

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4916892号
(P4916892)

(45) 発行日 平成24年4月18日(2012.4.18)

(24) 登録日 平成24年2月3日(2012.2.3)

(51) Int. Cl. F I
G06F 11/34 (2006.01)
 G06F 11/34 B
 G06F 11/34 D

請求項の数 11 (全 19 頁)

(21) 出願番号	特願2007-560 (P2007-560)	(73) 特許権者	390019839
(22) 出願日	平成19年1月5日(2007.1.5)		三星電子株式会社
(65) 公開番号	特開2007-188497 (P2007-188497A)		Samsung Electronics
(43) 公開日	平成19年7月26日(2007.7.26)		Co., Ltd.
審査請求日	平成19年1月5日(2007.1.5)		大韓民国京畿道水原市靈通区梅灘洞416
審判番号	不服2010-25976 (P2010-25976/J1)		416, Maetan-dong, Yeongtong-gu, Suwon-si,
審判請求日	平成22年11月17日(2010.11.17)		Gyeonggi-do, Republic of Korea
(31) 優先権主張番号	10-2006-0002867	(74) 代理人	100070150
(32) 優先日	平成18年1月10日(2006.1.10)		弁理士 伊東 忠彦
(33) 優先権主張国	韓国 (KR)	(74) 代理人	100091214
			弁理士 大貫 進介
		(74) 代理人	100107766
			弁理士 伊東 忠重

最終頁に続く

(54) 【発明の名称】 トランザクション処理のためのログ情報管理システムおよび方法

(57) 【特許請求の範囲】

【請求項1】

バッファ上の更新されるページが新しく割り当てられるページであるか否かを判断するインデックス/レコード/カタログ管理手段と、

フリーページのリストと未完了フリーページのリストとを管理して、前記更新されるページが新しく割り当てられるページである場合、前記フリーページのリストまたは前記未完了フリーページのリストから新規ページを割り当てる格納管理手段であって、前記フリーページは、データが新しく割当可能なページであり、前記未完了フリーページは、まだ完了しない任意のトランザクション内で削除されたページであり、前記新規ページは、前記フリーページのリストで使用可能なフリーページを割り当て、前記フリーページのリスト

10

に使用可能なフリーページがない時には、前記未完了フリーページのリストから前記フリーページを割り当てる格納管理手段と、

前記新規ページに対応するバッファページに対する識別情報を設定するバッファ管理手段であって、ページログシーケンス番号を前記識別情報とするバッファ管理手段とを含み、

【請求項2】

前記バッファ管理手段は、前記識別情報に、生成可能な最も大きいページログシーケンス番号を含める請求項1に記載のログ情報管理システム。

20

【請求項 3】

前記新規ページは、前記フリーページのリストから割り当てられる請求項 1 に記載のログ情報管理システム。

【請求項 4】

前記新規ページが前記未完了フリーページのリストから割り当てられた場合には前記格納管理手段が前記ディスクから該当ページを指定された、前記ディスク上のバッファ空間に読み込む請求項 1 に記載のログ情報管理システム。

【請求項 5】

前記バッファ管理手段によって管理される、前記ディスク上のバッファで任意のページに対するページログシーケンス番号が、生成可能な最も大きい既に設定された値に該当する場合、前記格納管理手段は前記ページを前記ディスクに記録する前に前記ページログシーケンス番号の値を既に設定された任意のページログシーケンス番号に変更する請求項 1 に記載のログ情報管理システム。

10

【請求項 6】

前記インデックス/レコード/カタログ管理手段は、前記バッファ管理モジュールによって管理される、前記ディスク上のバッファにあるページに対するページログシーケンス番号が、生成可能な最も大きい既に設定された値に該当する場合に、データ値が更新されたページに対するログ情報を生成しない請求項 1 に記載のログ情報管理システム。

【請求項 7】

インデックス/レコード/カタログ管理手段、格納管理手段、及びバッファ管理手段を含むシステムが行うログ情報管理方法であって、

20

バッファ上の更新されるページが新しく割り当てられるページであるか否かを前記インデックス/レコード/カタログ管理手段によって判断する (a) ステップと、

前記更新されるページが新しく割り当てられるページである場合、フリーページのリストまたは未完了フリーページのリストから新規ページを前記格納管理手段によって割り当てる (b) ステップであって、前記フリーページは、データが新しく割当可能なページであり、前記未完了フリーページは、まだ完了しない任意のランザクション内で削除されたページであり、前記新規ページは、前記フリーページのリストで使用可能なフリーページを割り当て、前記フリーページのリストに使用可能なフリーページがない時には、前記未完了フリーページのリストから前記フリーページを割り当てるステップと、

30

前記新規ページに対応するバッファページに対する識別情報を前記バッファ管理手段によって設定する (c) ステップとを含み、ページログシーケンス番号を前記識別情報とし

、前記バッファページに対するデータ値変更については、ディスクのログファイル上に格納されるログ情報が発生しないログ情報管理方法。

【請求項 8】

前記 (c) ステップは、前記識別情報に、生成可能な最も大きいページログシーケンス番号を前記バッファ管理手段によって含めるステップを含む請求項 7 に記載のログ情報管理方法。

【請求項 9】

40

前記 (b) ステップは、前記新規ページが前記未完了フリーページのリストから割り当てられた場合に前記割り当てられた新規ページがディスクから指定された、前記ディスク上のバッファ空間に前記格納管理手段によってロードされるステップを含む請求項 7 に記載のログ情報管理方法。

【請求項 10】

前記バッファページに対するページログシーケンス番号が、生成可能な最も大きい既に設定された値に該当する場合、前記バッファページをディスクに記録する前に前記ページログシーケンス番号の値を既に設定された任意のページログシーケンス番号に前記格納管理手段によって変更する (d) ステップをさらに含む請求項 7 に記載のログ情報管理方法。

50

【請求項 11】

前記バッファページに対するページログシーケンス番号が、生成可能な最も大きい既に設定された値に該当する場合、前記バッファページに対するログ情報を前記格納管理手段によって生成しない(d)ステップをさらに含む請求項7に記載のログ情報管理方法。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、データベース管理システムでのログ情報管理に関し、より詳しくは、新規に割り当てられたデータページに関して発生するログ情報の大きさを最小化するログ情報管理システムおよび方法に関するものである。

10

【背景技術】**【0002】**

データベース管理システムにおける「ログ」または「ロギング」という一般的にアプリケーションプログラムが行う作業によって、変更中であるデータがシステム異常またはユーザの不注意によって正常な値を有することができなくなることを防止するために変更状態のデータを別途に格納することを意味する。

【0003】

このようなロギングによりデータベース管理システムは取り消しまたは復旧作業を行うことによって、非正常的に終了したアプリケーションプログラムによって操作中であったデータを本来の正常な状態に戻せるようになる。

