

(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
2015年10月1日(01.10.2015)



(10) 国際公開番号
WO 2015/145587 A1

- (51) 国際特許分類:
G06F 12/00 (2006.01) G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2014/058382
- (22) 国際出願日: 2014年3月25日(25.03.2014)
- (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Murakumo (MURAKUMO CORPORATION) [JP/JP]; 〒1530061 東京都目黒区中目黒一丁目1番45号 Tokyo (JP).
- (72) 発明者: 山田 浩之 (YAMADA, Hiroyuki); 〒1530061 東京都目黒区中目黒一丁目1番45号 株式会社Murakumo内 Tokyo (JP).
- (74) 代理人: 畑添 隆人 (HATAZOE, Takahito); 〒1020072 東京都千代田区飯田橋二丁目1番4号 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, LR, LS, LT, 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.

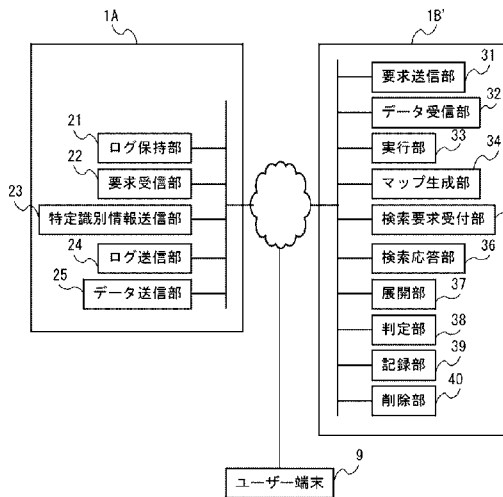
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーロシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 国際調査報告 (条約第 21 条(3))

(54) Title: DATABASE SYSTEM, INFORMATION PROCESSING DEVICE, METHOD, AND PROGRAM

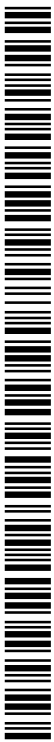
(54) 発明の名称: データベースシステム、情報処理装置、方法およびプログラム



- 9 User terminal
- 21 Log maintaining unit
- 22 Request receiving unit
- 23 Specific identification information transmitting unit
- 24 Log transmitting unit
- 25 Data transmitting unit
- 31 Request transmitting unit
- 32 Data receiving unit
- 33 Execution unit
- 34 Map generating unit
- 35 Search request accepting unit
- 36 Search response unit
- 37 Expansion unit
- 38 Determination unit
- 39 Recording unit
- 40 Deleting unit

(57) Abstract: A database system wherein: a first node, in accordance with a data transmission request from a second node, transmits to the second node identification information, specific identification information that represents a transaction log at a prescribed point in time, and the transaction log after the prescribed point in time; the first node transmits the requested data to the second node from the prescribed point in time onward; the second node, in accordance with a search request, transmits a data transmission request to the first node; the second node, when data received from the first node is expanded in a memory, executes for the expanded data instructions that pertain to the transaction log that is newer than the transaction log that is represented by the specific information; and the second node, on the basis of the data for which the instructions are executed, responds to the search request and, according to prescribed conditions, deletes the data for which the instructions are executed.

(57) 要約: データベースシステムにおいて、第一のノードは、第二のノードからのデータの送信要求に応じて、所定の時点におけるトランザクションログを示す特定識別情報、所定の時点より後のトランザクションログおよび識別情報を第二のノードに送信し、要求されたデータを所定の時点以降に第二のノードに送信し、第二のノードは、検索要求に応じて第一のノードにデータの送信要求を送信し、第一のノードから受信されたデータがメモリに展開された場合に、特定識別情報が示すトランザクションログよりも新しいトランザクションログに係る命令を展開されたデータに対して実行し、命令が実行されたデータに基づいて検索要求へ応答し、命令が実行されたデータを所定の条件に従って削除することとした。



WO 2015/145587 A1

明 細 書

発明の名称：

データベースシステム、情報処理装置、方法およびプログラム

技術分野

[0001] 本発明は、データベースを管理するための技術に関する。

背景技術

[0002] 従来、自己の障害に備えて他のサーバノードから送信されるトランザクションログを受信し、受信されたトランザクションログをデータベースに反映するデータベース管理システムが提案されている（特許文献1を参照）。

[0003] また、マスターが、複数のトランザクションログを並列に作成し、並列に作成した複数のトランザクションログを他のマスター又はスレーブに送信し、他のマスターまたはスレーブが、当該複数のトランザクションログをデータベースに並列に適用する技術が提案されている（特許文献2を参照）。

先行技術文献

特許文献

[0004] 特許文献1：特表2012-532376号公報

特許文献2：特開2012-133417号公報

発明の概要

発明が解決しようとする課題

[0005] 従来、複数のノードを有するデータベースシステムにおいて、ユーザー端末からのクエリを処理するための負荷を複数のノードで分担するための方法が種々提案されている。例えば、（1）ログベースのレプリケーションを全てのノードに対して行うことによって全ノード間でデータを同期し、何れのノードでもクエリへの応答を可能とする方法や、（2）各ノードがデータの一部のみを複製し、複製した部分に対するリクエストを処理する方法、方法、等が提案されている。

[0006] しかし、（1）全ノード間でデータを同期する方法では、全てのノードが

大きなストレージまたはメモリを有する必要がある、また、同期のために用いるネットワークリソースも大きくなる。また、(2)各ノードがデータの一部のみを複製し、複製した部分に対するリクエストを処理する方法では、クエリの処理に必要なデータが複数のノードに分かれて保持されている場合に、データを統合するためのリソースが必要となるという問題があった。

[0007] 本発明は、上記した問題に鑑み、データベースシステムにおいて、ストレージやメモリを節約しながら、複数のノードにおいてユーザー端末からのクエリを受け付けることを課題とする。

課題を解決するための手段

[0008] 本発明は、上記した課題を解決するために、以下の手段を採用した。即ち、本発明は、複数のノードを有するデータベースシステムであって、前記複数のノードのうち、第一のノードは、該第一のノードによって管理されているデータベースのトランザクションログを、該トランザクションログに係る命令の順序を把握可能な識別情報とともに保持するログ保持手段と、前記複数のノードのうち、前記第一のノードと異なる第二のノードから、該第一のノードによって管理されているデータの送信要求を受信する要求受信手段と、前記送信要求に応じて、前記ログ保持手段によって保持されている前記識別情報のうち、要求されたデータに係る、所定の時点におけるトランザクションログを示す特定識別情報を、前記第二のノードに送信する特定識別情報送信手段と、少なくとも前記所定の時点より後の前記トランザクションログおよび前記識別情報を、互いに関連づけて前記第二のノードに送信するログ送信手段と、前記送信要求に応じて、前記データベースによって管理されているデータのうち要求されたデータを、前記所定の時点以降に、前記第二のノードに送信するデータ送信手段と、を備え、前記第二のノードは、ユーザー端末からの検索要求を受け付ける検索要求受付手段と、前記検索要求に応じて、前記第一のノードに対して、前記データベース中の少なくとも前記検索要求に係るデータの前記送信要求を送信する要求送信手段と、前記第一のノードから、前記トランザクションログ、該トランザクションログの識別情

報、前記特定識別情報および前記データを受信する受信手段と、受信された前記データが該第二のノードのメモリに展開されてデータの検索または処理に供される状態となった場合に、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに係る命令を、メモリに展開された前記データに対して実行する実行手段と、前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータに基づいて、前記検索要求への応答を前記ユーザー端末に対して返す検索応答手段と、前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータを、前記検索応答手段による応答に用いた後、所定の条件に従って自動的に該第二のノードから削除する削除手段と、を備える、データベースシステムである。

[0009] 本発明に係るデータベースシステムでは、第一のノードは、検索要求を受けた第二のノードからの要求に応じて、所定の時点におけるトランザクションログを示す特定識別情報、所定の時点以降のトランザクションログ、および所定の時点以降のデータを第二のノードに送信する。そして、第二のノードは、受信されたデータをメモリに展開して検索等に供される状態とし（例えば、データを含むページを所謂キャッシュとし）、所定の時点以降のトランザクションログに係る命令を、メモリに展開されたデータ（キャッシュ）に対して実行する。即ち、本発明では、検索要求に応じてデータを取得し、検索等の用に供されるためにメモリに展開された時点でトランザクションログに係る命令を実行（トランザクションログを反映）することで、レプリケーションに用いるストレージやメモリを節約しながら、ユーザー端末からのクエリを受け付けることとしている。

[0010] また、前記削除手段は、前記データに関する処理に関する時点から所定時間経過した場合に、前記データを削除してもよい。

[0011] また、前記第二のノードは、前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータを、該第二の

ノードにおける保持対象とするか否かを判定する判定手段を更に備え、前記削除手段は、前記判定手段によって前記保持対象と判定されなかったデータを、前記所定の条件に従って自動的に削除してもよい。

[0012] また、前記第二のノードは、前記判定手段によって前記保持対象と判定されたデータを、不揮発性の記憶装置に記録する記録手段を更に備えてもよい。

[0013] また、前記データ送信手段は、前記データを、前記トランザクションログに係る命令の順序が互いに依存関係にあるレコードが同一の管理単位に入るように区切られた所定の管理単位毎に送信してもよい。

[0014] また、前記データベースシステムは、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに基づいて、命令の対象となるデータが収容された前記所定の管理単位と命令の内容との関係を示すマップを生成するマップ生成手段を更に備え、前記実行手段は、前記マップを参照して、前記トランザクションログに係る命令を、前記所定の管理単位毎に実行してもよい。また、前記データベースシステムは、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに基づいて、命令の対象となるデータと命令の内容との関係を示すマップを生成するマップ生成手段を更に備え、前記実行手段は、前記マップを参照して、前記トランザクションログに係る命令を、命令の対象となるデータに対して実行してもよい。

[0015] また、前記データ送信手段は、前記第一のノードのメモリに展開されてデータの検索または処理に供されている状態のデータを、前記第二のノードに送信してもよい。

[0016] また、前記データ送信手段は、前記データベースの管理情報を、前記第二のノードに送信し、前記要求送信手段は、前記管理情報を参照して、前記第一のノードに対して、前記データベース中のデータを指定して送信要求を送信してもよい。

[0017] また、前記第二のノードは、前記第一のノードから受信したデータを直接メモリに展開して、データの検索または処理に供する展開手段を更に備え、前記実行手段は、受信したデータが前記展開手段によってメモリに展開されたことを受けて、前記トランザクションログに係る命令を前記データに対して実行してもよい。

[0018] また、前記特定識別情報送信手段は、前記特定識別情報として、ログ保持手段によって保持されている前記識別情報のうち、最新のトランザクションログを示す識別情報を送信してもよい。

[0019] また、前記データは、テーブル単位またはページ単位で送受信されてもよい。

[0020] また、前記実行手段は、前記トランザクションログに係る命令を、前記識別情報によって把握される命令の順序に応じて、受信された前記データに対して実行してもよい。

[0021] なお、本発明は、コンピューターシステム、情報処理装置、コンピューターによって実行される方法、またはコンピューターに実行させるプログラムとして把握することが可能である。また、本発明は、そのようなプログラムをコンピューターその他の装置、機械等が読み取り可能な記録媒体に記録したものとしても把握できる。ここで、コンピューター等が読み取り可能な記録媒体とは、データやプログラム等の情報を電氣的、磁氣的、光学的、機械的、または化学的作用によって蓄積し、コンピューター等から読み取ることができる記録媒体をいう。

発明の効果

[0022] 本発明によれば、データベースシステムにおいて、ストレージやメモリを節約しながら、複数のノードにおいてユーザー端末からのクエリを受け付けることが可能となる。

図面の簡単な説明

[0023] [図1]実施形態に係るシステムのハードウェア構成の概略を示す図である。

[図2]実施形態に係るシステムの機能構成の概略を示す図である。

[図3]実施形態において、第一のノードおよび第二のノードによって実行される管理情報送受信処理の流れを示すフローチャートである。

[図4]実施形態において、第一のノードおよび第二のノードによって実行されるデータベース送受信処理の流れを示すフローチャートである。

[図5]実施形態において、第二のノードによって実行されるマップ生成処理の流れを示すフローチャートである。

[図6]実施形態において、第二のノードによって実行される更新実行処理の流れを示すフローチャートである。

[図7]実施形態において、第二のノードによって実行される検索要求対応処理の流れを示すフローチャート（1）である。

[図8]実施形態において、第二のノードによって実行される検索要求対応処理の流れを示すフローチャート（2）である。

[図9]実施形態に係る第一のノードおよび第二のノードの機能構成の概略を示す図である。

[図10]実施形態において、第一のノードおよび第二のノードによって実行されるトランザクションログ送受信開始処理の流れを示すフローチャートである。

[図11]実施形態において、第二のノードによって実行されるオンデマンド処理の流れを示すフローチャートである。

[図12]実施形態において、第二のノードによって実行される削除処理の流れ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

[0024] 以下、本開示に係るシステム、情報処理装置、方法およびプログラムの実施の形態を、図面に基づいて説明する。但し、以下に説明する実施の形態は、実施形態を例示するものであって、本開示に係るシステム、情報処理装置、方法およびプログラムを以下に説明する具体的構成に限定するものではない。実施にあたっては、実施形態に応じた具体的構成が適宜採用され、また、種々の改良や変形が行われてよい。

[0025] 本実施形態では、本開示に係るシステム、情報処理装置、方法およびプログラムを、追記型のデータベースシステムにおいて実施した場合の実施の形態について説明する。追記型のデータベースシステムとは、データの更新の際に、古いデータを新しいデータで上書きすることなく、新しいデータを追記することでデータの更新を行うタイプのデータベースシステムである。但し、本開示に係るシステム、情報処理装置、方法およびプログラムは、複数のノードを有するシステムにおいて、あるノードが管理するデータを他のノードにおいて用いるための技術について広く用いることが可能であり、本開示の適用対象は、本実施形態において示した例に限定されない。

[0026] <第一の実施形態>

はじめに、第一の実施形態について説明する。

[0027] <<システムの構成>>

図1は、本実施形態に係るシステムのハードウェア構成の概略を示す図である。本実施形態に係るシステムは、ユーザー端末9からの検索要求（クエリ）に応答するための複数のノード（情報処理装置）1を備える。複数のノード1は、ネットワークを介して互いに通信可能に接続される。複数のノード1のうち、データベースが既に構築されているノード1の何れかを、本実施形態におけるオリジナルのデータベースを有する第一のノード1Aとすることが出来る。また、本実施形態において第一のノード1Aからデータを受信する第二のノード1Bは、データベース用のソフトウェアがインストールされているが、データベースのコンテンツが未構築のノードである。本実施形態では、このようなシステムにおいて、第一のノード1Aが管理するデータベースのコンテンツ（データ）を、第二のノード1Bにおける検索の用に供するための技術を説明する。本実施形態では、ノードを区別することなくノード一般について述べる場合には「ノード1」と記載し、ノードを区別して述べる場合には「第一のノード1A」、「第二のノード1B」のように添字を付して記載する。

[0028] なお、本開示に係るシステムは、階層構造を有するデータベースシステム

や、マルチマスターのデータベースシステムに適用可能であるが、この際、データベースシステムにおけるマスターノードおよびスレーブノードの何れであっても、第一のノード1Aまたは第二のノード1Bとすることが出来る。

[0029] 第一のノード1Aおよび第二のノード1Bは、CPU (Central Processing Unit) 11A、11B、RAM (Random Access Memory) 12A、12BおよびROM (Read Only Memory) 13A、13B等からなる制御部10A、10Bと、補助記憶装置14A、14Bと、通信インターフェース15A、15Bと、を備えるコンピューターである。但し、ノード1の具体的なハードウェア構成に関しては、実施の形態に応じて適宜省略や置換、追加が可能である。また、ノード1は、単一の装置に限定されない。ノード1は、所謂クラウドや分散コンピューティングの技術等を用いた、複数の装置によって実現されてよい。

