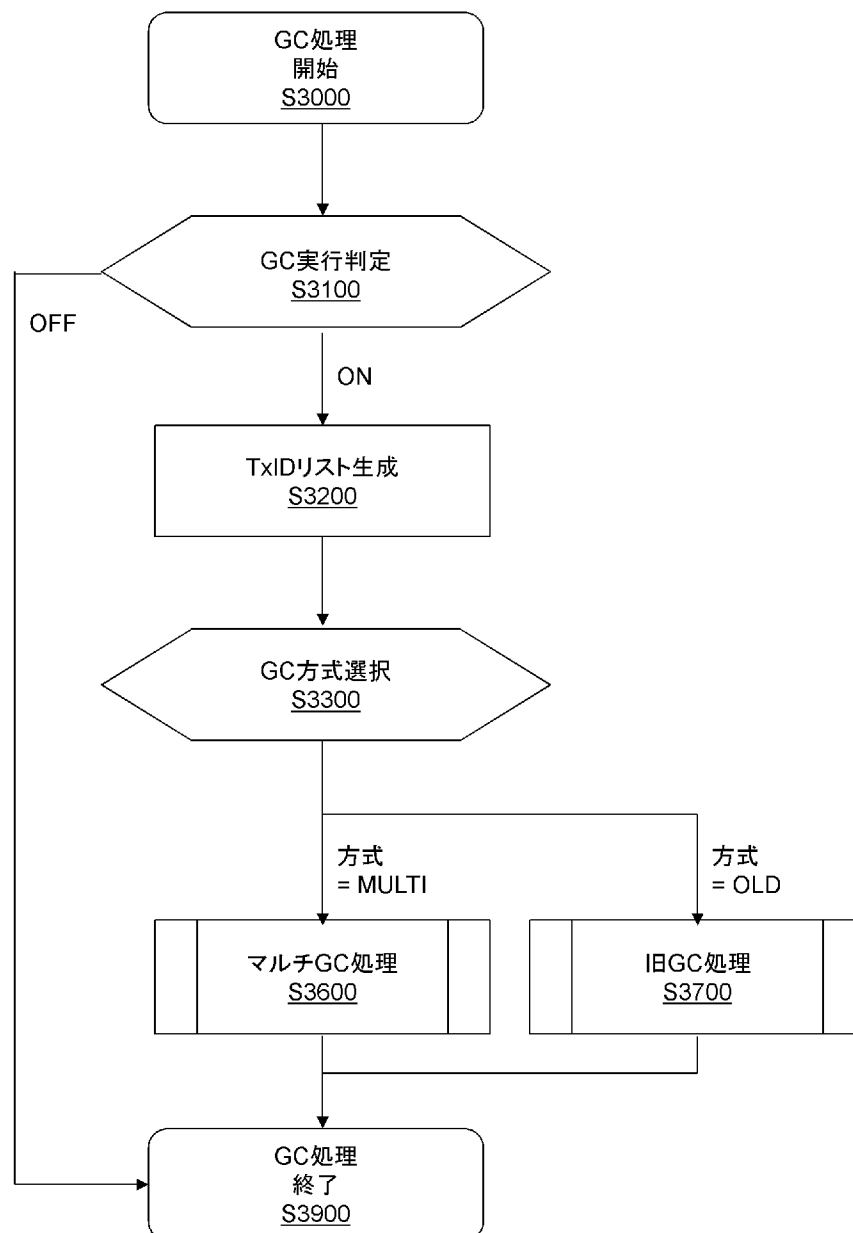
FIG. 8

GC設定情報 - 6000

801	実行有無	ON
802	方式	MULTI
803	間隔閾値	30
804	開始閾値	01:00