20

【0004】

図1は、従来技術に係るログ情報管理システムの構成を示すブロック図である。

【0005】

図1を参照すれば、トランザクションマネージャ110は、トランザクションの開始、終了、取り消し時にログマネージャ130を用いてログ情報を残し、現在進行しているトランザクションの情報を保持する。復旧マネージャ120は特定トランザクションの取り消し時にログマネージャ130を介してログ情報を確保した後、データに行なわれた変更を戻す役割を担当する。特に、復旧マネージャ120は、ログ情報管理システムが非正常的に終了した時に未完了であったトランザクションの影響をデータベースから削除して完了したトランザクションの修正事項をデータベースに反映して、データの一貫性を保持するようにする。ログマネージャ130は、他のモジュールで発生させたデータベースに対する変更されたログ情報を取り集めてバッファ160に反映し、復旧マネージャ120から要請されたログ情報をディスク180から読み込んで提供する。

30

【0006】

インデックス/レコード/カタログマネージャ140は、ログ情報管理システムを構成する主要データのインデックス、レコード、カタログ情報を各々管理するモジュールであって、バッファマネージャ150に必要なデータ(インデックス、レコード、カタログ)をバッファ160にロードさせて必要な値を読み込んだり変更したりするように要請する。ログ情報管理システムにおいて、大部分のログ情報はインデックス/レコード/カタログマネージャ140によって発生する。

40

【0007】

バッファマネージャ150は、バッファ160を管理しながら、他のモジュールの要請によってディスク180に格納されたインデックス、レコード、またはカタログに対するログ情報を入れたページをバッファにロードさせたり、バッファ160にロードされたページをディスク180に格納したりする。ディスク180でデータページを読み込んだり記録したりする作業は格納マネージャ170によって行われる。

【0008】

バッファ160は、メモリの一部であって、ログ情報管理システムが確保して専用する空間である。バッファ160にはログ情報が記録され、図2ではこのようなログ情報の例

50

としてログレコードを示している。

【0009】

ログレコード200は、ログヘッダ210、以前データイメージ220、そして更新データイメージ230を含む。

【0010】

ログヘッダ210は、ログシーケンス番号(以下、「LSN」)、トランザクション識別情報、以前LSN情報、ページ識別情報、オフセット情報、データ長情報などを含む。

【0011】

LSNは該当ログレコードを識別する情報であり、トランザクション識別情報は該当ログレコードが意味する変更を発生させたトランザクションの識別情報である。以前LSN情報は、トランザクション識別情報が示すトランザクション内で該当ログ直前に発生させたログレコードの識別情報である。ページ識別情報は、該当ログレコードが含んでいる変更作業がどのページに対して行われたかを示し、オフセット情報は、ページ識別情報によって識別されるページのどの位置からデータ変更が発生したかを示し、データ長情報は変更されたデータの大きさを示す情報である。

10

【0012】

以前データイメージ220は変更される以前のデータ値を、更新データイメージ230は変更された以後のデータ値を示す。

【0013】

バッファマネージャ150はディスク180からページ単位でデータを読み込むため、バッファ160も同様にページ単位で分割して管理される。

20

【0014】

格納マネージャ170は、バッファマネージャ150の要請によりディスク180の特定ページを読み込んでバッファ160にロードしたりバッファ160の特定ページをディスク180に記録したりする。また、格納マネージャ170は新しいページが要請される時、該当時点で用いられていないディスクページをデータ格納用として割り当てて、既存のディスクページをそれ以上データ格納用として用いない場合には、該当ディスクページを空のディスクページのうちの1つとして管理する。格納マネージャ170は、ページ使用現況を管理する役割をするページ使用管理モジュールをさらに含むことができる。

30

【0015】

アプリケーション190は、クエリープロセッサまたはクエリーエンジンとして動作することができる。アプリケーション190は、トランザクションマネージャ110にトランザクションの開始、取り消しまたは終了を知らせることができ、トランザクションの境界内、即ちトランザクションの開始とトランザクション終了/取り消しの間でインデックス/レコード/カタログマネージャ140を介してログ情報を読み込んだり変更したりすることができる。

【0016】

図1に示された各構成要素間の動作過程を説明すれば次の通りである。

【0017】

まず、アプリケーション190がトランザクションの開始をトランザクションマネージャ110に要請すれば、トランザクションマネージャ110は、新しい1つのトランザクションを生成し、生成されたトランザクションに対する情報をトランザクションが終了するまで保持する。

40

【0018】

そして、アプリケーションがインデックス/レコード/カタログマネージャ140にデータ更新作業を要請すれば、インデックス/レコード/カタログマネージャ140は必要なページをバッファマネージャ150に要請し、バッファ160内に確保した後に必要な更新作業を行う。各更新作業が行われる時毎にインデックス/レコード/カタログマネージャ140は、変更前のデータと変更後のデータに対するログ情報を生成し、ログマネージャ130に伝送する。データ更新作業に対するより具体的な過程は、後述することにする

50

る。

【0019】

アプリケーション190は、データ更新作業がエラーなしに行われたと判断した場合、トランザクションマネージャ110にトランザクションを終了するようにし、そうではない場合にはトランザクションを中断するように要請する。

【0020】

トランザクションが終了する場合に、トランザクションマネージャ110は、ログマネージャ130にトランザクションの終了を意味するログ(以下、「Commit_Log」)情報を生成するように指示する。そして、トランザクションマネージャ110は、ログマネージャ130に前記Commit_Log情報を含むすべてのログ情報をディスク180に記録するように要請する。前記Commit_LogのLSN値より小さいLSN値を有するすべてのログ情報がバッファ160内にある場合、ログマネージャ130は、バッファマネージャ150に前記ログ情報をディスク180のログファイルに記録するように要請する。

10

【0021】

アプリケーション190がトランザクションを中断すると決定した場合には、トランザクションマネージャ110が復旧マネージャ120にトランザクションの取り消しを要請する。復旧マネージャ120は該当トランザクション内で既に変更したデータ値を復旧するために、以前のデータ値を有するログ情報をログマネージャ130に要請する。前記要請は、バッファマネージャ150と格納マネージャ170に順に伝えられて処理される。

20

【0022】

そして、復旧マネージャ120は、以前の値に復旧しなければならないデータページをバッファマネージャ150に要請し、バッファマネージャ150は要請されたページを格納マネージャ170を用いてバッファ160に読み込む。復旧マネージャ120は、ログマネージャ130に要請したログ情報を参照して、変更が発生した部分を探して以前のデータ値(ビフォーイメージ)に復元する。

【0023】

トランザクションが終了したり取り消されたりした後、トランザクションマネージャ110は該当トランザクションの情報をトランザクションテーブルから削除する。

【0024】

一方、アプリケーション190がインデックス/レコード/カタログマネージャ140にデータ更新を要請する場合に発生する動作過程を具体的に説明することにする。

30

【0025】

インデックス/レコード/カタログマネージャ140は、データ更新が物理的にディスク180に及ぼす影響を把握して、ディスク180内のどのディスクページを修正すべきかを決定する。

【0026】

インデックス/レコード/カタログマネージャ140は、アプリケーション190から要請された更新作業がディスクページの削除であるか(「デリート」モード)、新しいページにデータを記録することであるか(「INSERT」モード)、または既存ページに記録されたデータ値の変更であるか(「モディファイ」モード)を判断する。