[0030] なお、本実施形態では、データベースの各レコードは、ページ単位で管理されており、ストレージ（例えば、補助記憶装置14A、14B）とメモリ等（例えば、RAM12A、12B）との間のデータのやり取りは、ページ単位で行われる。ここで、メモリ等に展開されてデータの検索または処理に供されている状態のデータは、キャッシュと称される。一方、ストレージ上のデータは、データの検索または処理に供されていない状態にある。即ち、データベースの内容を検索や更新等の処理の対象とする場合、各ノードは、対象のレコードを含むページをストレージから読み出してキャッシュとし、対象のレコードに係る処理を行う。また、各ノードは、対象のレコードの更新を行った後に、当該レコードを含むキャッシュのページをストレージに書き込むことで、データベースの更新を保存する。なお、ページがストレージに書き込まれるタイミングは、実施の形態に応じて適宜決定される。ページには、1または複数のレコードが格納される。また、ページは、ブロックと称されてもよく、本発明の「所定の管理単位」に相当する。

[0031] 本実施形態では、後述するノード間の送受信処理においても、ページ単位でデータが要求され、ページ単位でデータが送受信される。但し、データの読出／書込および送受信をページ単位で行うのは、本開示を実施する際に採用可能な具体的構成の一例であり、本開示の技術的範囲は、ページ単位での読出／書込および送受信に限定されない。データの読出／書込および送受信は、例えばテーブル単位で行われてもよいし、その他の管理単位で行われてもよい。

[0032] 図2は、本実施形態に係る第一のノード1Aおよび第二のノード1Bの機能構成の概略を示す図である。本実施形態に係る第一のノード1Aは、CPU11Aが、RAM12Aに展開された各種プログラムを解釈および実行して、ノード1Aに備えられた各種ハードウェアを制御することで、ログ保持部21、要求受信部22、特定識別情報送信部23、ログ送信部24およびデータ送信部25を備えるコンピューターとして機能する。本実施形態では、第一のノード1Aの機能が汎用のCPU11Aによって実行される例について説明しているが、これらの機能は、その一部または全部が、1または複数の専用のプロセッサによって実現されてもよい。

[0033] ログ保持部21は、第一のノード1Aによって管理されているデータベースにおけるトランザクションのログ（以下、「トランザクションログ」）を、トランザクションに含まれる命令の時系列上の順序を把握可能な識別情報であるLSN(Log Sequence Number)とともに保持する。命令には、データベースの更新に係る更新命令と、トランザクションの管理に係る管理命令とが含まれる。また、トランザクションログは、ノードにおいてトランザクションが発生する毎に生成され、LSNが付されてログ保持部21によって保持される。

[0034] 要求受信部22は、第二のノード1Bによって送信された、第一のノード1Aによって管理されているデータの送信要求を受信する。

[0035] 特定識別情報送信部23は、ログ保持部21によって保持されているLSNのうち、所定の時点におけるトランザクションログを示す特定識別情報（

特定L S N) を、要求受信部 2 2 によって受信されたデータ送信要求に応じて第二のノード 1 B に送信する。ここで送信される特定L S N は、要求されたページ毎に決定されてよい。なお、本実施形態において、特定識別情報送信部 2 3 は、特定L S N として、最新のトランザクションログを示すL S N を送信する。特定L S N を最新のトランザクションログを示すL S N とすることで、ログ送信部 2 4 によって送信すべきトランザクションログの量を低減させ、また、第二のノード 1 B において反映すべきトランザクションログの量を低減出来る。但し、送信される特定L S N は、第二のノード 1 B において反映すべきトランザクションログを判別するための、所定の時点におけるトランザクションログを示すものであればよく、最新のトランザクションログを示すL S N に限定されない。

[0036] ログ送信部 2 4 は、少なくとも、特定L S N が示す所定の時点より後のトランザクションログおよびL S N を、互いに関連づけて第二のノード 1 B に送信する。トランザクションログの送信開始後、ログ送信部 2 4 は、新たに発生したトランザクションログについても、順次第二のノード 1 B に対して送信する。

[0037] データ送信部 2 5 は、要求受信部 2 2 によって受信されたデータ送信要求に応じて、データベースによって管理されているページのうち、要求に係るページであって、少なくとも特定L S N に係る所定の時点までの内容が反映されたページを、第二のノード 1 B 宛に送信する。本実施形態では、データ送信部 2 5 は、データを、トランザクションログに基づく更新命令の順序が互いに依存関係にあるレコードが同一の管理単位に入るように区切られた所定の管理単位（本実施形態では、ページ）毎に送信する。

[0038] なお、本実施形態では、データ送信部 2 5 は、第一のノード 1 A のメモリに展開されてデータの検索または処理に供されている状態のデータ、即ちキャッシュを、第二のノード 1 B 宛に送信する。送信の際、送信対象のページのキャッシュが無い場合には、第一のノード 1 A は、対象ページをストレージから読み出してキャッシュとする。これは、本実施形態に係るデータベ

システムでは、各ノード1において発生したデータの更新は即座にキャッシュに反映され、キャッシュがデータの最新の状態を表しているために、第二のノード1B宛に送信されるデータとしてキャッシュを採用することで、送信されるデータを、最新のトランザクションログを示すLSNである特定LSNに係る所定の時点までの内容が反映されたページとすることが出来るからである。

[0039] 但し、送信されるデータは、少なくとも特定LSNに係る所定の時点までの内容が反映されたものであればよく、本実施形態において採用されたキャッシュに限定されない。本開示が適用されるデータベースシステムが、ストレージにあるデータに、少なくとも特定LSNに係る所定の時点までの内容が反映されていることが保証できるものである場合には、ストレージから読み出されたデータがそのまま送信されてもよい。

[0040] また、本実施形態に係る第二のノード1Bは、CPU11Bが、RAM12Bに展開された各種プログラムを解釈および実行して、ノード1Bに備えられた各種ハードウェアを制御することで、要求送信部31、データ受信部32、実行部33、マップ生成部34、検索要求受付部35、検索応答部36および展開部37を備えるコンピューターとして機能する。本実施形態では、第二のノード1Bの機能が汎用のCPU11Bによって実行される例について説明しているが、これらの機能は、その一部または全部が、1または複数の専用のプロセッサによって実現されてもよい。

[0041] 要求送信部31は、第一のノード1Aに対して、データベース中のデータを指定してデータの送信要求を送信する。本実施形態では、データの指定は、ページ単位で行われる。

[0042] データ受信部32は、第一のノード1Aから、トランザクションログ、該トランザクションログのLSN、ページ毎の特定LSNおよびデータを受信する。なお、先述の通り、本実施形態では、データの送受信は、ページ単位で行われる。

[0043] 実行部33は、データの送信要求に応じて第一のノード1Aから送信され

、データ受信部 32 によって受信されたデータが、該第二のノード 1B のメモリに展開されてデータの検索または処理に供される状態（キャッシュ）となった場合に、受信されたトランザクションログのうち、少なくとも特定 LSN に係るトランザクションログよりも新しいトランザクションログに含まれる更新命令を、メモリに展開されたデータに対して実行する。ここで、トランザクションログに含まれる更新命令は、LSN によって把握される順序に従って実行される。

[0044] マップ生成部 34 は、受信されたトランザクションログのうち、少なくとも特定 LSN が示すトランザクションログよりも新しいトランザクションログに基づいて、トランザクションログに含まれる更新命令の対象となるデータが収容されているページと当該更新命令との関係を示すマップを生成する。

[0045] なお、上述の通り本実施形態では、データはページをもって管理されるため、マップ生成部 34 は、更新命令の対象となるデータが収容されたページと更新命令との関係を示すマップを生成するが、マップ生成手段は、その他の管理単位（例えば、テーブル等）と更新命令との関係を示すマップを生成してもよい。また、マップ生成手段は、このような管理単位を用いることなく、更新命令の対象となるレコードと更新命令との関係を示すマップを生成してもよい。

[0046] 検索要求受付部 35 は、ユーザー端末 9 からの検索要求（クエリ）を受け付ける。

[0047] 検索応答部 36 は、第一のノード 1A から得られ、トランザクションログに基づく更新命令が反映されたデータに基づいて、検索要求（クエリ）への応答をユーザー端末 9 に対して返す。

[0048] 展開部 37 は、第一のノード 1A から受信したデータをメモリに展開して、データの検索または処理に供する。本実施形態において、展開部 37 は、クエリや、後述する更新実行処理における求めに応じて、第一のノード 1A から受信されたデータをメモリに読み出すことでキャッシュとする。

[0049] <<処理の流れ>>

次に、本実施形態に係る処理の詳細を説明する。なお、本実施形態において説明される処理の具体的な内容および順序等は、実施する上での一例である。具体的な処理内容および順序等は、実施の形態に応じて適宜選択されてよい。

[0050] 図3は、本実施形態において、第一のノード1Aおよび第二のノード1Bによって実行される管理情報送受信処理の流れを示すフローチャートである。本フローチャートに示された処理は、第二のノード1Bにおいて、データベースの構築開始の指示が受け付けられたことを契機として開始される。

[0051] ステップS101およびステップS102では、管理情報要求が送受信される。要求送信部31は、第一のノード1Aに対して、データベースの管理情報（システムカタログ等）を指定してデータの送信要求を送信する（ステップS101）。ここで、管理情報は、データベース中のテーブルやレコードの位置を特定可能な情報を含んでいる。本実施形態に係るデータベースはページ単位で管理されているため、管理情報は、データベースのテーブルとページとの関係を特定可能な情報を含む。また、本実施形態では、管理情報の読出／書込および送受信についても、他のテーブルと同様ページ単位で行われる。但し、管理情報がテーブル単位等その他の単位に従って要求されてもよいことは、先述したデータの読出／書込および送受信と同様である。要求受信部22は、第二のノード1Bによって送信された要求を受信する（ステップS102）。その後、処理はステップS103へ進む。

[0052] ステップS103およびステップS104では、データの送信要求に対する応答（Acknowledgement。以下、「ACK」と称する）が送受信される。このACKは、管理情報の特定LSNを含む。先述の通り、LSNとは、トランザクションログに含まれる命令（更新命令）の順序を把握可能な識別情報であり、特定LSNとは、ログ保持部21によって保持されているLSNのうち、所定の時点におけるトランザクションログ（本実施形態では、最新のトランザクションログ）を示すLSNである。即ち、特定

識別情報送信部 23 は、ログ保持部 21 によって保持されている L S N のうち、要求された管理情報に係る特定 L S N を含む A C K を、第二のノード 1 B に対して送信する（ステップ S 103）。データ受信部 32 は、第一のノード 1 A から、特定 L S N を含む A C K を受信する（ステップ S 104）。その後、処理はステップ S 105 へ進む。

[0053] ステップ S 105 およびステップ S 106 では、トランザクションログの送受信が開始される。ログ送信部 24 は、所定の時点より後のトランザクションログおよび L S N を、互いに関連づけて第二のノード 1 B に送信する（ステップ S 105）。以降、第一のノード 1 A において生成されたトランザクションログは、順次、継続的に第二のノード 1 B に対して送信される。ここで、「所定の時点」とは、ステップ S 103 で送信された特定 L S N が示すトランザクションログが生成された時点である。データ受信部 32 は、第一のノード 1 A によって送信された、トランザクションログおよび該トランザクションログに係る L S N の受信を開始する（ステップ S 106）。

[0054] ログの送受信開始後、第一のノード 1 A において新たに発生したトランザクションログは、発生次第、第二のノード 1 B に対して送信され、第二のノード 1 B によって受信される。即ち、第二のノード 1 B は、ログの送信開始後、所定の時点以降のトランザクションログを全て受信することとなる。その後、処理はステップ S 107 へ進む。

[0055] なお、先述の通り、処理順序は本フローチャートに示した例に限定されない。管理情報送受信処理では、対象のページに係るトランザクションログのうち、特定 L S N が示す時点以降のトランザクションログが全て第二のノード 1 B によって受信されればよい。このため、例えば、ステップ S 103 およびステップ S 104 の送受信処理と、ステップ S 105 およびステップ S 106 の送受信処理との処理順序は、入れ替わってもよい。

[0056] ステップ S 107 では、管理情報について、マップ生成処理が開始される。マップ生成部 34 は、管理情報に関するページを、マップ生成処理の対象とする。即ち、本ステップ以降、管理情報に対して実行される更新命令と、

当該更新命令の実行対象（対象ページ、対象テーブルまたは対象レコード等）との組合せがマッピングされたマップの生成が開始される。マップ生成処理は、本フローチャートに示された処理と並行して実行される。マップ生成処理の詳細は、図5を用いて説明する。その後、処理はステップS108へ進む。

[0057] ステップS108およびステップS109では、データベースの管理情報が送受信される。データ送信部25は、ステップS102で受信された要求を受けて、データベースの管理情報（システムカタログ等）を、第二のノード1Bに対して送信する（ステップS108）。この際、データ送信部25は、送信対象のデータのキャッシュを、第二のノード1B宛に送信する。送信の際、送信対象のデータのキャッシュが無い場合には、第一のノード1Aは、対象データをストレージから読み出してキャッシュとし、少なくとも特定LSNに係る所定の時点までの内容が反映されたデータとしてから、第二のノード1B宛に送信する。第二のノード1Bのデータ受信部32は、第一のノード1Aから、管理情報を受信する（ステップS109）。その後、処理はステップS110へ進む。

[0058] なお、本フローチャートでは、管理情報の送信（ステップS108）が、マップの生成開始（ステップS107）よりも後に記載されているが、管理情報は、特定LSNに係る所定の時点以降のものが送信されればよく、送信のタイミングは本フローチャートに示された例に限定されない。

[0059] ステップS110では、更新実行処理が開始される。実行部33は、管理情報に関するページを、更新実行処理の対象とする。即ち、本ステップ以降、マップ生成処理においてマップに記録された、管理情報に未反映の更新命令が実行される。更新実行処理は、本フローチャートに示された処理と並行して実行される。更新実行処理の詳細は、図6を用いて説明する。その後、処理はステップS111へ進む。

[0060] ステップS111では、検索要求（クエリ）の受付が開始される。第二のノード1Bは、検索要求対応処理を開始させることで、ユーザー端末9から

送信されるクエリの受付を開始する。以降、検索要求受付部35は、ユーザー端末9からのクエリを受け付け、クエリが受け付けられると、クエリに応じるために必要なページが、管理情報を参照することで取得される。検索要求対応処理は、本フローチャートに示された処理と並行して実行される。検索要求対応処理の詳細は、図7および図8を用いて説明する。その後、本フローチャートに示された処理は終了する。

[0061] なお、本実施形態では、管理情報の複製についても、データベースの他のテーブル同様、マップ生成処理、更新実行処理および検索要求対応処理を用いて複製する例について説明したが、管理情報の複製には、その他の方法が採用されてもよい。管理情報の複製には、例えば、チェックポイントを作成してチェックポイントのスナップショットを複製する従来の方法等を採用することが出来る。

[0062] 図4は、本実施形態において、第一のノード1Aおよび第二のノード1Bによって実行されるデータベース送受信処理の流れを示すフローチャートである。本フローチャートに示された処理は、第二のノード1Bが、第一のノード1Aからデータベースの管理情報を受信することで、第一のノード1Aによって管理されているデータベースのページ構成を把握可能となったことを契機として開始される。

[0063] ステップS201およびステップS202では、データの送信要求が送受信される。要求送信部31は、管理情報のキャッシュを参照して、第一のノード1Aに対して、データベース中のページを指定してデータの送信要求を送信する(ステップS201)。ページの指定には、例えば、ページ番号が用いられてよい。データの送信要求は、管理情報から目的のデータを含むページが分かる場合には、ページを指定したものであってもよいし、データベース中の具体的なレコードを指定したものであってもよい。本実施形態において、データはページ単位で要求されるが、テーブル単位等その他の単位に従って要求されてもよいことは先述した通りである。要求受信部22は、第二のノード1Bから、第一のノード1Aによって管理されているデータの送

信要求を受信する（ステップS202）。その後、処理はステップS203へ進む。

[0064] ステップS203およびステップS204では、データの送信要求に対するACKが送受信される。このACKは、ページ番号および当該ページの特定LSNを含む。即ち、特定識別情報送信部23は、ログ保持部21によって保持されているLSNのうち、要求されたページに係る特定LSNを含むACKを、第二のノード1Bに対して送信する（ステップS203）。データ受信部32は、第一のノード1Aから、特定LSNを含むACKを受信する（ステップS204）。その後、処理はステップS205へ進む。