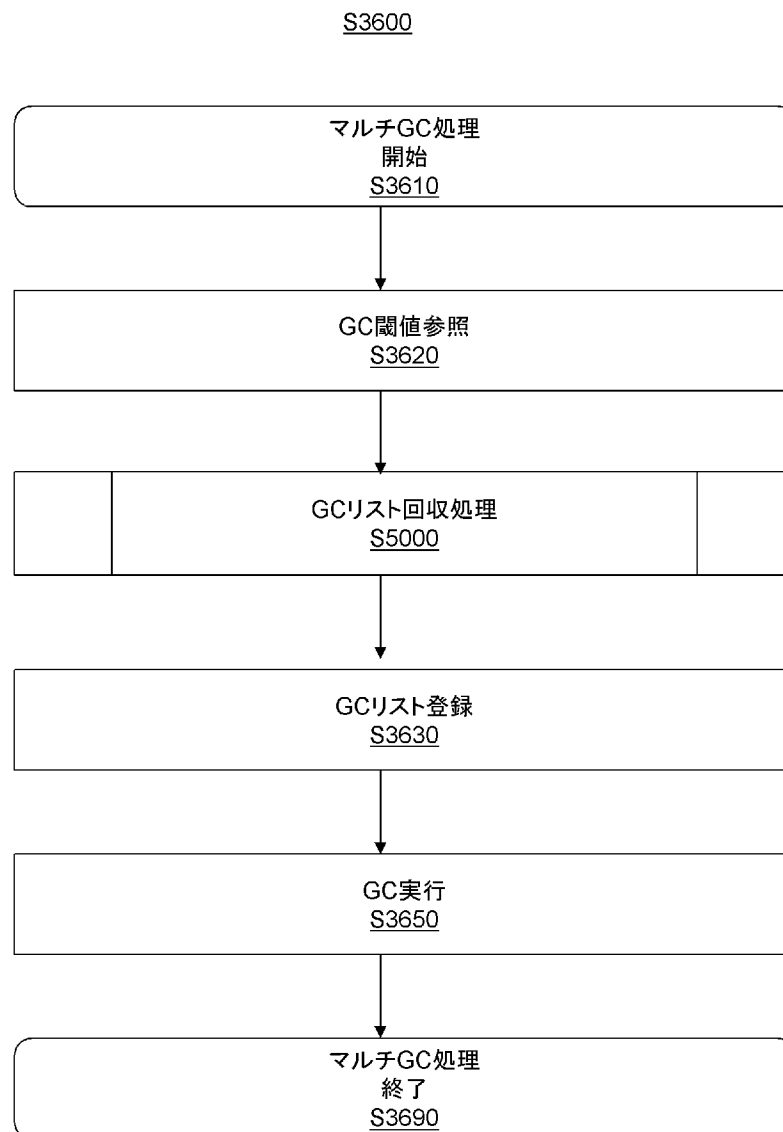
[図9]

FIG. 9



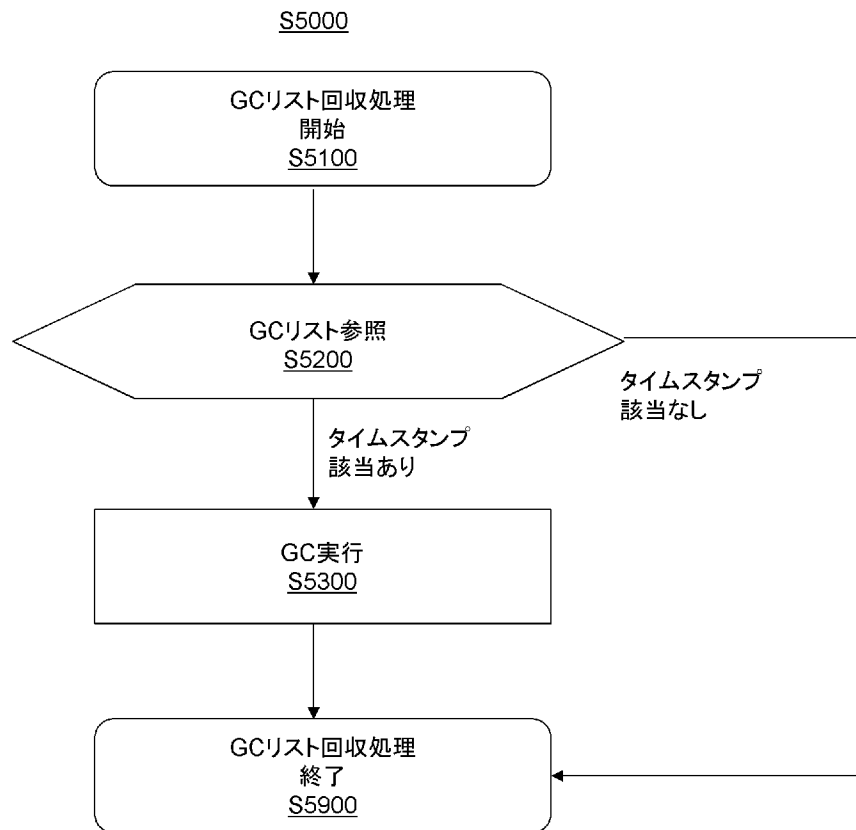
[図10]