40

【0027】

「デリート」モードの場合、バッファマネージャ150は、指定されたディスクページがバッファ160内にある場合、該当ページをバッファ160から削除する。そして、バッファマネージャ150は、指定されたディスクページの削除を格納マネージャ170に要請し、格納マネージャ170は該当ページをそれ以上用いないページとして分類する。この時、ディスクページの使用に変更が発生したため、格納マネージャ170は、変更内訳に対するログ情報を生成し、ログマネージャ130に伝送する。

【0028】

「INSERT」モードの場合、バッファマネージャ150は格納マネージャ170に新

50

しいページを割り当ててを要請する。格納マネージャ170が新しいページを割り当てた後にディスクページ使用の変更内訳に対するログ情報を生成してログマネージャ130に伝送する。そして、格納マネージャ170は、要請されたディスクページをバッファマネージャ150が指定したバッファ160の空間にコピーした後、インデックス/レコード/カタログマネージャ140は、該当ページに対するデータ値の挿入を行い、これに対するログ情報を発生させてログマネージャ130に伝送する。

【0029】

「モディファイ」モードの場合、インデックス/レコード/カタログマネージャ140は、変更するディスクページをバッファマネージャ150に要請する。バッファマネージャ150は、該当ディスクページがバッファ160に存在するか否かを検査する。データ値を更新するページが既にバッファ160内に存在すれば、インデックス/レコード/カタログマネージャ140は、該当ページに対するデータ値の変更を行う。

10

【0030】

しかし、データ値を更新するページがバッファ160内に存在しない場合には、格納マネージャ170が要請されたディスクページをバッファマネージャ150が指定したバッファ160の空いた空間にコピーする。この場合、使用可能な空いた空間がない場合にバッファマネージャ150は適切なページを選定して、ディスク180に送ってバッファ160内に空いた空間を確保する。そして、格納マネージャ170が要請されたディスクページをバッファマネージャ150が指定したバッファ160の空いた空間にコピーする。そして、インデックス/レコード/カタログマネージャ140は、該当ページに対するデータ値の変更を行い、データ値変更に対するログ情報を発生させて、ログマネージャ130に伝送する。

20

【0031】

一方、データベースでは、ログレコードの大量挿入やインデックスの生成時多くのページが新規に割り当てられ、割り当てられたページに対するデータ値が変更される。ところで、上記のような従来のログ情報管理システムでは、すべてのデータ値の変更に対して以前データイメージ(ビフォーイメージ)と更新データイメージ(アフターイメージ)とがログ情報として残る。このような特性のために、あるページがインデックス、レコードまたはカタログ情報を格納するためのページとして新しく割り当てられた後に、ページ全体に対し新しいデータ値が書き込まれるようになれば、従来技術ではヘッダー情報を無視するとしても、新しく割り当てられたページサイズの2倍に達するログ情報が発生する。即ち、N個のデータページが新規割り当てられた後に各ページ全体に亘って値が書き込まれる時、2N個のページに達するログ情報が発生するようになる。ところで、一度生成されたログ情報は、すべてトランザクションが完了する以前にディスクに記録されなければならないため、ディスクの入出力過程でログ情報管理システムの性能を低下させられる。特に大量のレコード挿入作業を迅速に処理しなければならないシステムでは、大きな問題ともなり得る。

30

【0032】

したがって、新規に割り当てられたデータページに関して、発生するログ情報の大きさを最小化することによって、大量のレコード挿入または新しいインデックス生成作業が発生してもシステムの性能が低下することを防止することができるようになっている。

40

【特許文献1】韓国公開特許第10-2002-0030223号公報

【発明の開示】

【発明が解決しようとする課題】

【0033】

本発明は、トランザクション発生時の新規割当ページに対するログ情報を最小化することによって大量のレコード挿入またはインデックス生成作業に対する応答時間を短縮して、ログ情報管理システムの性能を改善することを目的とする。

【0034】

本発明の目的は、以上で言及した目的に制限されず、言及していないまた他の目的は下

50

記記載によって当業者が明確に理解できるものである。

【課題を解決するための手段】

【0035】

前記目的を達成するために、本発明の一態様に係るログ情報管理システムは、更新されるページが新しく割り当てられるページであるか否かを判断するインデックス/レコード/カタログ管理モジュールと、フリーページリストと未完了フリーページリストとを管理して、前記更新されるページが新しく割り当てられるページである場合、前記フリーページリストまたは前記未完了フリーページリストから新規ページを割り当てる格納管理モジュールと、前記新規ページに対応するバッファページに対する識別情報を設定するバッファ管理モジュールとを含み、前記バッファページに対するデータ値変更についてはログ情報が発生しない。

10

【0036】

また、前記目的を達成するために、本発明の他の態様に係るログ情報管理方法は、更新されるページが新しく割り当てられるページであるか否かを判断する(a)ステップと、前記更新されるページが新しく割り当てられるページである場合、フリーページリストまたは未完了フリーページリストから新規ページを割り当てる(b)ステップと、前記新規ページに対応するバッファページに対する識別情報を設定する(c)ステップとを含み、前記バッファページに対するデータ値変更についてはログ情報が発生しない。

【発明の効果】

【0037】

本発明に従えば、ログ情報管理システムにおいて大量で新規ページが割り当てられても割り当てられたページに対するログ情報を最小化することができるため、ディスクの書き込みコストを減少させられる効果がある。

20

【0038】

また、本発明に従えば、新しく割り当てられたページをディスクから読み込まないため、ディスクの読み込み費用コストを減少させられる効果がある。

【0039】

また、本発明に従えば、新しく割り当てられたページに関するログ情報が最小化されるため、データを復旧する場合にも応答速度が速くなる効果がある。

【発明を実施するための最良の形態】

30

【0040】

その他、実施形態の具体的な事項は、詳細な説明および図面に含まれている。

【0041】

本発明の利点および特徴、そしてそれらを達成する方法は、添付する図面と共に詳細に後述する実施形態を参照すれば明確になる。しかし、本発明は以下にて開示する実施形態に限定されず、互いに異なる多様な形態によって実施され、単に本実施形態は本発明の開示が完全なものとなるようにし、本発明が属する技術分野で通常の知識を有する者に発明の範疇を完全に知らせるために提供されるものであって、本発明は請求項の範疇によってのみ定義されるものである。

【0042】

40

以下、本発明の実施形態によって、ログ情報管理システムおよび方法を説明するためのブロック図または処理フローチャートに対する図面を参考にして本発明について説明する。この時、処理フローチャートの各ブロックとフローチャートの組合せは、コンピュータプログラムインストラクションによって行われることが理解できる。これらコンピュータプログラムインストラクションは、汎用コンピュータ、特殊用コンピュータ、またはその他のプログラム可能なデータプロセッシング装備のプロセッサに搭載することができるため、コンピュータまたはその他のプログラム可能なデータプロセッシング装備のプロセッサを介して行われるそのインストラクションがフローチャートブロックで説明する機能を行う手段を生成するようになる。これらコンピュータプログラムインストラクションは、特定方式で機能を具現するために、コンピュータまたはその他のプログラム可能なデータプロ