[0065] ステップS205では、要求に係るページ等について、マップ生成処理が開始される。マップ生成部34は、ステップS201において要求したページを、マップ生成処理の対象とする。即ち、本ステップ以降、要求されたページに対して実行される更新命令と、当該更新命令の実行対象（対象ページ、対象テーブルまたは対象レコード等）との組合せがマッピングされたマップの生成が開始される。ここで開始されるマップ生成処理の内容は、ステップS107において管理情報に対して開始されたものと概略同様であり（図5を参照）、本フローチャートに示された処理と並行して処理される。その後、処理はステップS206へ進む。

[0066] なお、本実施形態では、管理情報やページ毎にマップの生成開始タイミングが異なる例について説明しているが（ステップS107やステップS205を参照）、マップの生成開始タイミングは、データベース全体において同時であってもよい。

[0067] ステップS206およびステップS207では、要求に係るページが送受信される。データ送信部25は、ステップS201で受信された要求に応じて、データベースによって管理されているページのうち要求されたものを、第二のノード1Bに送信する（ステップS206）。この際、データ送信部25は、送信対象のデータのキャッシュを、第二のノード1B宛に送信する。送信の際、送信対象のページのキャッシュが無い場合には、第一のノード

1 Aは、対象ページをストレージから読み出してキャッシュとし、少なくとも特定LSNに係る所定の時点までの内容が反映されたページとしてから、第二のノード1 B宛に送信する。第二のノード1 Bのデータ受信部3 2は、第一のノード1 Aから、データを受信する（ステップS 2 0 7）。データの送受信単位は、要求に応じたものであればよく、ページ単位、テーブル単位およびレコード単位の何れでもよい。その後、処理はステップS 2 0 8へ進む。

[0068] なお、本フローチャートでは、データの送信（ステップS 2 0 6）が、マップの生成開始（ステップS 2 0 5）よりも後に記載されているが、データは、特定LSNに係る所定の時点以降のものが送信されればよく、送信のタイミングは本フローチャートに示された例に限定されない。

[0069] ステップS 2 0 8では、更新実行処理が開始される。実行部3 3は、ステップS 2 0 7で受信されたデータを、更新実行処理の対象とする。即ち、本ステップ以降、マップ生成処理でマップに記録された、未実行の更新命令が実行される。その後、本フローチャートに示された処理は終了する。

[0070] なお、本実施形態において、第二のノード1 Bは、第一のノード1 Aのデータベースにある全てのデータを取得するため、図4に示されたデータベース送受信処理は、管理情報に示された全てのページについて取得が完了するまで、要求するページを順次変更しながら繰り返し実行される。即ち、要求送信部3 1は、ステップS 2 0 1からステップS 2 0 8の処理を繰り返すことで、管理情報を参照して把握される全てのページまたはテーブルについて、データの送信要求を送信する。但し、第二のノード1 Bが、第一のノード1 Aのデータベースにある全てのデータを取得する必要がある場合には、要求送信部3 1は、必要な一部のページやテーブルについてのみ、データの送信要求を行ってもよい。また、本実施形態では、複数のページを取得する場合に、ステップS 2 0 1からステップS 2 0 8の処理を繰り返す例について説明したが、複数のページを取得する場合に、ステップS 2 0 1において複数ページ分まとめて要求することとしてもよい。

- [0071] 本フローチャートに示された処理において第一のノード1 Aから受信されたページには、未反映のトランザクションログが存在する可能性がある。このため、マップ生成処理、更新実行処理および検索要求対応処理（詳細については後述する）が本フローチャートに示された処理と並行して行われることで、未反映のトランザクションログが反映される。これらの処理が開始されるタイミングは、上記説明した通りである。なお、要求に係るページやテーブルに関するトランザクションログは、ステップS 1 0 5以降順次受信されているトランザクションログに含まれている。
- [0072] なお、本実施形態では、第一のノード1 Aから受信されたデータは、一旦ストレージに保存される。但し、このような処理に代えて、展開部3 7は、第一のノード1 Aから受信したデータを、ストレージ（本実施形態では、補助記憶装置1 4 B）を介さずに直接メモリに展開して、データの検索または処理に供してもよい。受信データが直接メモリに展開された場合、当該受信データに係る更新命令は、後述するマップ生成処理に従って、即座に反映される。
- [0073] 次に、第二のノード1 Bにおいて実行されるデータベースの管理処理の流れを説明する。本実施形態において、第二のノード1 Bのデータベースは、マップ生成処理、更新実行処理および検索要求対応処理によって管理される。第二のノード1 Bは、これらの処理を実行するためのインスタンスを適宜複数立ち上げて並列実行することによって、マップの生成、マッピングされた更新命令の反映、および検索要求（クエリ）の処理の夫々について、並列処理することを可能としている。
- [0074] 図5は、本実施形態において、第二のノード1 Bによって実行されるマップ生成処理の流れを示すフローチャートである。本フローチャートに示された処理は、マップ生成処理の対象となったデータ（ステップS 1 0 7およびステップS 2 0 5を参照）に対して、繰り返し実行される。
- [0075] ステップS 3 0 1では、トランザクションログに含まれる命令が、LSNを参照して時系列順に参照される。第二のノード1 Bは、未反映のトランザ

クションログに含まれる命令を、時系列において最も古い時点のもの（LSNが最も小さいもの）から順に参照する。ここで参照されるトランザクションログは、ステップS105およびステップS106の処理において送受信が開始され、第二のノード1Bによって受信されたものである。その後、処理はステップS302へ進む。

[0076] ステップS302およびステップS303では、命令が更新命令であるか否かが判定され、更新命令以外の命令が実行される。第二のノード1Bは、ステップS301で参照された命令が、更新命令であるか否かを判定する（ステップS302）。参照された命令が更新命令以外の命令（例えば、コミット等の管理命令）であった場合、命令はそのまま実行される（ステップS303）。一方、参照された命令が更新命令であった場合、処理はステップS304へ進む。

[0077] ステップS304では、参照された命令が示すページが、マップ生成処理の対象であるか否かが判定される。第二のノード1Bは、対象となっているデータが、ステップS107やステップS205等において、マップ生成処理の対象に設定されているか否かを判定する。ここで、対象データが、マップ生成処理の対象に設定されていないと判定された場合、本フローチャートに示された処理は終了する。一方、対象データが、マップ生成処理の対象に設定されていると判定された場合、処理はステップS305へ進む。

[0078] ステップS305では、対象データ（ページ）のキャッシュが存在するか否かが判定される。第二のノード1Bは、参照した命令が更新命令である場合、当該命令を参照した時点でその対象がデータベース内に存在するような更新命令（例えば、DELETE命令）については、当該更新命令の対象のキャッシュが存在するか否かを判定する。一方、当該命令を参照した時点でその対象がデータベース内に存在しないような更新命令（例えば、INSERT命令）については、当該更新命令の対象が挿入される領域のデータのキャッシュが存在するか否かを判定することで、当該更新命令の対象のキャッシュが存在するか否かを判定する。なお、レコードがページ単位で管理され

ている場合、第二のノード1 Bは、更新命令の対象を含むページのキャッシュが存在するか否かを判定する。対象のページのキャッシュが存在する場合、処理はステップS 3 0 6へ進む。一方、対象のページのキャッシュが存在しない場合、処理はステップS 3 0 7へ進む。

[0079] ステップS 3 0 6では、トランザクションログに係る更新命令が実行される。実行部3 3は、受信されたデータのうち、展開部3 7によってメモリに展開されてキャッシュとなったデータに対して、トランザクションログに係る更新命令を実行する。その後、処理はステップS 3 0 8へ進む。

[0080] ステップS 3 0 7では、マップが生成される。トランザクションログに含まれる更新命令の対象となっているページのキャッシュが存在しない場合（ステップS 3 0 5で「NO」に進んだ場合）、マップ生成部3 4は、トランザクションログに含まれる更新命令の対象となるデータが収容されたページと更新命令との関係を示す情報を、マップに記録することで、更新命令の実行を待機させる。換言すれば、マップ生成部3 4は、更新命令の実行を待機させる際に、当該更新命令をマッピング（マップ生成）の対象とし、待機させた更新命令の内容を当該レコードに関連付けて記録するマップを生成する。但し、全ての更新命令をマップ生成の対象としてもよい。本実施形態では、マップには、待機させた更新命令の内容およびLSNのうちの少なくとも一方が、更新命令の対象であるレコードを含むページ（具体的にはページ番号）に関連付けられて記録される。その後、処理はステップS 3 0 8へ進む。

[0081] ステップS 3 0 8では、マップ生成処理を終了するか否かが判定される。マップ生成処理が終了されない場合、処理はステップS 3 0 1へ戻り、トランザクションログにおける時系列順（LSNの順）で次の命令が参照され、ステップS 3 0 4からステップS 3 0 7に示された処理の対象となる。通常、第一のノード1 Aにおいて新たに発生したトランザクションログを第二のノード1 Bに反映するために、マップ生成処理は、第二のノード1 Bが運用されている間、繰り返し実行される。即ち、本実施形態において、第二のノ

ード1 Bは、受信されたトランザクションログを、古いものから順に検査し、トランザクションログに含まれる更新命令の対象となっているページがオンメモリ（キャッシュ）である場合に、キャッシュに対してトランザクションログに応じた更新命令を実行する。一方、第二のノード1 Bが停止される等の理由でマップ生成処理が終了される場合、本フローチャートに示された処理は終了する。

[0082] マップ生成処理で生成されたマップは、対象ページがキャッシュとなった場合に、更新実行処理または検索要求対応処理において参照され、更新実行処理または検索要求対応処理において、当該マップに待機させられた更新命令が実行される。即ち、本実施形態に係るシステムによれば、第一のノードから受信されたデータに対するトランザクションログの反映を、対象データが実際に用いられる時点（換言すれば、キャッシュとなる時点）まで遅延させることが出来る。

[0083] ここで、更新命令の内容を記録したトランザクションログは、ノード1 BのRAM 1 2 Bまたは補助記憶装置1 4 Bに保持されている。更新実行処理または検索要求対応処理では、当該マップに記録されるLSNに基づいてRAM 1 2または記憶装置1 4に保持されているトランザクションログへのアクセスが行われ、実行する更新命令の内容が特定される。

[0084] なお、マップ生成処理では、参照した順番に応じて、更新命令の内容またはLSNがマップに記録される。これにより、後述する更新実行処理または検索要求対応処理では、待機させられた更新命令の実行順序を認識することができる。但し、仮に、更新命令の内容またはLSNが参照された順番に応じてマップに記録されなかったとしても、更新実行処理または検索要求対応処理では、待機させられた更新命令のLSNにより、当該更新命令の実行順序を認識することができる。

[0085] 図6は、本実施形態において、第二のノード1 Bによって実行される更新実行処理の流れを示すフローチャートである。本フローチャートに示された処理は、マップに待機している更新命令が存在する間、繰り返し実行される

- 。
- [0086] ステップS401では、待機している更新命令の有無が判定される。第二のノード1Bは、マップを参照し、待機している更新命令があるか否かを判定する。待機している更新命令がマップに記録されていない場合、本フローチャートに示された処理は終了する。一方、待機している更新命令がマップに記録されている場合、第二のノード1Bは、更新命令の対象となるデータが含まれる管理単位を、処理の対象として選択して、処理をステップS402へ進める。
- [0087] ステップS402では、更新対象が存在するか否かが判定される。第二のノード1Bは、ステップS401において選択された処理の対象が、第二のノード1Bに存在するか否か、換言すれば、処理の対象となるページが、既に第一のノード1Aから受信されているか否か、を判定する。更新対象のページが存在しない場合、処理はステップS401へ戻る。即ち、第二のノード1Bは、マップを参照して、第二のノード1Bに存在するページについて待機している更新命令を待ち受ける。更新対象のページが存在する場合、処理はステップS404へ進む。
- [0088] ステップS403では、更新対象ページに関する更新命令がマップから抽出される。第二のノード1Bは、ステップS402において存在すると判定された更新対象のページに対する更新命令を、マップから全て抽出する。このようにすることで、更新対象ページに関するマップ上の更新命令を全て実行し、更新命令が実行されたページをそのままキャッシュとして、データの検索または処理に供することが出来る。その後、処理はステップS404へ進む。
- [0089] ステップS404では、展開部37は、自身が処理する対象を、補助記憶装置14BからRAM12Bに読み出すことで、キャッシュとする。その後、処理はステップS405へ進む。
- [0090] ステップS405およびステップS406では、マップに待機している更新命令および抽出された更新命令が実行され、実行済みの更新命令がマップ

から削除される。実行部33は、マップを参照して、自身が選択した処理対象に関連付けられている更新命令を実行する（ステップS405）。即ち、実行部33は、ステップS401で発見された更新命令、およびステップS403で抽出された更新命令を、所定の管理単位であるページ毎に、マップに記録されている順に実行する。例えば、更新命令が削除命令である場合、実行部33は、対象のレコードに削除ポインタを付与する。そして、第二のノード1Bは、自身が選択した処理対象について待機している更新命令を実行した後に、実行した更新命令に関する記録をマップから削除する（ステップS406）。その後、処理はステップS407へ進む。

[0091] ステップS407では、更新実行処理を終了するか否かが判定される。更新実行処理が終了されない場合、処理はステップS401へ戻る。即ち、更新実行処理では、処理対象として選択したページに複数の更新命令が待機している場合、マップに記録されている順（LSN順）に、当該待機している更新命令が実行される。これにより、そのページに関して、ノード1とレプリケーション元のコンピューターとの間で、レプリケーションが実行されている最中におけるデータベースの整合性を保つことができる。一方、第二のノード1Bが停止される等の理由で更新実行処理が終了される場合、本フローチャートに示された処理は終了する。

[0092] なお、第二のノード1Bは、更新実行処理を実行するためのインスタンスを複数立ち上げてよく、このようなインスタンスを生成若しくは削除または起動若しくは停止することで、更新実行処理を実行するインスタンスの数を増減させてもよい。例えば、第二のノード1Bは、マップで記録されている待機させられている更新命令の管理単位の数（本実施形態では、ページ数）に応じて、起動するインスタンスの数を増減させてもよい。この場合は、更新実行処理の処理能力を要求に応じて可変にでき、無駄なインスタンスが存在することを避けることができるため、リソースを効率よく活用することが可能になる。

[0093] また、例えば、更新実行処理を実行するためのインスタンスを、予め一定

数起動させておいてもよい。この場合、処理の要求が生じた場合に改めてインスタンスを起動しなくてもよいため、生じた要求に迅速に対応することが可能になる。

[0094] 図7および図8は、本実施形態において、第二のノード1Bによって実行される検索要求対応処理の流れを示すフローチャートである。本フローチャートに示された処理は、第二のノード1Bが、ユーザー端末9等からクエリを受信したことを契機として開始される。

[0095] ステップS501では、クエリが受け付けられる。検索要求受付部35は、ユーザー端末9からのクエリを受け付ける。その後、処理はステップS502へ進む。

[0096] ステップS502では、管理情報のキャッシュが存在するか否かが判定される。検索応答部36は、メモリ上に、管理情報のキャッシュが存在するか否かを判定する。なお、本実施形態では、レコードはページ単位で管理されているため、検索応答部36は、検索対象での管理情報を含むページのキャッシュが存在するか否かを判定する。

[0097] 本実施形態では、管理情報を含むページがRAM12Bに読み出された段階で、当該ページに対して、待機させられた更新命令が実行される（図6および後述するステップS504からステップS507に示す処理を参照）。また、キャッシュが存在するページに対する更新命令は即座に実行される（図5を用いて説明したマップ生成処理を参照）。このため、本実施形態において、管理情報を含むページのキャッシュが既に存在する場合（ステップS502における判定結果が「YES」の場合）、当該ページに対する更新命令の実行が待機されていることはない。従って、管理情報を含むページのキャッシュが存在する場合、検索応答部36は、記憶領域からの読み出しを行わず、既にキャッシュが存在する管理情報を参照の対象とし、処理はステップS508へ進む。一方、管理情報のキャッシュが存在しない場合、処理はステップS503へ進む。