FIG. 10



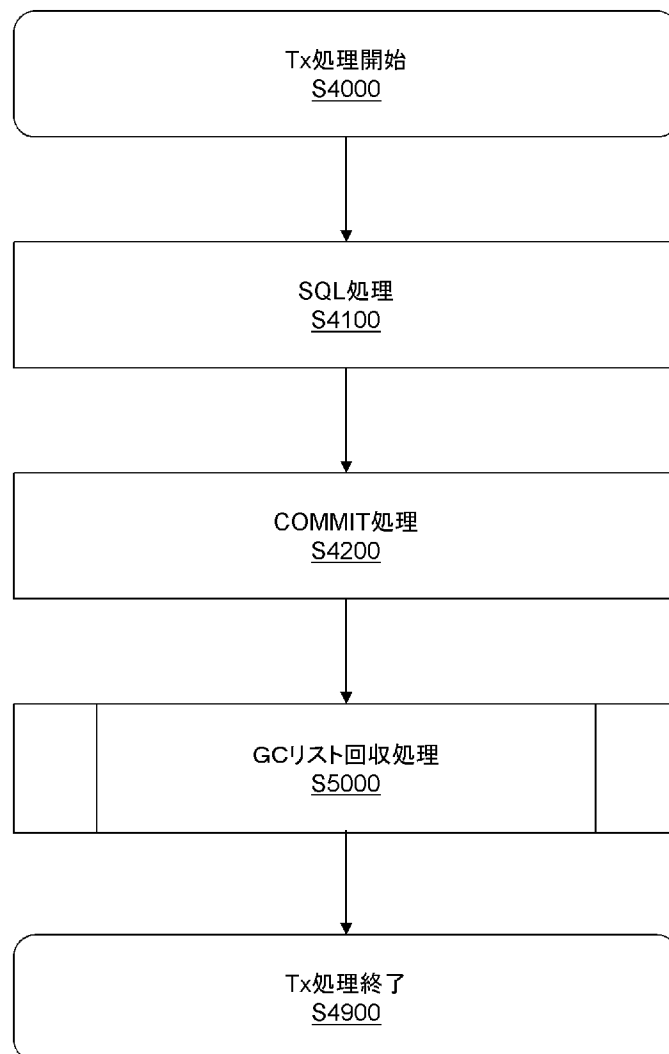
[図11]

FIG. 11



[図12]

FIG. 12



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/051414

A. CLASSIFICATION OF SUBJECT MATTER

G06F12/00(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

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	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

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 Y A	EP 2738697 A1 (SAP AG), 04 June 2014 (04.06.2014), paragraphs [0006], [0017], [0031]; fig. 4 & US 2014/0149353 A1	1, 12, 13 10 2-9, 11
Y	US 2011/0252000 A1 (MICROSOFT CORP.), 13 October 2011 (13.10.2011), fig. 4 (Family: none)	10
A	US 2011/0153566 A1 (MICROSOFT CORP.), 23 June 2011 (23.06.2011), entire text; all drawings (Family: none)	1-13

☒ 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
03 March 2015 (03.03.15)

Date of mailing of the international search report
10 March 2015 (10.03.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/051414

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2013-522778 A (NuoDB, Inc.), 13 June 2013 (13.06.2013), entire text; all drawings & US 2011/0231447 A1 & EP 2548135 A2 & CN 103026355 A & RU 2012144349 A	1-13
A	JP 2007-501468 A (Oracle International Corp.), 25 January 2007 (25.01.2007), entire text; all drawings & US 2005/0033720 A1 & EP 1652084 A2 & CN 1864137 A	1-13

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

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

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 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 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 Y A	EP 2738697 A1 (SAP AG) 2014. 06. 04, [0006], [0017], [0031], Fig. 4 & US 2014/0149353 A1	1, 12, 13 10 2-9, 11
Y	US 2011/0252000 A1 (MICROSOFT CORPORATION) 2011. 10. 13, Fig. 4 (ファミリーなし)	10
A	US 2011/0153566 A1 (MICROSOFT CORPORATION) 2011. 06. 23, 全文, 全図 (ファミリーなし)	1-13

☒ 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

0 3 . 0 3 . 2 0 1 5

国際調査報告の発送日

1 0 . 0 3 . 2 0 1 5

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）

野田 佳邦

5 U

3 4 5 0

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 6 5

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	JP 2013-522778 A (ヌオディービー インコーポレイテッド) 2013.06.13, 全文, 全図 & US 2011/0231447 A1 & EP 2548135 A2 & CN 103026355 A & RU 2012144349 A	1-13
A	JP 2007-501468 A (オラクル・インターナショナル・コーポレイシ ョン) 2007.01.25, 全文, 全図 & US 2005/0033720 A1 & EP 1652084 A2 & CN 1864137 A	1-13