50

セシング装置が備えられるコンピュータによって利用可能または読取可能なメモリに格納されることも可能であるため、そのコンピュータによって利用可能または読取可能なメモリに格納されたインストラクションは、フローチャートブロックで説明する機能を行うインストラクション手段を含む製造品目の生産に用いることも可能である。コンピュータプログラムインストラクションは、コンピュータまたはその他のプログラム可能なデータプロセシング装置上に搭載することも可能であるため、コンピュータまたはその他のプログラム可能なデータプロセシング装置上で一連の動作ステップが行われ、コンピュータで実行されるプロセスを生成して、コンピュータまたはその他のプログラム可能なデータプロセシング装置の機能を行うインストラクションはフローチャートブロックで説明する各ステップの機能を実行することも可能である。

10

【0043】

また、各ブロックは、特定された論理的機能を実行するための1つ以上の実行可能なインストラクションを含むモジュール、セグメント、またはコードの一部を示すことができる。また、複数の代替実行例ではブロックで言及された機能が順序から外れて行われることも可能であることに注目しなければならない。例えば、連続して図示されている2つのブロックは、実質的に同時に行われることも可能で、またはそのブロックが時には該当機能により逆順に行われることも可能である。

【0044】

図3は、本発明の一実施形態に係るログ情報管理システムの構成を示すブロック図である。

20

【0045】

図3を参照すれば、ログ情報管理システム300は、トランザクション管理モジュール310、復旧管理モジュール320、ログ管理モジュール330、インデックス/レコード/カタログ管理モジュール340、バッファ管理モジュール350、バッファモジュール360、格納管理モジュール370、ディスクモジュール380、およびアプリケーション390を含む。ここで、バッファモジュール360はデータバッファ362とログバッファ364を含み、格納管理モジュール370はページ使用管理モジュール372を含む。

【0046】

この時、本実施形態で用いられる「モジュール」という用語は、ソフトウェアまたはFPGAまたはASICのようなハードウェア構成要素を意味し、モジュールはある役割を行う。ところが、モジュールは、ソフトウェアまたはハードウェアに限定される意味ではない。モジュールは、アドレッシングできる格納媒体にあるように構成することもでき、1つまたはそれ以上のプロセッサを再生させるように構成することもできる。したがって、一例としてモジュールはソフトウェア構成要素、オブジェクト指向ソフトウェア構成要素、クラス構成要素およびタスク構成要素のような構成要素と、プロセス、関数、属性、プロシージャ、サブルーチン、プログラムコードのセグメント、ドライバ、ファームウェア、マイクロコード、回路、データ、データベース、データ構造、テーブル、アレイ、および変数を含む。構成要素とモジュールの中で提供されている機能は、さらに小さい数の構成要素およびモジュールに結合したり、追加的な構成要素とモジュールにさらに分離したりすることができる。それだけでなく、構成要素およびモジュールはデバイスまたは保安マルチメディアカード内の1つまたはそれ以上のCPUを再生させるように具現されることもできる。

30

40

【0047】

アプリケーション390は、クエリープロセッサまたはクエリーエンジンとして動作することができる。アプリケーション390は、トランザクション管理モジュール310にトランザクションの開始、取り消しまたは終了を知らせることができ、トランザクションの境界内、即ちトランザクションの開始とトランザクション終了/取り消しの間でインデックス/レコード/カタログ管理モジュール340を介してログ情報を読み込んだり変更したりすることができる。

50

【 0 0 4 8 】

トランザクション管理モジュール 3 1 0 は、トランザクションの開始、終了、取り消し時にログ管理モジュール 3 3 0 を用いてログ情報を残し、現在進行しているトランザクションの情報を維持保持する。

【 0 0 4 9 】

復旧管理モジュール 3 2 0 は、特定トランザクションの取り消し時にログ管理モジュール 3 3 0 を介してログ情報を確保した後、データに行なわれた変更を戻す役割を担当する。特に、復旧管理モジュール 3 2 0 は、ログ情報管理システム 3 0 0 が非正常的に終了した時に未完了であったトランザクションの影響をデータベースから削除し、完了したトランザクションの修正事項をデータベースに反映して、データの一貫性を維持保持するようにする。

10

【 0 0 5 0 】

ログ管理モジュール 3 3 0 は、他のモジュールで発生させたデータベースに対する変更したログ情報を取り集めてバッファモジュール 3 6 0 に反映し、復旧管理モジュール 3 2 0 から要請されたログ情報をディスクモジュール 3 8 0 から読み込んで提供する。

【 0 0 5 1 】

インデックス/レコード/カタログ管理モジュール 3 4 0 は、ログ情報管理システム 3 0 0 を構成する主要データのインデックス、レコード、カタログ情報を管理するモジュールであって、バッファ管理モジュール 3 5 0 に必要なデータ(インデックス、レコード、カタログ)をバッファモジュール 3 6 0 にロードさせて必要な値を読み込んだり変更したりするように要請する。ログ情報管理システム 3 0 0 において、大部分のログ情報はインデックス/レコード/カタログ管理モジュール 3 4 0 によって発生する。

20

【 0 0 5 2 】

また、インデックス/レコード/カタログ管理モジュール 3 4 0 は、データ値に対する更新が発生する場合、ログ情報を残す前に更新が発生したページが新しく割り当てられたページであるかを確認する。データ値に対する更新が発生したページが該当トランザクション内で新しく割り当てられたページとして判定されれば、インデックス/レコード/カタログ管理モジュール 3 4 0 はデータ値が更新されたとしても、これに対するログ情報を残さない。

【 0 0 5 3 】

バッファ管理モジュール 3 5 0 は、バッファモジュール 3 6 0 を管理しながら、他のモジュールの要請によりディスクモジュール 3 8 0 に格納されたインデックス、レコードまたはカタログに対するログ情報を入れたページをバッファモジュール 3 6 0 のデータバッファ 3 6 2 にロードさせたり、データバッファ 3 6 2 にロードされたページをディスクモジュール 3 8 0 に格納したりする。ディスクモジュール 3 8 0 からデータページを読み込んだり記録したりする作業は、格納管理モジュール 3 7 0 によって行われる。

30

【 0 0 5 4 】

また、バッファ管理モジュール 3 5 0 は、ログ管理モジュール 3 3 0 の要請によってページ L S N 値として M A X _ L S N 値を有するデータページのみをデータバッファ 3 6 2 から選択し、ディスクモジュール 3 8 0 にすべて記録する。ページ L S N は、各ページを識別する識別情報として理解され得る。

40

【 0 0 5 5 】

バッファモジュール 3 6 0 は、メモリの一部としてログ情報管理システム 3 0 0 が確保して専用する空間である。バッファモジュール 3 6 0 は、ログ情報が格納されるログバッファ 3 6 4 とページが格納されるデータバッファ 3 6 2 とを含む。

【 0 0 5 6 】

格納管理モジュール 3 7 0 は、バッファ管理モジュール 3 5 0 の要請によってディスクモジュール 3 8 0 の特定ページを読み込み、バッファモジュール 3 6 0 にロードしたり、バッファモジュール 3 6 0 の特定ページをディスクモジュール 3 8 0 に記録したりする。また、格納管理モジュール 3 7 0 は新しいページが要請される時、該当時点で用いられて