[0098] ステップS503では、管理情報がストレージから読み出される。本実施

形態では、データベースの構築開始の時点で管理情報送受信処理が行われるため（図3を参照）、クエリを受けた時点で管理情報がキャッシュに存在しない場合にも、第二のノード1Bは、ストレージに管理情報を有している。展開部37は、管理情報のキャッシュが存在しない場合、管理情報をストレージ（補助記憶装置14B）からRAM12Bに読み出す。その後、処理はステップS504へ進む。

[0099] ステップS504では、待機している更新命令の有無が判定される。第二のノード1Bは、マップを参照し、RAM12Bに読み出された管理情報のページについて待機している更新命令があるか否かを判定する。RAM12Bに読み出されたページについて待機している更新命令がマップに記録されていない場合、処理はステップS508へ進む。一方、RAM12Bに読み出したページについて待機している更新命令がマップに記録されている場合、処理はステップS505に進む。

[0100] ステップS505では、更新対象ページに関する更新命令がマップから抽出される。第二のノード1Bは、ステップS503において存在すると判定された更新対象のページ（ここでは、管理情報を含むページ）に対する更新命令を、マップから全て抽出する。このようにすることで、更新対象ページに関するマップ上の更新命令を全て実行し、更新命令が実行されたページをそのままキャッシュとして、データの検索または処理に供することが出来る。その後、処理はステップS506へ進む。

[0101] ステップS506およびステップS507では、マップに待機している更新命令および抽出された更新命令が実行され、実行済みの更新命令がマップから削除される。実行部33は、マップを参照して、RAM12Bに読み出した管理情報に関連づけられている更新命令を実行する（ステップS506）。即ち、本実施形態において、第二のノード1Bは、クエリに応じて（ステップS501）、データ受信部32によって受信された管理情報（ステップS207）をメモリに展開し（ステップS503）、マップに基づいて更新命令を管理情報に対して実行する。

- [0102] ここで、実行部33は、マップに記録されている順に、待機している更新命令を実行する。そして、第二のノード1Bは、RAM12Bに読み出した管理情報について待機している更新命令を全て実行した後に、実行した更新命令に関する記録をマップから削除する（ステップS507）。その後、処理はステップS508へ進む。
- [0103] ステップS508では、管理情報が参照されてページが特定される。検索応答部36は、受信されたクエリにより検索範囲に指定された記憶領域を特定するために、管理情報を参照する。本実施形態では、管理情報を参照することで、クエリを処理するために必要なデータを含むページが特定される。その後、処理はステップS511へ進む。
- [0104] ステップS511では、検索対象のデータ（ページ）のキャッシュが存在するか否かが判定される。検索応答部36は、受信したクエリの内容に適合するレコードを抽出するために、当該クエリにより検索範囲に指定された記憶領域にアクセスする。そして、検索応答部36は、メモリ上に、クエリに係るレコードのキャッシュが存在するか否かを判定する。なお、本実施形態では、レコードはページ単位で管理されているため、検索応答部36は、検索対象でのレコードを含むページのキャッシュが存在するか否かを判定する。
- [0105] 本実施形態では、ページがRAM12Bに読み出された段階で、当該ページに対して、待機させられた更新命令が実行される（図6および後述する図8のステップS515からステップS518に示す処理を参照）。また、キャッシュが存在するページに対する更新命令は即座に実行される（図5を用いて説明したマップ生成処理を参照）。このため、本実施形態において、検索範囲に含まれるページのキャッシュが既に存在する場合（ステップS511における判定結果が「YES」の場合）、当該ページに対する更新命令の実行が待機されていることはない。従って、対象のページのキャッシュが存在する場合、検索応答部36は、記憶領域からの読み出しを行わず、既にキャッシュが存在するページを検索処理の対象とし、処理はステップS519

へ進む。一方、対象のページのキャッシュが存在しない場合、処理はステップS 5 1 2へ進む。

[0106] ステップS 5 1 2では、検索対象のページが第二のノード1 Bのストレージ（データベース）に存在するか否かが判定される。上述の通り、本実施形態では、ページの送信要求は逐次的に行われるため（ステップS 2 0 1を参照）、クエリを受けた時点で、第二のノード1 Bがクエリに係るページを有していない場合がある。ページがストレージに存在しないと判定された場合、該当するページを第一のノード1 Aから取得するために、処理はステップS 5 1 4へ進む。一方、ページがストレージに存在すると判定された場合、処理はステップS 5 1 3へ進む。

[0107] ステップS 5 1 4では、図4に示されたデータベース送受信処理が実行される。ステップS 5 1 2においてページがストレージに存在しないと判定された場合、要求送信部3 1は、第一のノード1 Aに対して、データベース中の少なくともクエリに係るデータを指定してデータの送信要求を送信する。本実施形態では、図4に示されたデータベース送受信処理が繰り返されることで管理情報に示されたページが順番に取得されるが、ステップS 5 0 1で検索要求を受けて検索対象となったページについては、この順番を無視して、優先的にデータ送信要求される。即ち、本実施形態によれば、検索要求を受けて検索対象となったページを、優先的にリクエスト（データ送信要求）の対象とすることが出来る。その後、処理はステップS 5 1 1へ戻る。

[0108] そして、第一のノード1 Aから対象ページが取得されるまで、ステップS 5 1 1からステップS 5 1 4に示された処理が繰り返され、ステップS 5 0 1で受け付けられたクエリへの対応は、対象ページが第一のノード1 Aから取得されるまで待機される。

[0109] ステップS 5 1 3では、検索対象のページがストレージから読み出される。展開部3 7は、検索対象のデータ（ページ）のキャッシュが存在しないが、第二のノード1 Bのストレージに存在する場合、そのページをストレージ（補助記憶装置1 4 B）からRAM 1 2 Bに読み出し、検索処理の対象とす

る。その後、処理はステップS 5 1 5へ進む。

[0110] なお、展開部3 7は、検索範囲に含まれるページを補助記憶装置1 4 BからRAM 1 2 Bに読み出す場合に、例えば、1つ1つ対象のページを読み出してもよいし、複数のページを一度に読み出してもよい。この場合に、実行部3 3は、1または複数のページを読み出す毎に、読み出した1または複数のページに対して、後述するステップS 5 1 5以降の処理を実行する。

[0111] ステップS 5 1 5では、待機している更新命令の有無が判定される。第二のノード1 Bは、マップを参照し、RAM 1 2 Bに読み出されたページについて待機している更新命令があるか否かを判定する。RAM 1 2 Bに読み出されたページについて待機している更新命令がマップに記録されていない場合、処理はステップS 5 1 9へ進む。一方、RAM 1 2 Bに読み出したページについて待機している更新命令がマップに記録されている場合、処理はステップS 5 1 6に進む。

[0112] ステップS 5 1 6では、更新対象ページに関する更新命令がマップから抽出される。第二のノード1 Bは、ステップS 5 1 5において存在すると判定された更新対象のページに対する更新命令を、マップから全て抽出する。このようにすることで、更新対象ページに関するマップ上の更新命令を全て実行し、更新命令が実行されたページをそのままキャッシュとして、データの検索または処理に供することが出来る。その後、処理はステップS 5 1 7へ進む。

[0113] ステップS 5 1 7およびステップS 5 1 8では、マップに待機している更新命令および抽出された更新命令が実行され、実行済みの更新命令がマップから削除される。実行部3 3は、マップを参照して、RAM 1 2 Bに読み出したページに係る更新命令を実行する（ステップS 5 1 7）。即ち、本実施形態において、第二のノード1 Bは、クエリに応じて（ステップS 5 0 1）、データ受信部3 2によって受信されたデータ（ステップS 2 0 7）をメモリに展開し（ステップS 5 1 3）、マップに基づいて更新命令を該データに対して実行する。

- [0114] ここで、実行部33は、マップに記録されている順（LSN順）に、待機している更新命令を実行する。そして、第二のノード1Bは、RAM12Bに読み出したページについて待機している更新命令を全て実行した後に、実行した更新命令に関する記録をマップから削除する（ステップS518）。その後、処理はステップS519へ進む。
- [0115] このように、検索要求対応処理では、更新実行処理と同様、処理対象として選択したページに複数の更新命令が待機している場合、マップに記録されている順（LSN順）に、当該待機している更新命令が実行される。これにより、そのページに関して、第一のノード1Aと第二のノード1Bとの間で、データベースの整合性を保つことができる。
- [0116] ステップS519では、検索応答部36は、第一のノード1Aから得られ、更新命令が反映されたページに対して、受信したクエリに応じた検索処理を実行する。その後、処理はステップS520へ進む。
- [0117] ステップS520では、検索要求（クエリ）に対する応答が送信される。検索応答部36は、クエリに対する応答として、検索処理の結果をユーザー端末9に送信する。その後、本フローチャートに示された処理は終了する。
- [0118] なお、第二のノード1Bは、検索要求対応処理を実行するためのインスタンスを複数立ち上げてよく、このようなインスタンスを生成若しくは削除または起動若しくは停止することで、検索要求対応処理を実行するインスタンスの数を増減させてもよい。例えば、第二のノード1Bは、ユーザー端末9から受け付けたクエリの数に応じて起動するインスタンスの数を増減させてもよい。この場合、検索要求対応処理の処理能力を要求に応じて可変にでき、無駄なインスタンスが存在することを避けることができるため、リソースを効率よく活用することが可能になる。
- [0119] また、検索要求対応処理は、複数のインスタンスに分けて実行されてもよい。例えば、検索要求対応処理のうち、ステップS515からステップS518の処理を、他の処理とは異なるインスタンスで実行することとしてもよい。この場合、例えば、ステップS515からステップS518以外の処理

を実行するための第一のインスタンスにおいてページ読み出し（ステップS 5 1 3）が行われた後に、ページ更新のため、処理がステップS 5 1 5からステップS 5 1 8の処理を実行するための第二のインスタンスに引き継がれる。そして、第二のインスタンスにおけるステップS 5 1 8の処理が完了すると、処理は第一のインスタンスに戻り、ステップS 5 1 9の検索処理が実行される。

[0120] <<第一の実施形態に係るシステムの効果>>

本実施形態に係るシステムによれば、データベース全体を一括コピーすること無く、また、チェックポイントを作成すること無く、データベースの部分（ページ等の所定の管理単位）毎にデータを第一のノード1 Aから第二のノード1 Bに送信出来る。また、第二のノード1 Bは、トランザクションログに基づく更新命令同士の依存関係を気にすること無く、データベースの部分毎に並列にトランザクションログを反映することが出来る。また、マップを生成することで、トランザクションログに基づく更新命令の実行を遅延させることが出来、受信されたデータがメモリに展開された場合に、マップに基づいて、トランザクションログに基づく更新命令を、所定の管理単位毎に並列に実行することが出来る。

[0121] 結果として、本実施形態に係るシステムによれば、データベース全体の複製完了を待つこと無く、受信したデータを第二のノード1 Bにおけるクエリに供することが出来る。

[0122] また、本実施形態に係るシステムでは、第一のノード1 Aから第二のノード1 Bへのデータ送信を、データの送信要求に応じたものとする事で、データ取得を逐次的にする等、データ取得のタイミングを調整することが出来る。更に、第二のノード1 Bにおいて受け付けられた検索要求に応じてデータの送信要求を行ったり、トランザクションログの反映を行ったりすることで、検索対象となったデータを優先的に取得したり、優先的にトランザクションログを反映したりすることが可能となる。即ち、本実施形態に係るシステムによれば、第二のノード1 Bが全てのデータを受信していない状態で、

ユーザー端末からの検索要求を受け付けることが可能となる。

[0123] <第二の実施形態>

次に、第二の実施形態について説明する。第二の実施形態に係るシステムは、第二のノードにデータベース全体を保持することなく、第二のノードがクエリに応答可能とするために、第一の実施形態に係るシステムに構成を追加したものである。第一の実施形態では、全てのページについてのデータベース送受信処理が完了した場合、第二のノードにデータベース全体が保持されることとなるが、第二の実施形態では、データベース中のデータは、保持対象と判定されたものを除いて、第二のノードに恒久的には保持されない。第二の実施形態に係るシステムのうち、上述した第一の実施形態のシステムと共通する構成には、同一の符号を付し、説明を省略する。

[0124] <<システムの構成>>

第二の実施形態に係るシステムのハードウェア構成は、第一の実施形態に係るシステムのハードウェア構成と概略同様であるため、説明を省略する（図1を参照）。但し、第二の実施形態では、第二のノードの機能構成が第一の実施形態とは異なるため、第二のノードに符号「1 B'」を付して説明する。

[0125] 図9は、本実施形態に係る第一のノード1 Aおよび第二のノード1 B'の機能構成の概略を示す図である。本実施形態に係る第一のノード1 Aの機能構成は、第一の実施形態と概略同様であるため説明を省略する。本実施形態に係る第二のノード1 B'は、CPU 1 1 Bが、RAM 1 2 Bに展開された各種プログラムを解釈および実行して、ノード1 B'に備えられた各種ハードウェアを制御することで、要求送信部3 1、データ受信部3 2、実行部3 3、マップ生成部3 4、検索要求受付部3 5、検索応答部3 6、展開部3 7、判定部3 8、記録部3 9および削除部4 0を備えるコンピューターとして機能する。本実施形態では、第二のノード1 B'の機能が汎用のCPU 1 1 Bによって実行される例について説明しているが、これらの機能は、その一部または全部が、1または複数の専用のプロセッサによって実現されてもよ

い。

[0126] 要求送信部 31、データ受信部 32、実行部 33、マップ生成部 34、検索要求受付部 35、検索応答部 36 および展開部 37 については、第一の実施形態と概略同様であるため、説明を省略する。

[0127] 判定部 38 は、検索要求に応じて第一のノード 1A から得られ、トランザクションに基づく命令が実行されたデータを、第二のノード 1B' における保持対象とするか否かを判定する。

[0128] 記録部 39 は、判定部 38 によって保持対象と判定されたデータを、不揮発性の記憶装置である補助記憶装置 14B に記録する。

[0129] 削除部 40 は、検索要求に応じて第一のノード 1A から得られ、トランザクションに基づく命令が実行されたデータを、検索応答部 36 による応答に用いた後、所定の条件に従って自動的に第二のノードから削除する。

[0130] <<処理の流れ>>

次に、本実施形態に係る処理の詳細を説明する。なお、本実施形態において説明される処理の具体的な内容および順序等は、実施する上での一例である。具体的な処理内容および順序等は、実施の形態に応じて適宜選択されてよい。

[0131] なお、第一の実施形態において説明したデータベース送受信処理（図 4 を参照）、マップ生成処理（図 5 を参照）および更新実行処理（図 6 を参照）は、第二の実施形態でも実行される。これらの処理の詳細は、第一の実施形態において説明した通りであるため、説明を省略する。

[0132] 図 10 は、本実施形態において、第一のノード 1A および第二のノード 1B' によって実行されるトランザクションログ送受信開始処理の流れを示すフローチャートである。本フローチャートに示された処理は、第一の実施形態において説明した管理情報送受信処理（図 3 を参照）に代えて実行される処理であり、データベースの構築開始の指示が受け付けられたことを契機として開始される。

[0133] ステップ S801 およびステップ S802 では、トランザクションログ要

求が送受信される。要求送信部31は、第一のノード1Aに対して、トランザクションログの送信要求を送信する（ステップS801）。要求受信部22は、第二のノード1B'によって送信された要求を受信する（ステップS802）。その後、処理はステップS803へ進む。

[0134] ステップS803およびステップS804では、トランザクションログの送受信が開始される。ログ送信部24は、以降、第一のノード1Aにおいて新たに発生したトランザクションログおよびLSNを、発生次第、互いに関連づけて第二のノード1B'に順次送信する（ステップS803）。第二のノード1B'のデータ受信部32は、第一のノード1Aによって送信された、トランザクションログおよび該トランザクションログに係るLSNの受信を開始する（ステップS804）。即ち、第二のノード1B'は、ログの送信開始後、所定の時点以降のトランザクションログを全て受信することとなる。その後、処理はステップS805へ進む。

[0135] ステップS805では、検索要求（クエリ）の受付が開始される。第二のノード1B'は、後述するオンデマンド処理を開始させることで、ユーザー端末9から送信されるクエリの受付を開始する。以降、検索要求受付部35は、ユーザー端末9からのクエリを受け付け、クエリが受け付けられると、クエリに応じるために必要なページ（管理情報を含む）が取得される。その後、本フローチャートに示された処理は終了する。