50

いないディスクページをデータ格納用として割り当て、既存のディスクページをそれ以上データ格納用として用いない場合には、該当ディスクページを空のディスクページのうちの1つとして管理する。格納管理モジュール370は、ページ使用現況を管理する役割をするページ使用管理モジュール372を含む。

【0057】

一方、格納管理モジュール370は、新しく割り当てられたページをバッファモジュール360のデータバッファ362に搬入する時、該当ページをフリーページリストから割り当てられた場合、ディスクモジュール380からページを読み込む代わりに、データバッファ362にある指定されたバッファページのページLSNを変更する作業を行う。この時、フリーページとは、データが新しく割当可能なページを意味し、フリーページリストはこのようなフリーページを管理するデータ構造としてページ使用管理モジュール372によって管理される。

10

【0058】

また、データバッファ362内の任意のページをディスクモジュール380に搬出する時には指定されたバッファページのページLSNを任意の値に変更した後に記録する。このようなディスク入出力に関連した格納管理モジュール370の動作は、図7と図8で具体的に説明することにする。

【0059】

格納管理モジュール370に含まれたページ使用管理モジュール372はフリーページリストと未完了ノンコミテッドフリーページリストとを管理する。未完了フリーページリストは、まだ完了しない任意のトランザクション内で削除されたページの目録である。ディスクページに対する新規割当が要請されれば、ページ使用管理モジュール372は、先ずフリーページリストで使用可能なフリーページを探して割り当てる。フリーページリストに使用可能なフリーページがない時には、未完了フリーページリストからフリーページを探して割り当てるようになる。フリーページと未完了フリーページを更新する方法については、図9で具体的に説明することにする。

20

【0060】

図4は、本発明の一実施形態に係る更新トランザクション過程を示すフローチャートである。

【0061】

まず、アプリケーション390がトランザクションの開始をトランザクション管理モジュール310に要請すれば、トランザクション管理モジュール310は新しい1つのトランザクションを生成し、生成されたトランザクションに対する情報をトランザクションが終了するまで維持保持する(S405)。

30

【0062】

そして、アプリケーション390がインデックス/レコード/カタログ管理モジュール340にデータ更新作業を要請すれば、インデックス/レコード/カタログ管理モジュール340は必要なページをバッファ管理モジュール350に要請して、バッファモジュール360のデータバッファ362内に確保した後に必要な更新作業を行うようになる(S410)。この時、更新が発生したページのうち新しく割り当てられるデータページについてはログ情報が発生しない。一方、データバッファ364にロードされるページは各ページ内部にページLSN値を保有するようになるが、本発明では新しく割り当てられるページについてはこのページLSN値を特定の値(例えば、MAX_LSN値)に設定することによって不要なログ情報を生成しない。即ち、インデックス/レコード/カタログ管理モジュール340は、更新が発生したページのページLSN値がMAX_LSNである場合、ログ情報を生成しない。この時、「MAX_LSN」は該当ログ情報管理システムで生成可能な最も大きいLSN値を示す。

40

【0063】

データ更新作業に対するより具体的な過程は、図5で説明することにする。

【0064】

50

アプリケーション 390 は、データ更新作業がエラーなしに行われたと判断した場合、トランザクション管理モジュール 310 にトランザクションを終了するようにし、そうではない場合にはトランザクションを中断するように要請する (S 415)。

【0065】

アプリケーション 390 が更新を承認する場合 (S 415)、該当トランザクションによって削除され、未完了フリーページリストに入っていたページ情報をフリーページリストに反映する (S 420)。本発明では従来の方式とは異なり、あるトランザクション内で削除されたページを直ちにフリーページリストに入れずに、別の未完了フリーページリストで管理する。

【0066】

そして、データバッファ 362 に入っているデータページのうち、未だディスクモジュール 380 に記録されていない新規割当ページ (即ち、ページ LSN が MAX_LSN であるページ) をディスクモジュール 380 に記録する (S 425)。本発明では、新規割当ページに対してログ情報を残さない代わりに、トランザクションが完了する前にディスクモジュール 380 に新規割当ページを記録することによって、トランザクション完了後に発生し得る正常ではない終了においてもデータの一貫性が維持保持されるようになる。

【0067】

一方、新規割当ページをディスクモジュール 380 にページ LSN 値が MAX_LSN の状態で記録されるようになれば、他のトランザクション内で該当ページに対する更新を試みる時に、ページ LSN を新規に割り当てられたページとして誤認し得る。したがって、格納管理モジュール 370 はページ LSN が MAX_LSN の新規割当ページをディスクモジュール 380 に記録する時には、ページ LSN を任意の他の値に変えた後に記録することが好ましい。

【0068】

S 425 のステップが行われた後、トランザクション管理モジュール 310 はログ管理モジュール 330 に「Commit_Log」を生成するように指示する (S 430)。そして、前記 Commit_Log の LSN 値より小さいか同じである LSN 値を有するすべてのログ情報がログバッファ 364 内にある場合、ログ管理モジュール 330 はバッファ管理モジュール 350 に前記ログ情報をディスクモジュール 380 のログファイルに記録するように要請する (S 435)。

【0069】

アプリケーション 390 がデータ更新を承認しない場合、即ち、トランザクションを中断すると決定した場合には、トランザクション管理モジュール 310 は復旧管理モジュール 320 にトランザクションの取り消しを要請する。復旧管理モジュール 320 は、該当トランザクション内で既に変更したデータ値を復旧するために、以前のデータ値を有するログ情報をログ管理モジュール 330 に要請する (S 445)。前記要請は、バッファ管理モジュール 350 と格納管理モジュール 370 に順に伝えられて処理される。

【0070】

そして、復旧管理モジュール 320 は、ログ管理モジュール 330 に要請したログ情報を参照し、変更が発生した部分を探して以前のデータ値に復元する (S 450)。

【0071】

トランザクションが終了するか取り消された後、トランザクション管理モジュール 310 は、該当トランザクションの情報をトランザクションテーブルから削除する (S 440)。

【0072】

図 5 は、本発明の一実施形態によりデータを更新する過程を具体的に示すフローチャートである。

【0073】

まず、アプリケーション 390 は、インデックス/レコード/カタログ管理モジュール 340 にデータ更新を要請する (S 505)。そして、インデックス/レコード/カタロ

10

20

30

40

50

グ管理モジュール340は、前記要請された更新がディスクページを削除する更新であると判断した場合には(S510)、バッファ管理モジュール350が該当するディスクページがデータバッファ362に存在する場合に該当ページを削除し(S515)、格納管理モジュール370のページ使用管理モジュール372は、削除されたページ番号を未完了フリーページリストに挿入した後、このページを削除させたトランザクションが完了するか中断される前まで維持保持するようになる。このページを削除させたトランザクションが完了する場合、完了したトランザクションによって削除されたすべてのページは未完了フリーページリストから削除され、フリーページリストに挿入される。そして、トランザクションが中断される場合には、中断されたトランザクションによって削除されたすべてのページは未完了フリーページリストから削除され、フリーページリストには挿入されない。このような方法によって、要請されたディスクページが削除される(S520)。

10

【0074】

格納管理モジュール370は、ページ変更に対するログ情報を生成してログ管理モジュール330に伝送するようになる。

【0075】

一方、S510のステップで要請された更新がディスクページを削除する更新ではなく、新しいページまたは既存ページに対するデータ値の変更であるとインデックス/レコード/カタログ管理モジュール340によって判断された場合には(S510)、バッファ管理モジュール350は更新されるページがバッファモジュール360に既に存在するか否かを検査する(S530)。更新されるページがバッファモジュール360に既に存在する場合にはページを更新するが(S560)、そうではない場合にはバッファ管理モジュール350はバッファモジュール360に使用可能なバッファ空間があるか否かを検査する(S535)。

20

【0076】

S535のステップで使用可能なバッファ空間がない場合、ディスクモジュール380に記録するページが選択され、選択されたページに対するページLSNが処理されるが、このような動作を具体的に説明することにする(S540)。

【0077】

まず、データバッファ362から選択された特定のページに対するページLSNがMAX_LSNである場合、これは選択されたページがフリーページリストにあってトランザクション内で新しく割り当てられたことを意味する。したがって、この場合、格納管理モジュール370は選択されたページをディスクモジュール380に記録する前に、先ずページLSN値を既に設定された任意のLSN値に変更する。この時、好ましくは変更されたLSN値は、システム内で発生した最も大きいLSN値である。データバッファ362から選択された特定のページに対するページLSNがMAX_LSNでない場合に、格納管理モジュール370は選択されたページのページLSNを変更させずにディスクモジュール380に記録する。

30

【0078】

上記のような方法によって、バッファモジュール360に使用可能なバッファ空間が備えられれば、バッファ管理モジュール350は要請されたページが新しく割り当てられなければならないページであるかを検査する(S545)。

40

【0079】

要請されたページが新しく割り当てられなければならないページでない場合には、格納管理モジュール372は要請されたディスクページをバッファ管理モジュール350が指定したデータバッファ362にコピーし(S555)、インデックス/レコード/カタログ管理モジュール340は該当ページに対するデータ値の変更を行う(S560)。

【0080】

S545のステップで要請されたページが新しく割り当てられなければならないページである場合には、新しく割り当てられるページに対する処理動作が行われるが(S550)、以下にて具体的に説明することにする。

50

【 0 0 8 1 】

まず、ページ使用管理モジュール 3 7 2 は、フリーページリスト内に利用可能なフリーページがある場合にはフリーページリスト内で新規ページを割り当て、そうではない場合には未完了フリーページリストからページを割り当てる。

【 0 0 8 2 】

この時、新しく割り当てられたページがフリーページリストに割り当てられた場合、該当ディスクページがデータバッファ 3 6 2 にロードされ、再びディスクモジュール 3 8 0 に記録される前までは該当ページに対するすべてのデータ値変更についてはログ情報を発生させない。そして、格納管理モジュール 3 7 0 は、ディスクモジュール 3 8 0 で該当ディスクページを読み込まず、バッファ管理モジュール 3 5 0 は指定されたバッファページ 10
に対するページ L S N 値をこのページがフリーページリストで新規割り当てられたページであることを意味する M A X _ L S N 値として設定する。

【 0 0 8 3 】

フリーページリスト内にフリーページが 1 つもなく、新規割り当てられたページが未完了フリーページリストから割り当てられた場合には、ディスクモジュール 3 8 0 から該当ページを指定されたバッファ空間に読み込む。

【 0 0 8 4 】

S 5 6 0 のステップでページが更新されれば、更新されたページに対するログ情報が処理される (S 5 6 5) 。

【 0 0 8 5 】

従来技術ではあるページに対するデータ値が更新される毎に必ずログ情報が生成されたが、本発明ではインデックス/レコード/カタログ管理モジュール 3 4 0 が S 5 6 0 のステップでデータ値を更新した後、該当ページに対するログ情報生成が必要であるか否かを判断する。より具体的にインデックス/レコード/カタログ管理モジュール 3 4 0 は、バッファモジュール 3 6 0 に存在する該当ページのページ L S N 値が M A X _ L S N である場合には、ログ情報を生成しなくなる。 20

【 0 0 8 6 】

図 6 は、本発明の一実施形態によりディスクページを新しく割り当てる過程を示すフローチャートである。

【 0 0 8 7 】

図 6 を参照すれば、ページ使用管理モジュール 3 7 2 は、フリーページリストが空いているか否かを検査し (S 6 1 0) 、空いている場合には、未完了フリーページリストから一番目のページを選択して割り当てる (S 6 2 0) 。空いていない場合には、フリーページリストから一番目のページを選択して割り当てる (S 6 3 0) 。

【 0 0 8 8 】

図 7 は、本発明の一実施形態により新しく割り当てられたディスクページをバッファに搬入する過程を示すフローチャートである。

【 0 0 8 9 】

格納管理モジュール 3 7 0 は、要請を受けたページが新しく割り当てられるページであるかを検査し (S 7 1 0) 、新しく割り当てられるページである場合には、格納管理モジュール 3 7 0 は新しく割り当てられたページがフリーページリストに割り当てられたものか検査し (S 7 2 0) 、フリーページリストに割り当てられたページである場合には、データバッファ 3 6 2 の該当空間 (即ち、ディスクページをロードするバッファモジュール 3 6 0 内のバッファページ) のページ L S N を M A X _ L S N に設定することによって (S 7 3 0) 、新しく割り当てられるディスクページをバッファモジュール 3 6 0 に搬入する過程が終了する。 40

【 0 0 9 0 】

しかし、S 7 1 0 のステップで新しく割り当てられるページではないか、S 7 2 0 のステップで新しく割り当てられたページがフリーページリストに割り当てられたものでない場合 (例えば、新しく割り当てられたページが未完了フリーページリストで割り当てられ 50

た新規ページの場合)には、格納管理モジュール370は該当ディスクページをディスクモジュール380から読み込んだ後、バッファモジュール360のデータバッファ362にロードする。

【0091】

図7のような動作を行うために、格納管理モジュール370はバッファモジュール360を管理するバッファ管理モジュール350と通信を行うようになる。

【0092】

図8は、本発明の一実施形態によりバッファ内のページをディスクに搬出する過程を示すフローチャートである。

【0093】

図8を参照すれば、格納管理モジュール370は、ディスクモジュール380に記録するページのページLSNがMAX__LSNであるかを検査する(S810)。この時、MAX__LSNの場合には、ページLSNを既に設定されたLSNに変更し(S820)、バッファモジュール360のデータバッファ362内で要請されたバッファページをディスクモジュール380に記録する(S830)。S810のステップでMAX__LSNでない場合には、バッファモジュール360のデータバッファ362内で要請されたバッファページをディスクモジュール380に記録する(S830)。

【0094】

図8のような動作を行うために、格納管理モジュール370はバッファモジュール360を管理するバッファ管理モジュール350と通信を行うようになる。

【0095】

図9は、本発明の一実施形態によりフリーページリストと未完了フリーページリストとの更新過程を示す例示図であって、このような更新過程は格納管理モジュール370のページ使用管理モジュール372によって行われる。