[0136] 即ち、第二の実施形態では、第一の実施形態と異なり、データベースの構築開始の時点で管理情報送受信処理が行われない。

[0137] また、図4に示すデータベース送受信処理については、第一の実施形態において「第一のノード1Aによって管理されているデータベースのページ構成を把握可能となったことを契機として開始される」と説明したが、第二の実施形態では、第一の実施形態とは異なり、後述するオンデマンド処理の中で呼び出されるタイミングで実行される（後述するステップS605およびステップS514を参照）。

[0138] 図11は、本実施形態において、第二のノード1B'によって実行される

オンデマンド処理の流れを示すフローチャートである。本フローチャートに示された処理は、第二のノード1 B' が、ユーザー端末9等からクエリを受信したことを契機として開始される。即ち、本実施形態では、第一の実施形態において説明した検索要求対応処理（図7および図8を参照）に代えて、オンデマンド処理が実行される。

[0139] 但し、第二の実施形態では、上述の通り、データベースの構築開始の時点で管理情報送受信処理が行われていない。このため、検索要求が受け付けられた時点で、第二のノード1 B' が管理情報を有していない可能性がある。そして、第二のノード1 B' が管理情報を有していない場合、データの送信要求においてクエリに係るページを指定することが出来ない。このため、本実施形態に係るオンデマンド処理では、検索要求が受け付けられると、はじめに、第二のノード1 B' が管理情報を有しているか否かが確認される。そして、第二のノード1 B' が管理情報を有していない場合、これを第一のノード1 Aから取得してから、データベース送受信処理が実行される。以下、本実施形態に係るオンデマンド処理の詳細を、図11を参照しながら説明する。

[0140] ステップS601では、クエリが受け付けられる。検索要求受付部35は、ユーザー端末9からのクエリを受け付ける。その後、処理はステップS602へ進む。

[0141] ステップS602では、管理情報のキャッシュが存在するか否かが判定される。検索応答部36は、メモリ上に、管理情報のキャッシュが存在するか否かを判定する。なお、本実施形態では、レコードはページ単位で管理されているため、検索応答部36は、検索対象での管理情報を含むページのキャッシュが存在するか否かを判定する。

[0142] 本実施形態では、管理情報を含むページがRAM12Bに読み出された段階で、当該ページに対して、待機させられた更新命令が実行される（図6および後述するステップS606からステップS609に示す処理を参照）。また、キャッシュが存在するページに対する更新命令は即座に実行される（

図5を用いて説明したマップ生成処理を参照)。このため、本実施形態において、管理情報を含むページのキャッシュが既に存在する場合（ステップS602における判定結果が「YES」の場合）、当該ページに対する更新命令の実行が待機されていることはない。従って、管理情報を含むページのキャッシュが存在する場合、検索応答部36は、記憶領域からの読み出しを行わず、既にキャッシュが存在する管理情報を参照の対象とし、処理はステップS610へ進む。一方、管理情報のキャッシュが存在しない場合、処理はステップS603へ進む。

[0143] ステップS603では、管理情報を含むページが第二のノード1B'のストレージ（データベース）に存在するか否かが判定される。上述の通り、本実施形態では、データベースの構築開始の時点で管理情報送受信処理が行われていない可能性があるため、クエリを受けた時点で、第二のノード1B'がクエリに係るページを特定するための管理情報を有していない場合がある。管理情報がストレージに存在しないと判定された場合、管理情報を第一のノード1Aから取得するために、処理はステップS605へ進む。一方、ページがストレージに存在すると判定された場合、処理はステップS604へ進む。

[0144] ステップS605では、図4に示されたデータベース送受信処理が実行される。ステップS603において管理情報がストレージに存在しないと判定された場合、要求送信部31は、クエリに応じて、第一のノード1Aに対して、管理情報に係るデータを指定してデータの送信要求を送信する。上述の通り、本実施形態では、管理情報の送受信についても、他のテーブルと同様ページ単位で行われるため、図4に示されたデータベース送受信処理によって管理情報を取得することが出来る。その後、処理はステップS602へ戻る。そして、第一のノード1Aから管理情報が取得されるまで、ステップS602からステップS605に示された処理が繰り返される。

[0145] ステップS604では、管理情報がストレージから読み出される。展開部37は、管理情報のキャッシュが存在しないが、第二のノード1B'のスト

レージに存在する場合、管理情報をストレージ（補助記憶装置14B）からRAM12Bに読み出す。その後、処理はステップS606へ進む。

[0146] ステップS606では、待機している更新命令の有無が判定される。第二のノード1B'は、マップを参照し、RAM12Bに読み出された管理情報のページについて待機している更新命令があるか否かを判定する。RAM12Bに読み出されたページについて待機している更新命令がマップに記録されていない場合、処理はステップS610へ進む。一方、RAM12Bに読み出したページについて待機している更新命令がマップに記録されている場合、処理はステップS607に進む。

[0147] ステップS607では、更新対象ページに関する更新命令がマップから抽出される。第二のノード1B'は、ステップS603において存在すると判定された更新対象のページ（ここでは、管理情報を含むページ）に対する更新命令を、マップから全て抽出する。このようにすることで、更新対象ページに関するマップ上の更新命令を全て実行し、更新命令が実行されたページをそのままキャッシュとして、データの検索または処理に供することが出来る。その後、処理はステップS608へ進む。

[0148] ステップS608およびステップS609では、マップに待機している更新命令および抽出された更新命令が実行され、実行済みの更新命令がマップから削除される。実行部33は、マップを参照して、RAM12Bに読み出した管理情報に係る更新命令を実行する（ステップS608）。即ち、本実施形態において、第二のノード1B'は、クエリに応じて（ステップS601）、データ受信部32によって受信された管理情報（ステップS207）をメモリに展開し（ステップS604）、マップに基づいて更新命令を管理情報に対して実行する。

[0149] ここで、実行部33は、マップに記録されている順に、待機している更新命令を実行する。そして、第二のノード1B'は、RAM12Bに読み出した管理情報について待機している更新命令を全て実行した後に、実行した更新命令に関する記録をマップから削除する（ステップS609）。その後、

処理はステップS 6 1 0へ進む。

[0150] ステップS 6 1 0では、管理情報が参照されてページが特定される。検索応答部3 6は、受信されたクエリにより検索範囲に指定された記憶領域を特定するために、管理情報を参照する。本実施形態では、管理情報を参照することで、クエリを処理するために必要なデータを含むページが特定される。その後、処理は図8を用いて説明した検索要求対応処理のステップS 5 1 1からステップS 5 2 0に示す処理の内容と概略同様の処理へ進む。

[0151] 即ち、第一のノード1 Aから対象ページが取得されるまで、ステップS 5 1 1からステップS 5 1 4に示された処理が繰り返され、ステップS 6 0 1で受け付けられたクエリへの対応は、対象ページが第一のノード1 Aから取得されるまで待機される。本実施形態では、検索要求を受けて検索対象となることで、はじめてリクエスト（データ送信要求）の対象となる。その後、ステップS 5 1 9およびステップS 5 2 0に示された処理が実行され、クエリに対する応答が完了すると、本フローチャートに示された処理は終了する。

[0152] 図1 2は、本実施形態において、第二のノード1 B' によって実行される削除処理の流れを示すフローチャートである。本フローチャートに示された処理は、第二のノード1 B' において、メモリに展開されてキャッシュとなっているデータの単位（本実施形態では、ページ単位）毎に、定期的に行われる。

[0153] ステップS 7 0 1では、データが削除条件を満たしたか否かが判定される。削除部4 0は、第二のノード1 B' 上にあるキャッシュのうち、判定対象のページに係るキャッシュについて、所定の削除条件を満たしたか否かを判定する。ここで用いられる削除条件は、例えば、データ処理に関する時点（第一のノード1 Aから取得された時点や、クエリ応答や更新等のために最後に参照された時点等）から所定の時間が経過したこと、等である。但し、削除条件は、本実施形態における例示に限定されない。削除条件を満たさないと判定された場合、判定対象のページに係るキャッシュは削除されず、本フロ

ーチャートに示された処理は終了する。一方、削除条件を満たしたと判定された場合、処理はステップS 7 0 2へ進む。

[0154] ステップS 7 0 2では、データが保持の対象であるか否かが判定される。判定部38は、第二のノード1 B' によって保持されているデータのうち、判定対象のページについて、第二のノード1 B' における保持対象とするか否かを判定する。ここで、判定のための条件は、例えば、当該データがクエリにおいて参照される頻度等に基づいて決定されてよい。例えば、参照される頻度の高いデータを保持の対象とし、頻度の低いデータを保持の対象としないことで、クエリへの応答速度を犠牲にすること無く、第二のノード1 B' のストレージやメモリを節約することが出来る。また、例えば、予め管理者等によって指定されたテーブルやページを保持の対象として設定しておくことも可能である。但し、保持対象とするか否かの判定条件は、本実施形態における例示に限定されない。判定条件は、実施の形態に応じて適宜採用することが出来る。判定対象のページが、保持の対象となるための条件を満たさない場合、処理はステップS 7 0 4へ進む。一方、判定対象のページが、保持の対象となるための条件を満たしている場合、処理はステップS 7 0 3へ進む。

[0155] ステップS 7 0 3では、データがストレージに記録される。記録部39は、判定部38によって保持対象と判定されたデータ（本実施形態では、ページ単位で判定される）を、ストレージ（本実施形態では、補助記憶装置14 B）に記録する。その後、処理はステップS 7 0 5へ進む。

[0156] ステップS 7 0 4では、削除条件を満たしたデータがストレージから削除される。削除部40は、判定部38によって保持対象ではないと判定されたデータ（本実施形態では、ページ単位で判定される）が、第二のノード1 B' のストレージに記録されている場合に、これをストレージ（本実施形態では、補助記憶装置14 B）から削除する。このようにすることで、受信されてキャッシュとなる前に一旦ストレージに記録されたデータについても、ストレージから削除することが出来る。その後、処理はステップS 7 0 5へ進

む。

[0157] ステップS705では、削除条件を満たしたデータのキャッシュが削除される。削除部40は、削除条件を満たしたページのキャッシュを、第二のノード1B'のメモリから削除する。その後、本フローチャートに示された処理は終了する。

[0158] なお、図12に示したフローチャートでは、キャッシュとなっているデータについて削除の対象とするか否かを判定している。但し、ストレージに保持されているがキャッシュとなっていないデータについても、定期的に保持対象であるか否かを判定し、保持対象となるための条件（ステップS702を参照）を満たさなくなった場合にはストレージから削除することとしてもよい。このようにすることで、ストレージにのみ保持されているようなデータについても、時間の経過に従ってストレージから削除し、ストレージの容量を節約することが出来る。

[0159] <<第二の実施形態に係るシステムの効果>>

第二の実施形態に係るシステムによっても、第一の実施形態に係るシステムと同様の効果を得ることが出来る。また、第二の実施形態では、更に判定部38、記録部39および削除部40を備えることで、第二のノード1B'のストレージやメモリを節約しながら、ユーザー端末9に対して、あたかも第二のノード1B'がデータベース全体に係るデータを有しているかのようにサービスを提供することが出来る。

[0160] また、判定部38や削除部40によって用いられる条件を実施の形態に応じて適宜設定することで、ストレージリソースとネットワークリソースのバランスに応じて、リソースの無駄が少ないシステムを構築することが可能となる。

請求の範囲

[請求項1]

複数のノードを有するデータベースシステムであって、
前記複数のノードのうち、第一のノードは、

該第一のノードによって管理されているデータベースのトランザクションログを、該トランザクションログに係る命令の順序を把握可能な識別情報とともに保持するログ保持手段と、

前記複数のノードのうち、前記第一のノードと異なる第二のノードから、該第一のノードによって管理されているデータの送信要求を受信する要求受信手段と、

前記送信要求に応じて、前記ログ保持手段によって保持されている前記識別情報のうち、要求されたデータに係る、所定の時点におけるトランザクションログを示す特定識別情報を、前記第二のノードに送信する特定識別情報送信手段と、

少なくとも前記所定の時点より後の前記トランザクションログおよび前記識別情報を、互いに関連づけて前記第二のノードに送信するログ送信手段と、

前記送信要求に応じて、前記データベースによって管理されているデータのうち要求されたデータを、前記所定の時点以降に、前記第二のノードに送信するデータ送信手段と、

を備え、

前記第二のノードは、

ユーザー端末からの検索要求を受け付ける検索要求受付手段と、

前記検索要求に応じて、前記第一のノードに対して、前記データベース中の少なくとも前記検索要求に係るデータの前記送信要求を送信する要求送信手段と、

前記第一のノードから、前記トランザクションログ、該トランザクションログの識別情報、前記特定識別情報および前記データを受信する受信手段と、

受信された前記データが該第二のノードのメモリに展開されてデータの検索または処理に供される状態となった場合に、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに係る命令を、メモリに展開された前記データに対して実行する実行手段と、

前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータに基づいて、前記検索要求への応答を前記ユーザー端末に対して返す検索応答手段と、

前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータを、前記検索応答手段による応答に用いた後、所定の条件に従って自動的に該第二のノードから削除する削除手段と、

を備える、

データベースシステム。

[請求項2] 前記削除手段は、前記データに関する処理に関する時点から所定時間経過した場合に、前記データを削除する、

請求項1に記載のデータベースシステム。

[請求項3] 前記第二のノードは、前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータを、該第二のノードにおける保持対象とするか否かを判定する判定手段を更に備え、

前記削除手段は、前記判定手段によって前記保持対象と判定されなかったデータを、前記所定の条件に従って自動的に削除する、

請求項1または2に記載のデータベースシステム。

[請求項4] 前記第二のノードは、前記判定手段によって前記保持対象と判定されたデータを、不揮発性の記憶装置に記録する記録手段を更に備える、

請求項3に記載のデータベースシステム。

[請求項5] 前記データ送信手段は、前記データを、前記トランザクションログに係る命令の順序が互いに依存関係にあるレコードが同一の管理単位に入るように区切られた所定の管理単位毎に送信する、

請求項1から4の何れか一項に記載のデータベースシステム。

[請求項6] 受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに基づいて、命令の対象となるデータが収容された前記所定の管理単位と命令の内容との関係を示すマップを生成するマップ生成手段を更に備え、

前記実行手段は、前記マップを参照して、前記トランザクションログに係る命令を、前記所定の管理単位毎に実行する、

請求項5に記載のデータベースシステム。

[請求項7] 受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに基づいて、命令の対象となるデータと命令の内容との関係を示すマップを生成するマップ生成手段を更に備え、

前記実行手段は、前記マップを参照して、前記トランザクションログに係る命令を、命令の対象となるデータに対して実行する、

請求項1に記載のデータベースシステム。

[請求項8] 前記データ送信手段は、前記第一のノードのメモリに展開されてデータの検索または処理に供されている状態のデータを、前記第二のノードに送信する、

請求項1から7の何れか一項に記載のデータベースシステム。

[請求項9] 前記データ送信手段は、前記データベースの管理情報を、前記第二のノードに送信し、

前記要求送信手段は、前記管理情報を参照して、前記第一のノードに対して、前記データベース中のデータを指定して送信要求を送信す

る、

請求項 1 から 8 の何れか一項に記載のデータベースシステム。

[請求項10]

前記第二のノードは、

前記第一のノードから受信したデータを直接メモリに展開して、データの検索または処理に供する展開手段を更に備え、

前記実行手段は、受信したデータが前記展開手段によってメモリに展開されたことを受けて、前記トランザクションログに係る命令を前記データに対して実行する、

請求項 1 から 9 の何れか一項に記載のデータベースシステム。

[請求項11]

前記特定識別情報送信手段は、前記特定識別情報として、ログ保持手段によって保持されている前記識別情報のうち、最新のトランザクションログを示す識別情報を送信する、

請求項 1 から 10 の何れか一項に記載のデータベースシステム。

[請求項12]

前記データは、テーブル単位またはページ単位で送受信される、

請求項 1 から 11 の何れか一項に記載のデータベースシステム。

[請求項13]

前記実行手段は、前記トランザクションログに係る命令を、前記識別情報によって把握される命令の順序に応じて、受信された前記データに対して実行する、

請求項 1 から 12 の何れか一項に記載のデータベースシステム。

[請求項14]

データベースを管理する管理手段と、

前記データベースのトランザクションログを、該トランザクションログに係る命令の順序を把握可能な識別情報とともに保持するログ保持手段と、

他の情報処理装置から、前記データベースによって管理されているデータの送信要求を受信する要求受信手段と、

前記送信要求に応じて、前記ログ保持手段によって保持されている前記識別情報のうち、要求されたデータに係る、所定の時点におけるトランザクションログを示す特定識別情報を、前記他の情報処理装置

に送信する特定識別情報送信手段と、

少なくとも前記所定の時点より後の前記トランザクションログおよび前記識別情報を、互いに関連づけて前記他の情報処理装置に送信するログ送信手段と、

前記送信要求に応じて、前記データベースによって管理されているデータのうち要求されたデータを、前記所定の時点以降に、前記他の情報処理装置に送信するデータ送信手段と、

を備える情報処理装置。

[請求項15]

ユーザー端末からの検索要求を受け付ける検索要求受付手段と、

前記検索要求に応じて、データベースを管理する他の情報処理装置に対して、前記データベース中の少なくとも前記検索要求に係るデータの前記送信要求を送信する要求送信手段と、

前記他の情報処理装置から、該データベースのトランザクションログ、該トランザクションログに係る命令の順序を把握可能な識別情報、前記識別情報のうち所定の時点におけるトランザクションログを示す特定識別情報、および前記データベースによって管理されているデータを受信する受信手段と、

受信された前記データがメモリに展開されてデータの検索または処理に供される状態となった場合に、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに係る命令を、メモリに展開された前記データに対して実行する実行手段と、

前記検索要求の結果前記他の情報処理装置から得られ、前記トランザクションに基づく命令が実行されたデータに基づいて、前記検索要求への応答を前記ユーザー端末に対して返す検索応答手段と、

前記検索要求の結果前記他の情報処理装置から得られ、前記トランザクションに基づく命令が実行されたデータを、前記検索応答手段による応答に用いた後、所定の条件に従って自動的に該情報処理装置か

ら削除する削除手段と、

を備える情報処理装置。

[請求項16]

複数のノードを有するデータベースシステムにおいて、

前記複数のノードのうち、第一のノードが、

該第一のノードによって管理されているデータベースのトランザクションログを、該トランザクションログに係る命令の順序を把握可能な識別情報とともに保持するログ保持ステップと、

前記複数のノードのうち、前記第一のノードと異なる第二のノードから、該第一のノードによって管理されているデータの送信要求を受信する要求受信ステップと、

前記送信要求に応じて、前記ログ保持ステップで保持されている前記識別情報のうち、要求されたデータに係る、所定の時点におけるトランザクションログを示す特定識別情報を、前記第二のノードに送信する特定識別情報送信ステップと、

少なくとも前記所定の時点より後の前記トランザクションログおよび前記識別情報を、互いに関連づけて前記第二のノードに送信するログ送信ステップと、

前記送信要求に応じて、前記データベースによって管理されているデータのうち要求されたデータを、前記所定の時点以降に、前記第二のノードに送信するデータ送信ステップと、

を実行し、

前記第二のノードが、

ユーザー端末からの検索要求を受け付ける検索要求受付ステップと、

前記検索要求に応じて、前記第一のノードに対して、前記データベース中の少なくとも前記検索要求に係るデータの前記送信要求を送信する要求送信ステップと、

前記第一のノードから、前記トランザクションログ、該トランザ

クションログの識別情報、前記特定識別情報および前記データを受信する受信ステップと、

受信された前記データが該第二のノードのメモリに展開されてデータの検索または処理に供される状態となった場合に、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに係る命令を、メモリに展開された前記データに対して実行する実行ステップと、

前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータに基づいて、前記検索要求への応答を前記ユーザー端末に対して返す検索応答ステップと、

前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータを、前記検索応答ステップにおける応答に用いた後、所定の条件に従って自動的に該第二のノードから削除する削除ステップと、

を実行する、

方法。

[請求項17]

複数のノードを有するデータベースシステムにおいて、

前記複数のノードのうち、第一のノードを、

該第一のノードによって管理されているデータベースのトランザクションログを、該トランザクションログに係る命令の順序を把握可能な識別情報とともに保持するログ保持手段と、

前記複数のノードのうち、前記第一のノードと異なる第二のノードから、該第一のノードによって管理されているデータの送信要求を受信する要求受信手段と、

前記送信要求に応じて、前記ログ保持手段によって保持されている前記識別情報のうち、要求されたデータに係る、所定の時点におけるトランザクションログを示す特定識別情報を、前記第二のノードに

送信する特定識別情報送信手段と、

少なくとも前記所定の時点より後の前記トランザクションログおよび前記識別情報を、互いに関連づけて前記第二のノードに送信するログ送信手段と、

前記送信要求に応じて、前記データベースによって管理されているデータのうち要求されたデータを、前記所定の時点以降に、前記第二のノードに送信するデータ送信手段と、

として機能させ、

前記第二のノードを、

ユーザー端末からの検索要求を受け付ける検索要求受付手段と、

前記検索要求に応じて、前記第一のノードに対して、前記データベース中の少なくとも前記検索要求に係るデータの前記送信要求を送信する要求送信手段と、

前記第一のノードから、前記トランザクションログ、該トランザクションログの識別情報、前記特定識別情報および前記データを受信する受信手段と、

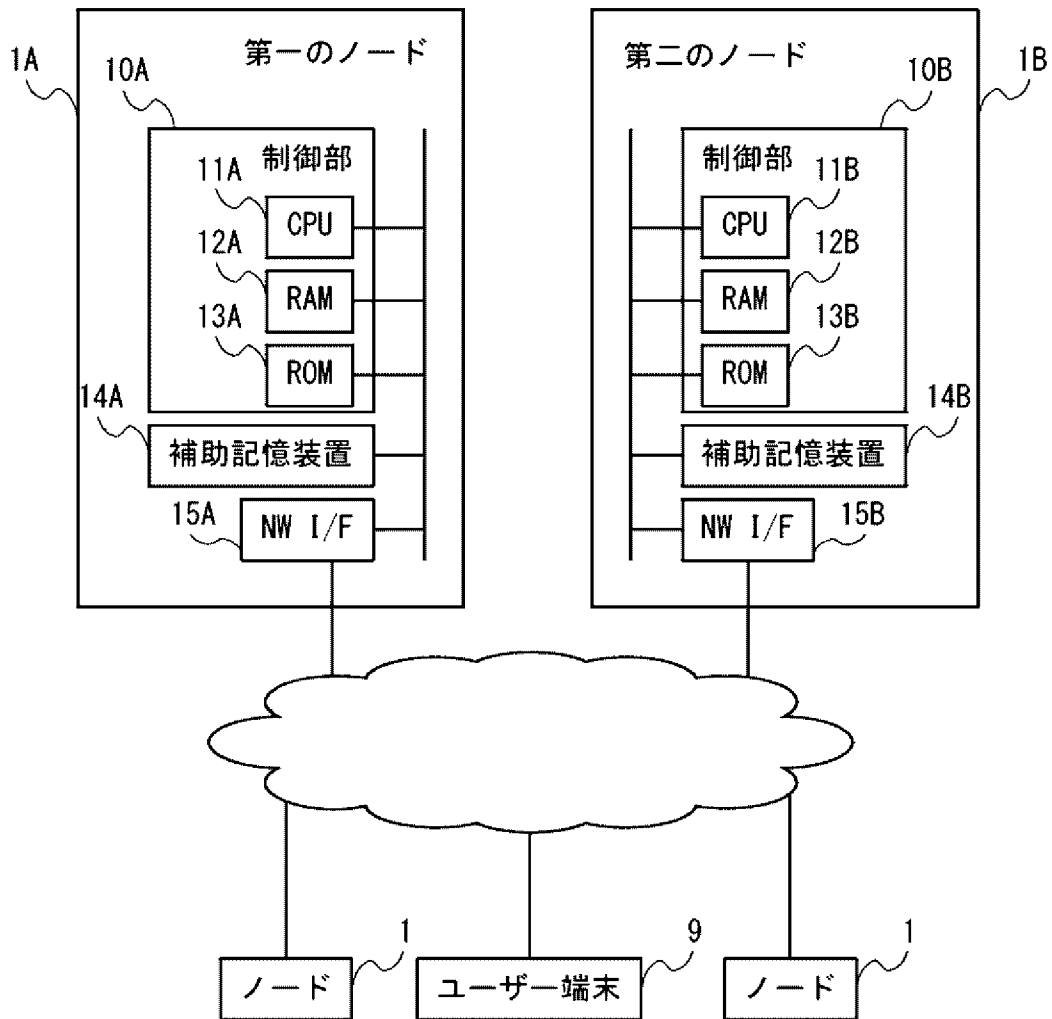
受信された前記データが該第二のノードのメモリに展開されてデータの検索または処理に供される状態となった場合に、受信された前記トランザクションログのうち、少なくとも前記特定識別情報が示すトランザクションログよりも新しい前記トランザクションログに係る命令を、メモリに展開された前記データに対して実行する実行手段と、

前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータに基づいて、前記検索要求への応答を前記ユーザー端末に対して返す検索応答手段と、

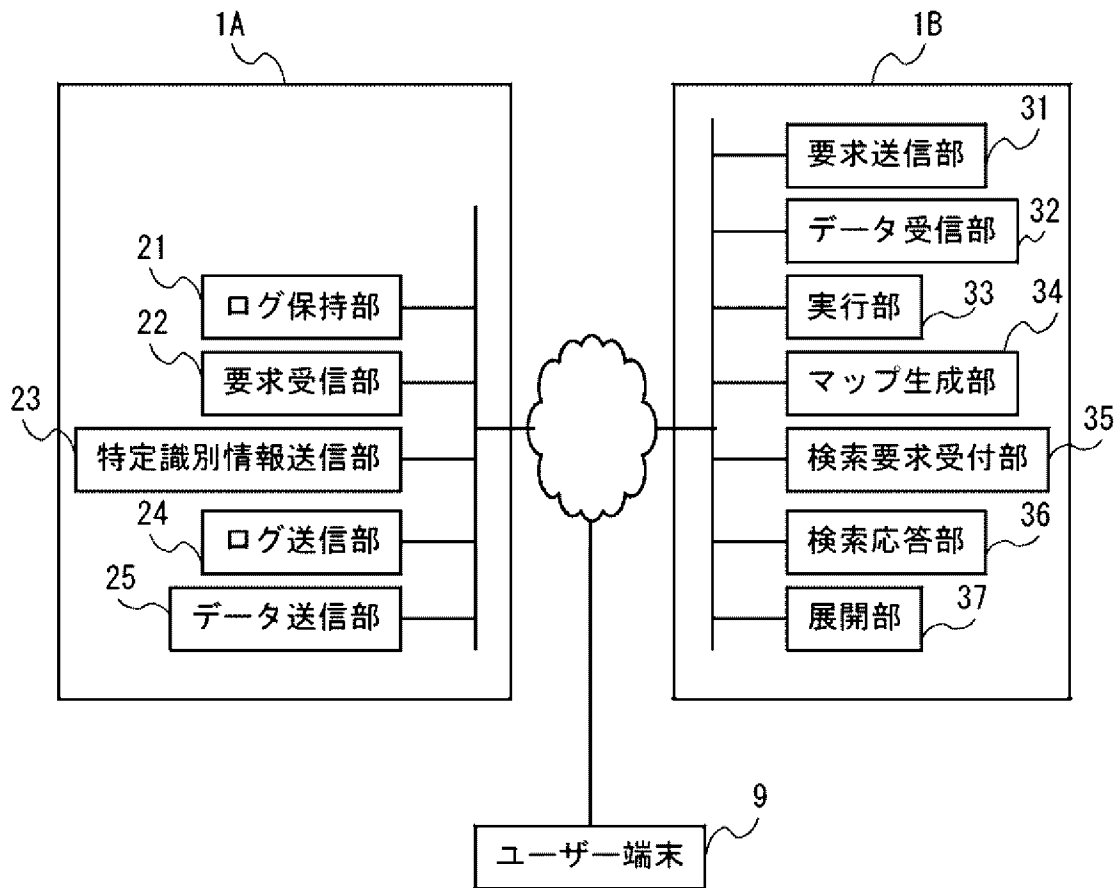
前記検索要求の結果前記第一のノードから得られ、前記トランザクションに基づく命令が実行されたデータを、前記検索応答手段による応答に用いた後、所定の条件に従って自動的に該第二のノードから

削除する削除手段と、
として機能させる、
プログラム。

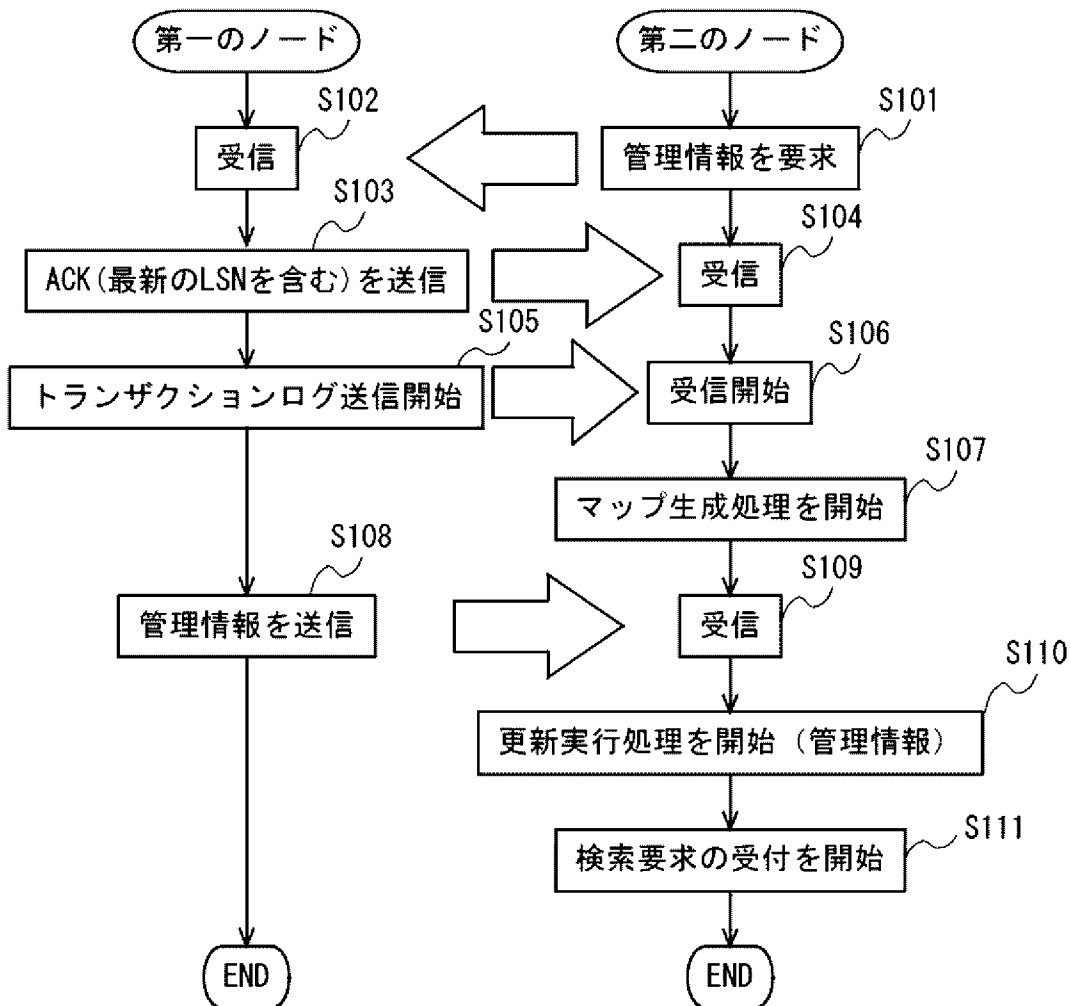
[図1]