【0096】

まず最初に、フリーページリストに5個のフリーページがあり、未完了フリーページリストは空いていることを示している(S910)。未完了フリーページリストが空いているとは、現在完了していないトランザクションから削除したディスクページが無いことを意味する。

【0097】

ページ#1とページ#3とが削除されれば、ページ#1とページ#3とは未完了フリーページリストに入り、フリーページリストには変化がない(S920)。

【0098】

次に、5個のディスクページが割り当てられれば、先ずフリーページリストにあるページから割り当てられるため、フリーページリストは空くようになり、未完了フリーページリストには変化がない(S930)。

【0099】

次に、1個のページがさらに割り当てられれば、今は未完了フリーページリストからページ#3を割り当てる。その結果、未完了フリーページリストにはページ#1だけがフリーページとして残る(S940)。

【0100】

最後に、ページ#1とページ#3とを削除したトランザクションが完了した後は、未完了フリーページリストのフリーページがフリーページリストに併合される。その結果、フリーページリストは、未完了フリーページリストにあったページ#1を有するようになり、未完了フリーページリストは空いているようになる(S950)。

【0101】

図10は、本発明の一実施形態に係る実験結果を示すグラフであって、3000個のレコードを新しく挿入した時の処理速度を比較している。

【0102】

この時、1つのレコードは、11個のフィールドから構成され、11個のフィールドの

10

20

30

40

50

うち9個のフィールドをインデックスとし、1つのレコードの大きさは210バイトとした。

【0103】

図10を参照すれば、従来技術に係わる場合には、3000個のレコードを新しく挿入する場合に約25秒を必要としたが、本発明に係わる場合には約10秒を必要とし、挿入性能が約2.5倍向上したことが分かる。また、本発明によりログ情報の量が減少したので復旧時所要時間も約1/50で減少した。

【0104】

以上、添付した図面を参照して本発明の実施形態を説明したが、本発明が属する技術分野で通常の知識を有する者であれば、本発明がその技術的思想や必須の特徴を変更せず、他の具体的な形態によって実施することができるということを理解できる。したがって、以上で記述した実施形態はすべての面で例示的なものであり、限定的なものではないことを理解しなければならない。

【図面の簡単な説明】

【0105】

【図1】従来技術に係るログ情報管理システムの構成を示すブロック図である。

【図2】従来技術に係る一般的なログレコードの構造を示す例示図である。

【図3】本発明の一実施形態に係るログ情報管理システムの構成を示すブロック図である。

。

【図4】本発明の一実施形態に係る更新トランザクション過程を示すフローチャートである。

【図5】本発明の一実施形態によりデータを更新する過程を具体的に示すフローチャートである。

【図6】本発明の一実施形態によりディスクページを新しく割り当てる過程を示すフローチャートである。

【図7】本発明の一実施形態により新しく割り当てられたディスクページをバッファに搬入する過程を示すフローチャートである。

【図8】本発明の一実施形態によりバッファ内のページをディスクに搬出する過程を示すフローチャートである。

【図9】本発明の一実施形態によりフリーページリストと未完了フリーページリストとの更新過程を示す例示図である。

【図10】本発明の一実施形態に係る実験結果を示すグラフである。

【符号の説明】

【0106】

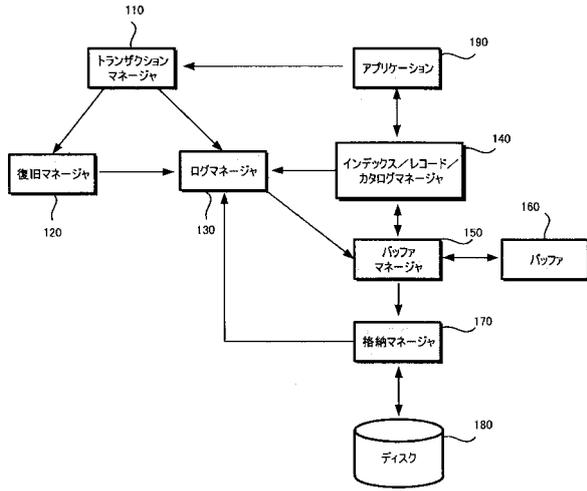
300	ログ情報管理システム	
310	トランザクション管理モジュール	
320	復旧管理モジュール	
330	ログ管理モジュール	
340	インデックス/レコード/カタログ管理モジュール	
350	バッファ管理モジュール	40
370	格納管理モジュール	
390	アプリケーション	