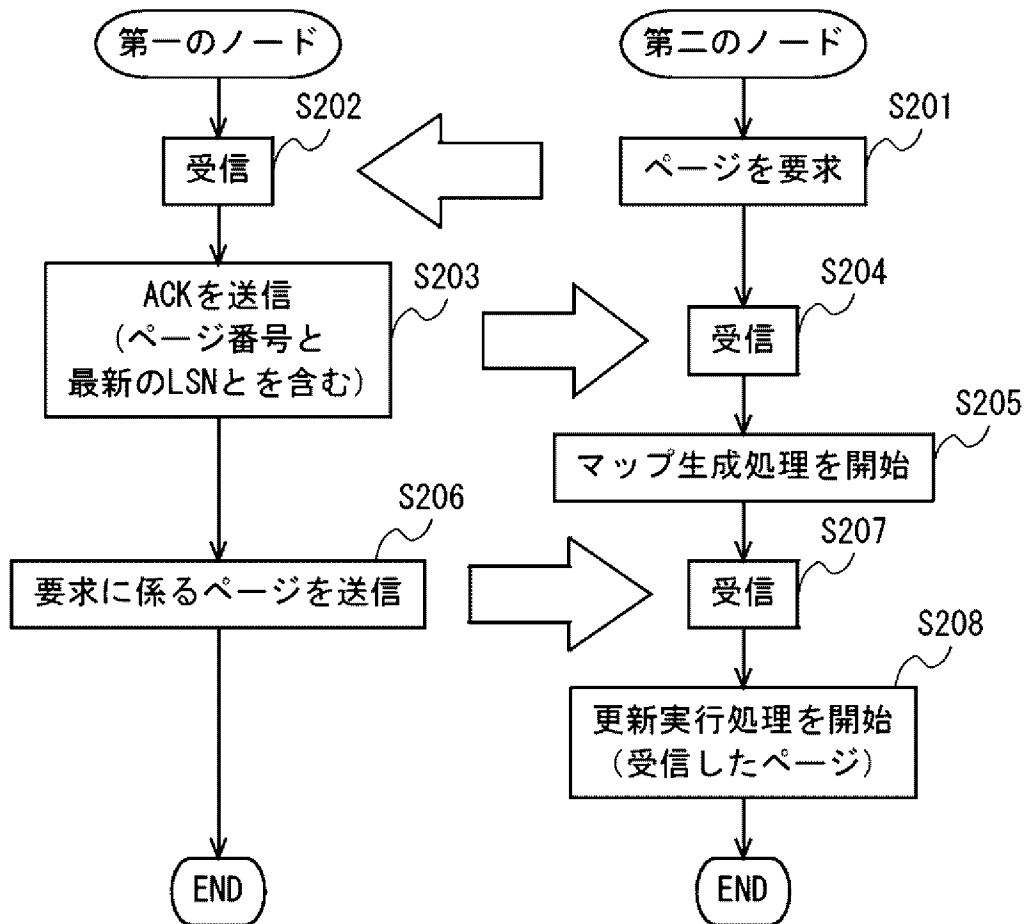
[図2]



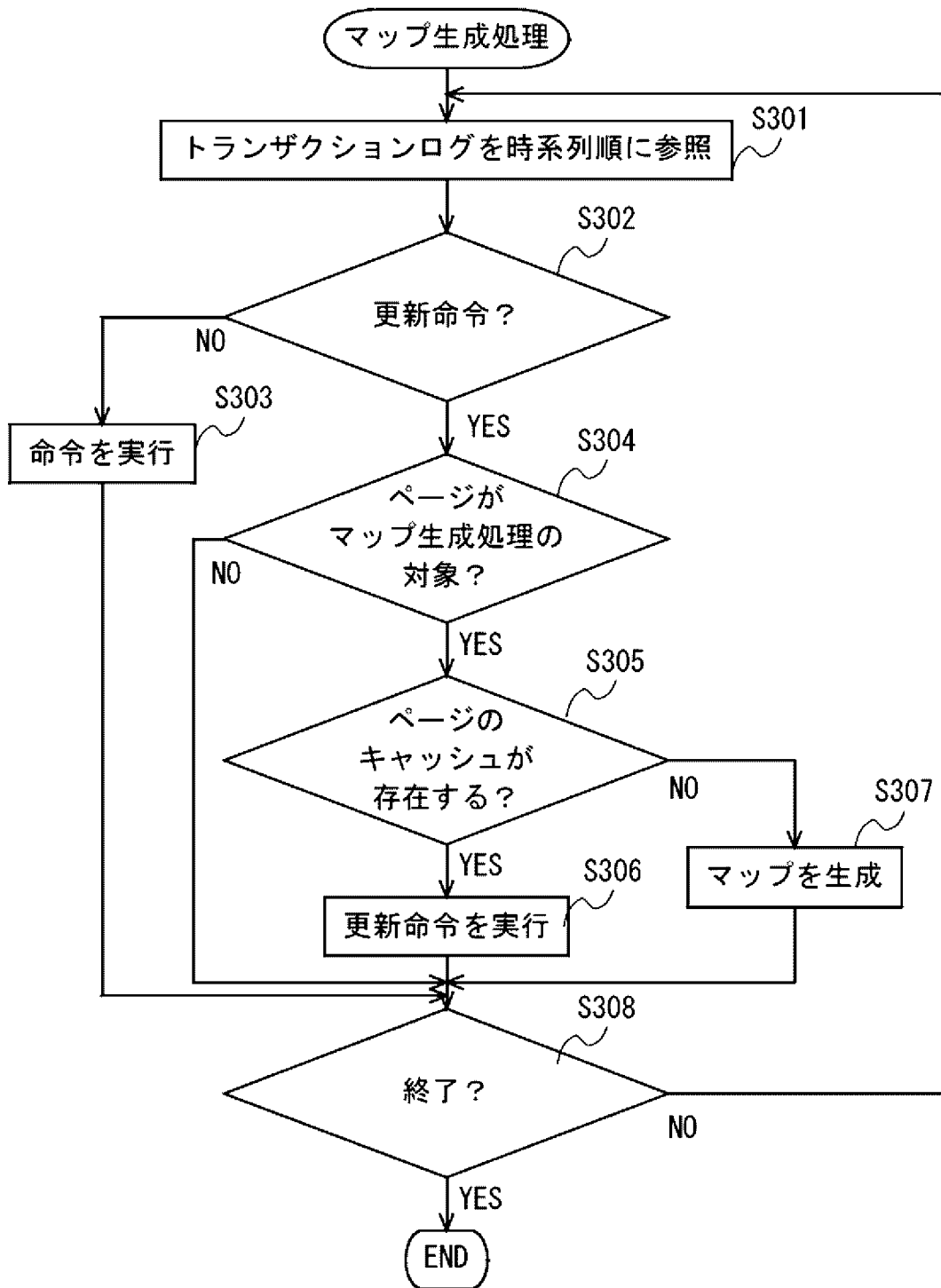
[図3]



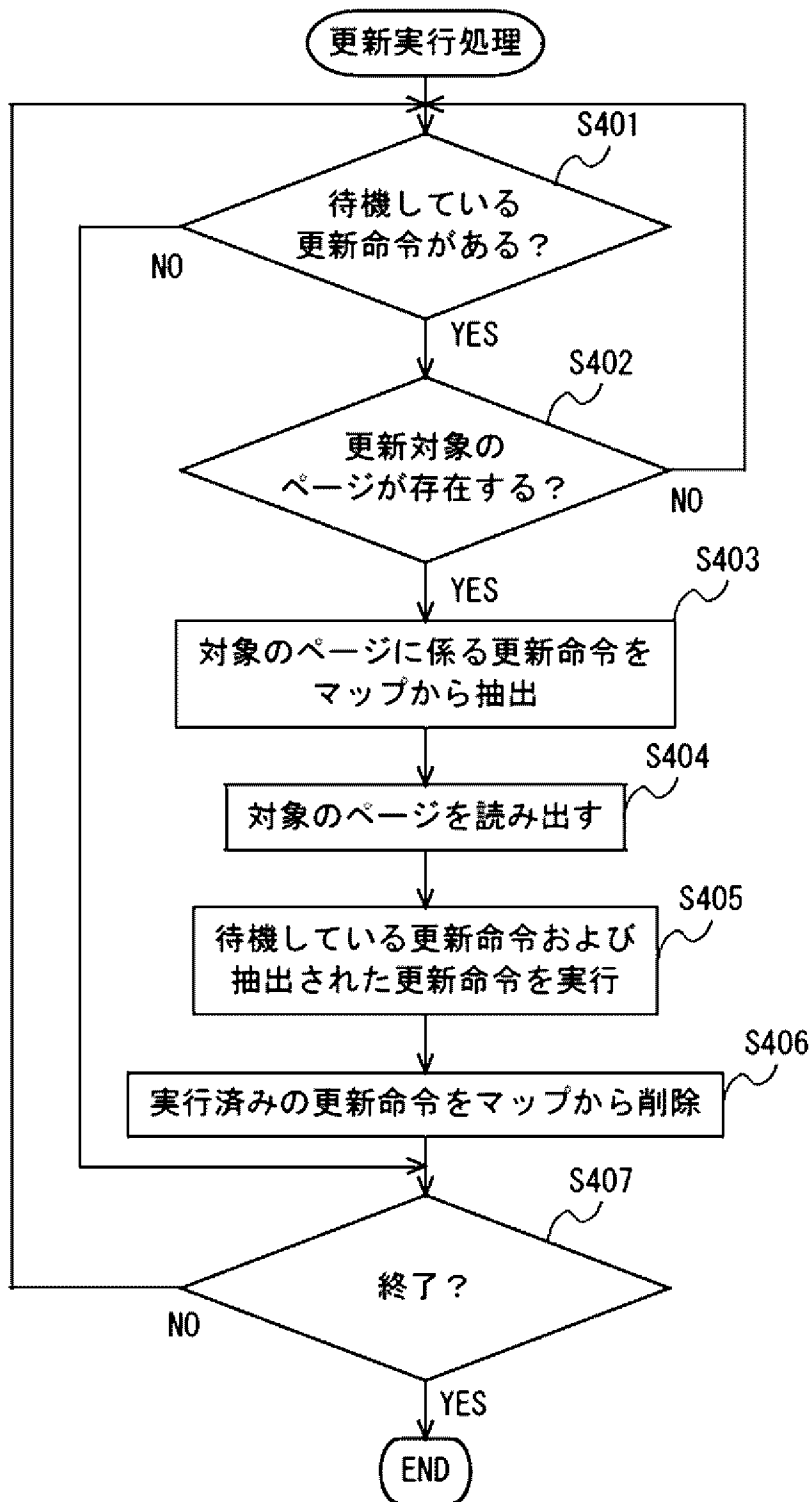
[図4]



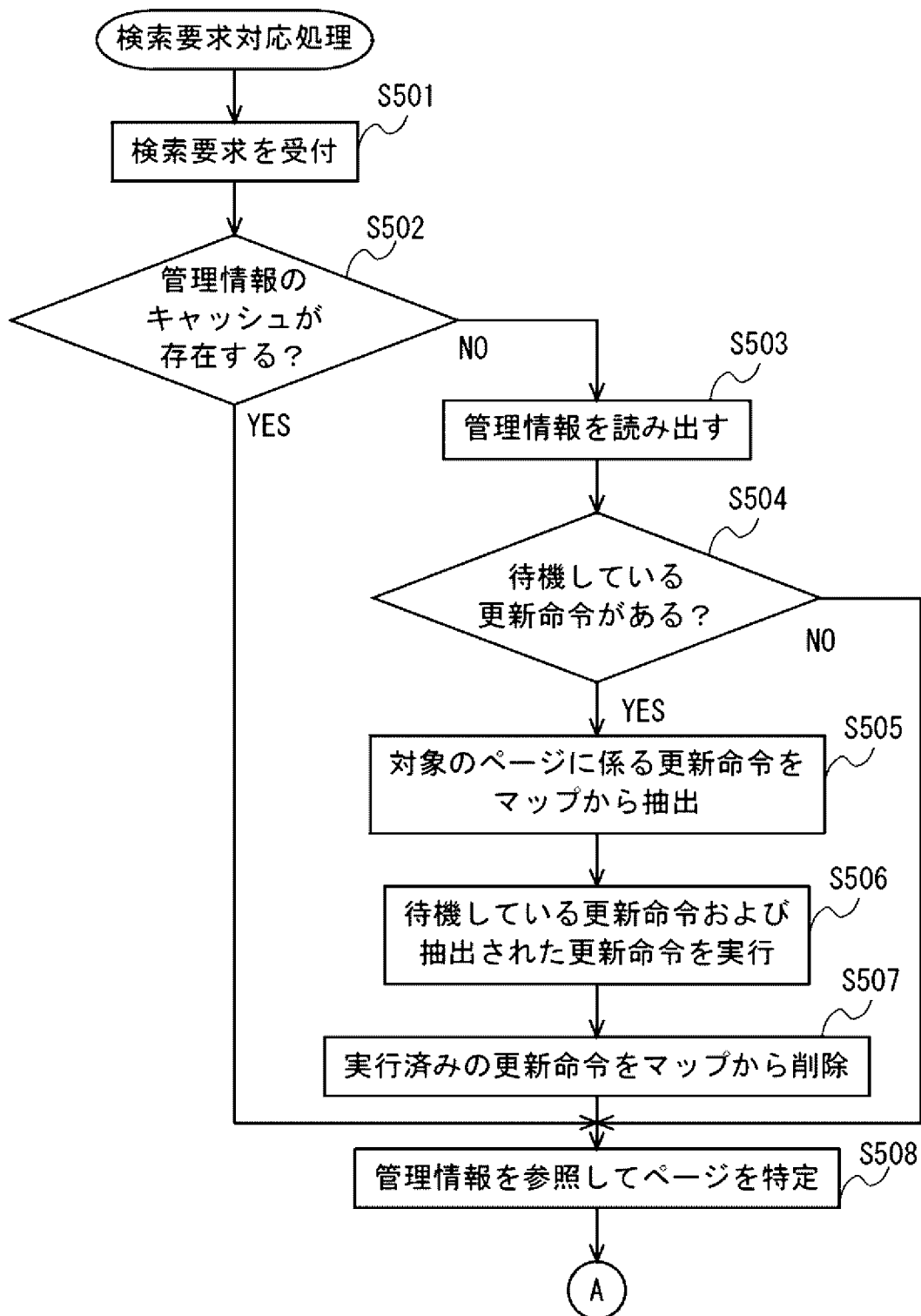
[図5]