10

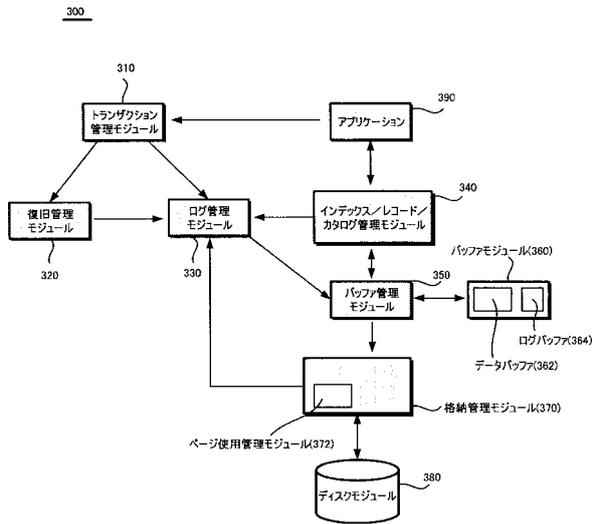
20

30

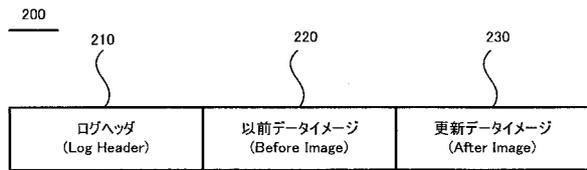
【図1】



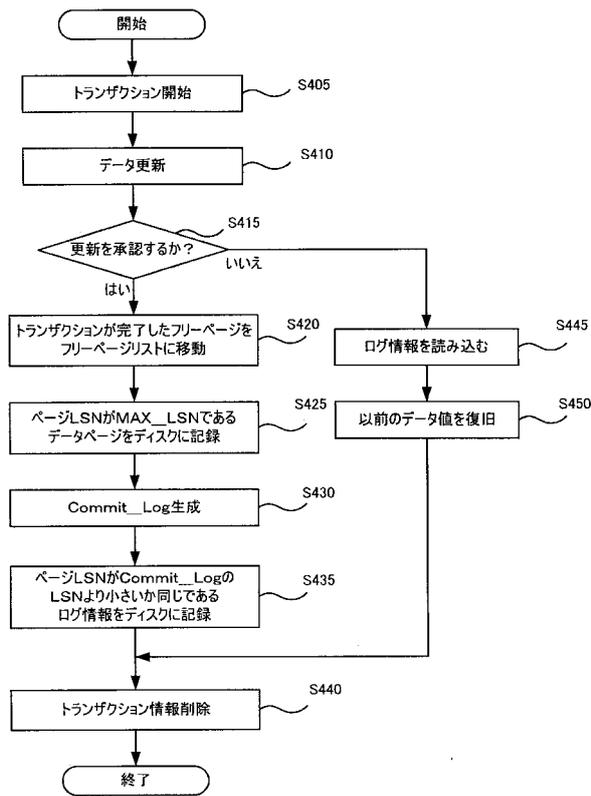
【図3】



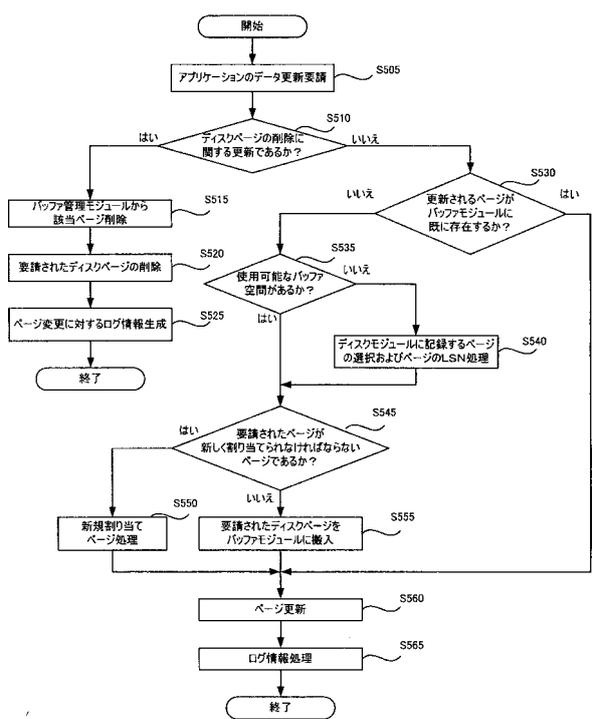
【図2】



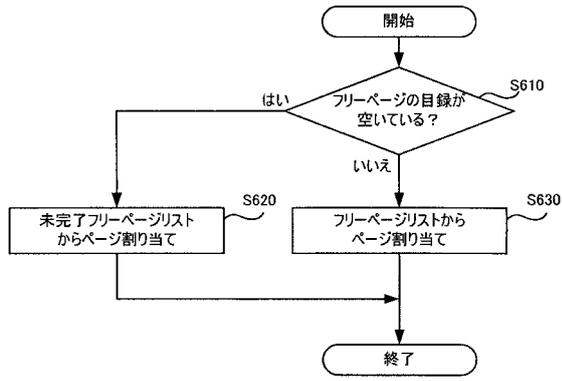
【図4】



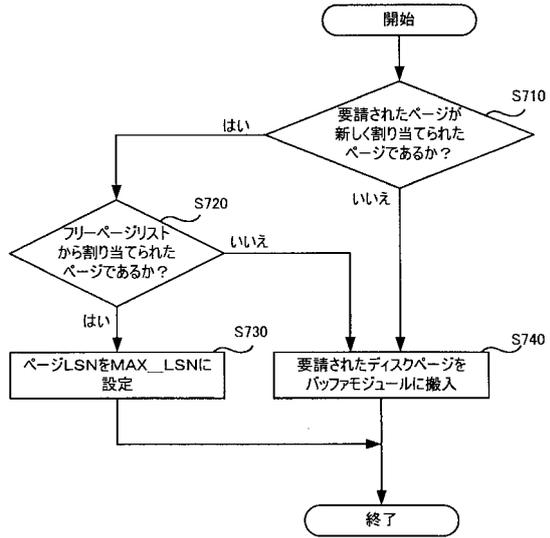
【図5】



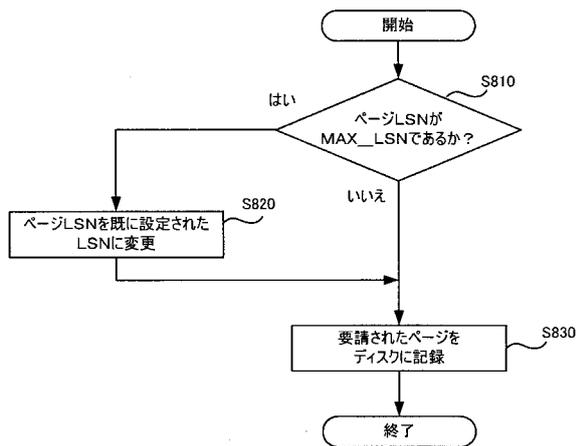
【図6】



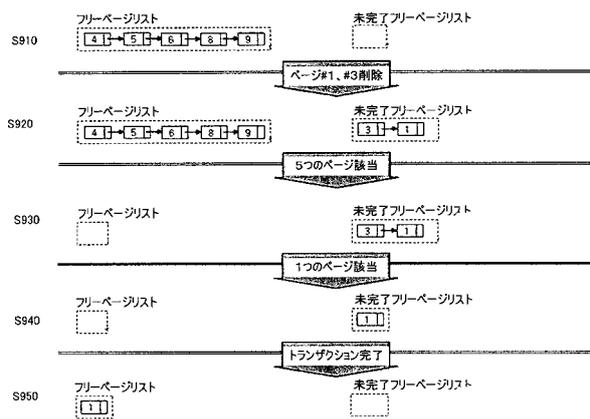
【図7】



【図8】



【図9】



【図10】

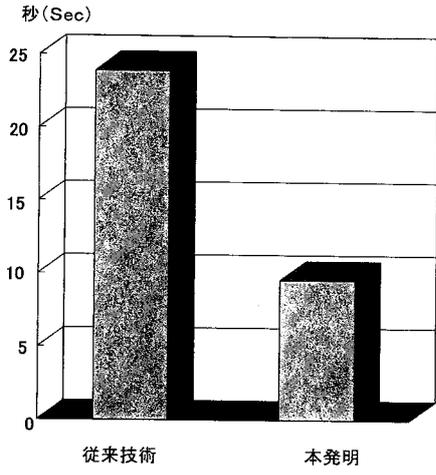


図 3000個のレコード挿入

フロントページの続き

- (72)発明者 禹 景 久
大韓民国ソウル特別市瑞草区瑞草洞1688-6番地 瑞草メイフラワーメンバースヴィル803号
- (72)発明者 姜 仁 鮮
大韓民国ソウル特別市瑞草区瑞草1洞1679-1番地 コーアプレジダンス303号
- (72)発明者 閔 庚 燮
大韓民国ソウル特別市西大門区弘濟1洞 現代アパート104棟903号(番地なし)
- (72)発明者 金 映 錫
大韓民国ソウル特別市江南区論 ひょん 洞22番地 論 ひょん アパート104棟909号

合議体

審判長 赤川 誠一
審判官 清木 泰
審判官 石井 茂和

- (56)参考文献 特開平4-148250(JP,A)
米国特許第6571259(US,B1)

- (58)調査した分野(Int.Cl., DB名)
G06F 12/00