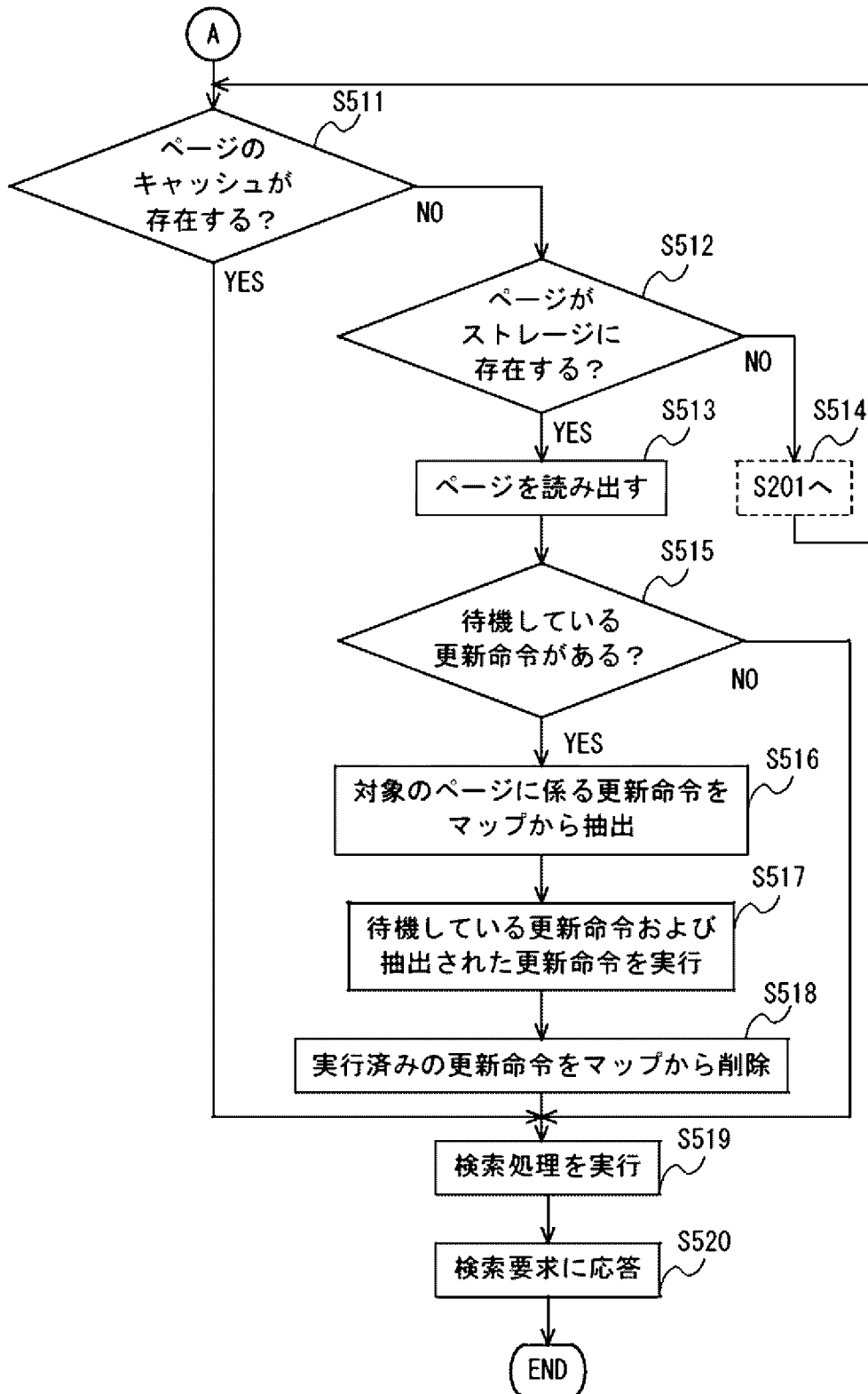
[図6]



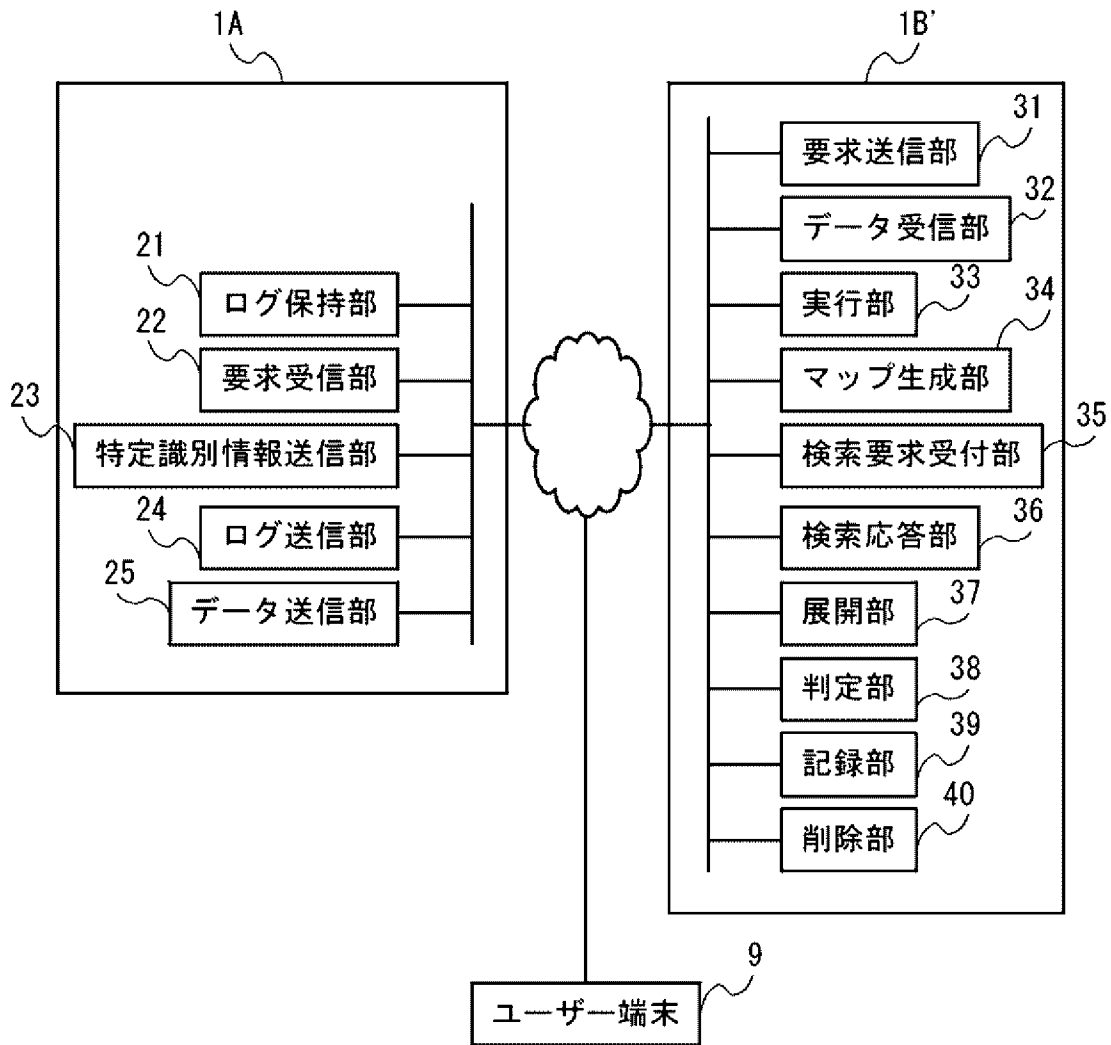
[図7]



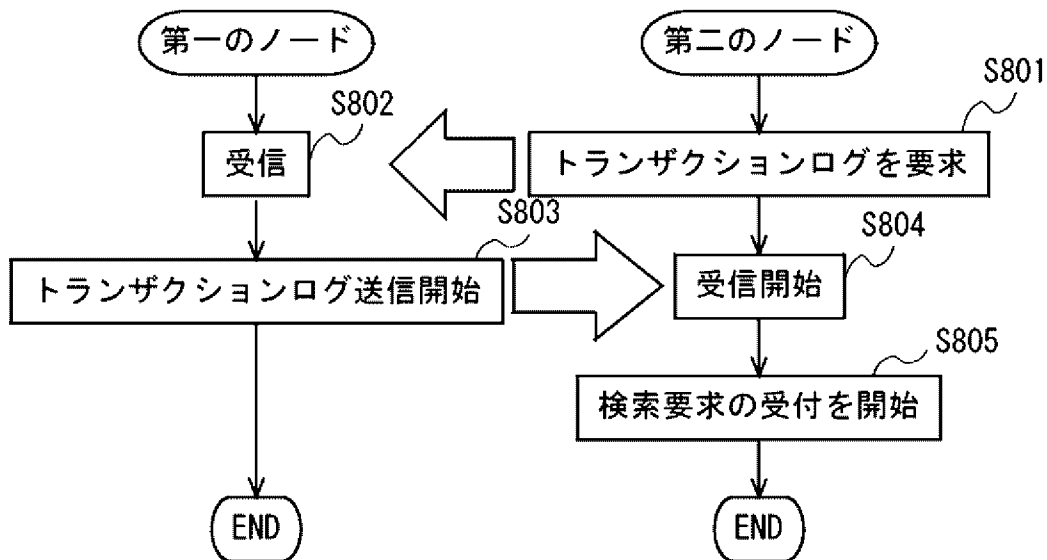
[図8]



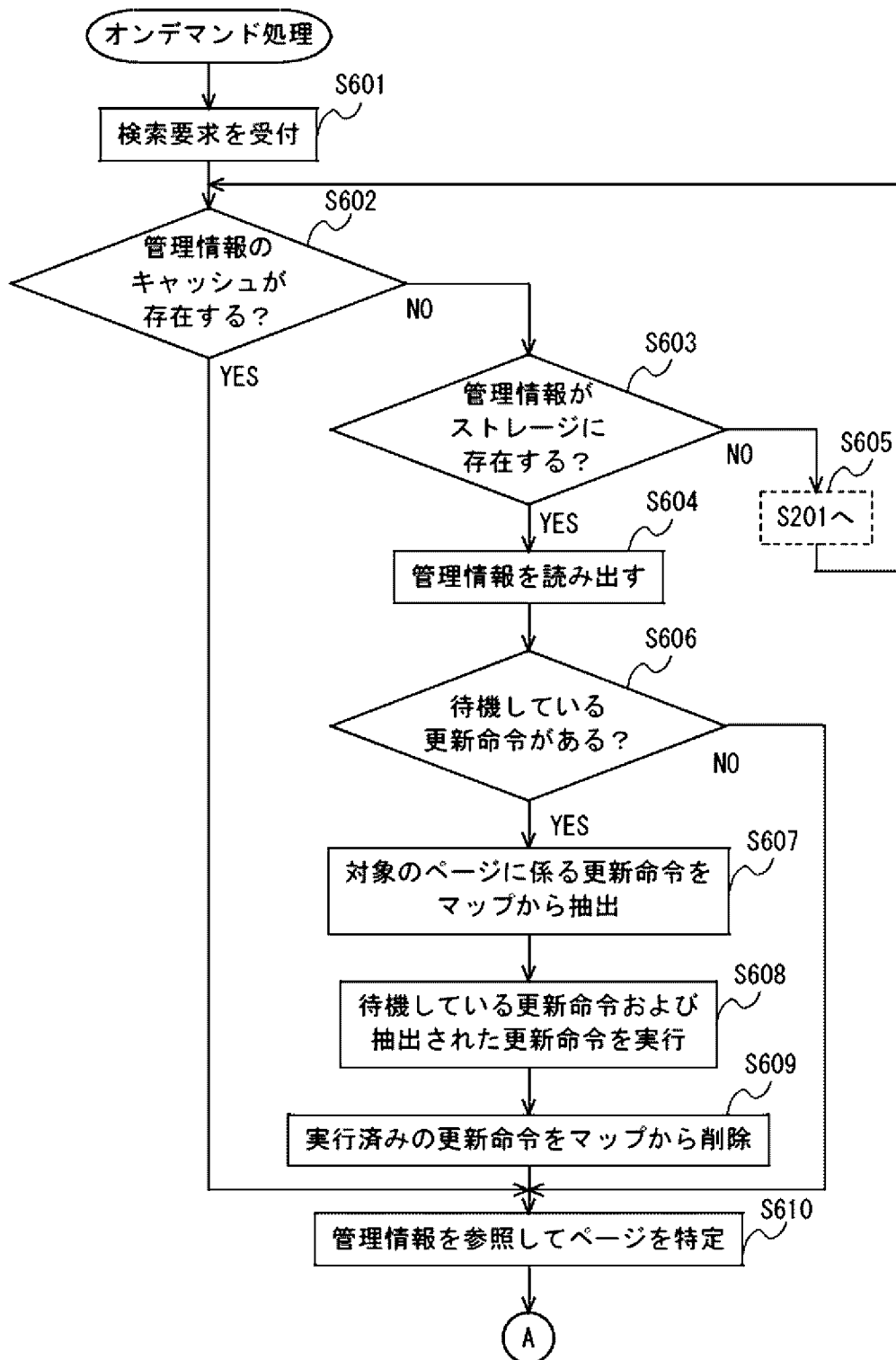
[図9]



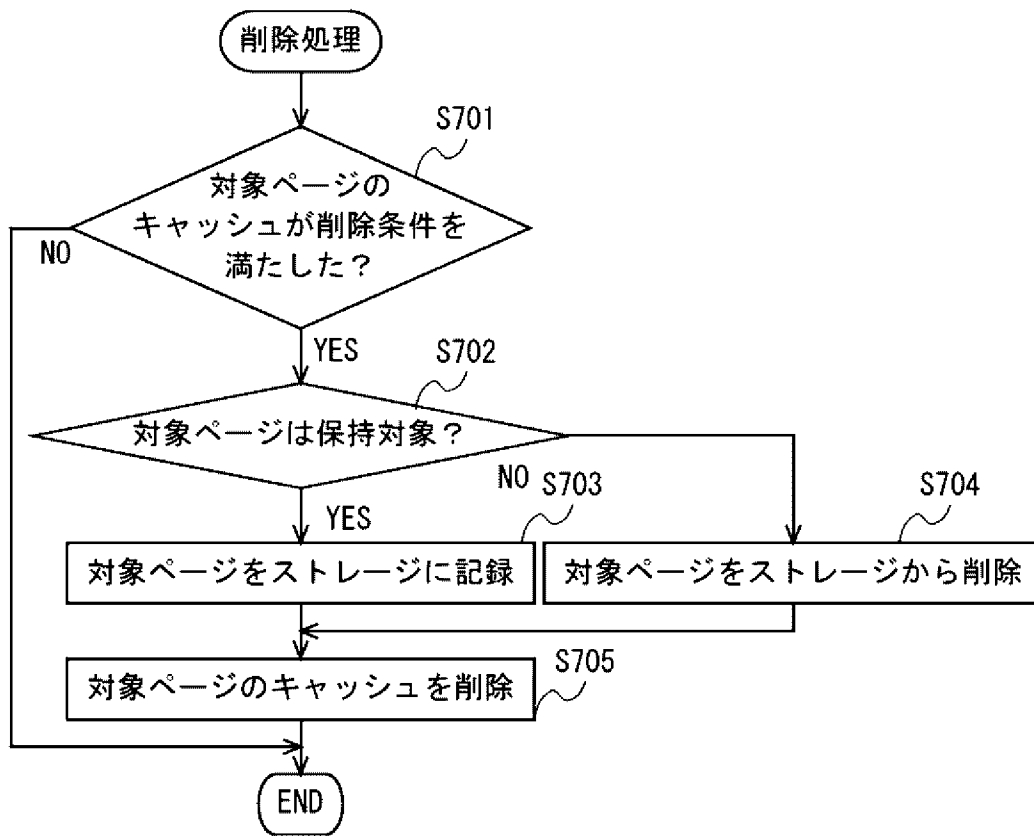
[図10]



[図11]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2014/058382

A. CLASSIFICATION OF SUBJECT MATTER
G06F12/00(2006.01)i, G06F17/30(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, G06F17/30

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	1996-2014
Kokai Jitsuyo Shinan Koho	1971-2014	Toroku Jitsuyo Shinan Koho	1994-2014

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 A	WO 2010/106991 A1 (Kokolink Corp.), 23 September 2010 (23.09.2010), entire text; all drawings (particularly, paragraphs [0007] to [0013], [0029] to [0030]; fig. 2) & US 2012/0011098 A1 & EP 2410431 A1	14 1-13, 15-17
A	JP 2013-15915 A (Fujitsu Ltd.), 24 January 2013 (24.01.2013), entire text; all drawings & US 2013/0006930 A1	1-17
A	JP 2009-15476 A (Hitachi, Ltd.), 22 January 2009 (22.01.2009), entire text; all drawings & US 2009/0013012 A1	1-17

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 27 June, 2014 (27.06.14)	Date of mailing of the international search report 08 July, 2014 (08.07.14)
---	--

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））
 Int.Cl. G06F12/00(2006.01)i, G06F17/30(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. G06F12/00, G06F17/30

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 1922-1996年
 日本国公開実用新案公報 1971-2014年
 日本国実用新案登録公報 1996-2014年
 日本国登録実用新案公報 1994-2014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 A A A	WO 2010/106991 A1 (株式会社ココリンク) 2010.09.23, 全文全図 (特に[0007]-[0013], [0029]-[0030]及び 図 2) & US 2012/0011098 A1 & EP 2410431 A1 JP 2013-15915 A (富士通株式会社) 2013.01.24, 全文全図 & US 2013/0006930 A1 JP 2009-15476 A (株式会社日立製作所) 2009.01.22, 全文全図 & US 2009/0013012 A1	14 1-13, 15-17 1-17 1-17

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	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 27.06.2014	国際調査報告の発送日 08.07.2014
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 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) 原 秀人 電話番号 03-3581-1101 内線 3565

5U	9644
----	------