(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113056734 A
(43)申请公布日 2021.06.29

(21)申请号 201980075736.6
(22)申请日 2019.11.13
(30)优先权数据
16/192,514 2018.11.15 US
(85)PCT国际申请进入国家阶段日
2021.05.14
(86)PCT国际申请的申请数据
PCT/CN2019/117927 2019.11.13
(87)PCT国际申请的公布数据
WO2020/098682 EN 2020.05.22
(71)申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72)发明人 薛洵 陈冲 佩尔-阿克·拉尔森 罗宾·格罗斯曼
(51)Int.Cl.
G06F 16/23(2006.01)

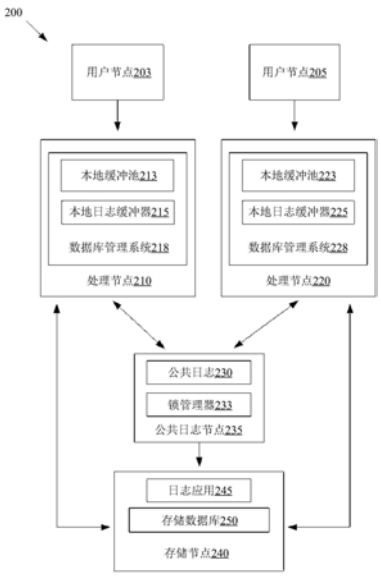
权利要求书3页 说明书15页 附图9页

(54)发明名称

管理共享数据库的系统和方法

(57)摘要

描述了用于管理共享数据库的方法和系统。一个或多个处理节点可以访问共享数据库。公共日志节点可以管理所述共享数据库。所述公共日志节点可以验证所述一个或多个处理节点所请求的数据库操作。在验证期间，所述公共日志节点可以检测所述一个或多个处理节点所请求的所述数据库操作之间发生的冲突。



CN 113056734 A

1.一种方法,其特征在于,包括:

第一计算设备接收对数据库的页面的记录执行修改操作的请求,其中,所述请求包括:

所述页面的标识,

与所述页面对应的基础版本号,以及

与所述修改操作对应的事务的标识;

所述第一计算设备确定是否已分配与所述页面对应的锁;

在确定已分配所述锁之后,所述第一计算设备确定所述锁是否分配给了第二计算设备;

所述第一计算设备将所述基础版本号与所述页面的最新验证版本号进行比较;以及

在确定所述锁已分配给了所述第二计算设备,并且所述基础版本号等同于所述最新验证版本号之后,所述第一计算设备发送所述修改操作已通过验证的指示。

2.根据权利要求1所述的方法,其特征在于,还包括:所述第一计算设备将所述事务的所述标识添加到与所述锁对应的事务列表中。

3.根据权利要求2所述的方法,其特征在于,所述锁包括:

所述页面的所述标识;

所述事务列表;以及

所述第二计算设备的指示。

4.根据权利要求2或3所述的方法,其特征在于,还包括:

所述第一计算设备从所述第二计算设备接收已提交所述事务的指示;

所述第一计算设备从所述事务列表中删除所述事务;

所述第一计算设备确定所述事务列表是否为空;以及

所述第一计算设备在确定所述事务列表为空之后,释放所述锁。

5.根据权利要求1至4中任一项所述的方法,其特征在于,所述基础版本号包括先前修改所述页面的处理节点的指示。

6.根据权利要求1至5中任一项所述的方法,其特征在于,还包括:

所述第一计算设备从所述第二计算设备接收对所述数据库的第二页面的记录执行第二修改操作的第二请求,其中,所述第二请求包括所述第二页面的基础版本号;

所述第一计算设备将所述第二页面的所述基础版本号与所述第二页面的最新验证版本号进行比较;以及

在确定所述第二页面的所述基础版本号与所述第二页面的所述最新验证版本号不同之后,所述第一计算设备向所述第二计算设备发送所述第二修改操作未通过验证的指示。

7.根据权利要求1至6中任一项所述的方法,其特征在于,还包括:

所述第一计算设备从第三计算设备接收对所述页面的所述记录执行第二修改操作的请求;

所述第一计算设备确定与所述页面对应的所述锁已分配给了所述第二计算设备;以及

在确定所述锁已分配给了所述第二计算设备之后,所述第一计算设备向所述第三计算设备发送所述第二修改操作未通过验证的指示。

8.根据权利要求1至7中任一项所述的方法,其特征在于,还包括:所述第一计算设备基于执行所述修改操作的所述请求,向包括所述数据库的第三计算设备发送对所述页面的所

述记录的更新。

9.一种方法,其特征在于,包括:

第一计算设备接收对数据库的页面的记录执行操作的请求,其中,所述请求包括:

所述页面的标识,

与所述页面对应的基础版本号,以及

与所述操作对应的事务的标识;

确定与所述页面对应的锁是否已分配给了第二计算设备;

在确定所述锁已分配给了所述第二计算设备之后,将所述基础版本号与所述页面的最新验证版本号进行比较;以及

在确定所述基础版本号与所述最新验证版本号不同之后,发送所述操作未通过验证的指示。

10.根据权利要求9所述的方法,其特征在于,还包括:所述第一计算设备从所述第二计算设备接收用于撤销所述事务的指令。

11.根据权利要求10所述的方法,其特征在于,还包括:

确定与所述事务对应的多个锁;

确定与所述多个锁对应的多个页码;以及

对于所述多个页码中的每个页码,撤销与所述事务对应的页面更新。

12.一种方法,其特征在于,包括:

计算设备接收对数据库的页面的记录执行修改操作的请求,其中,所述请求包括:

所述页面的标识,

与所述页面对应的基础版本号,以及

与所述修改操作对应的事务的标识;

所述计算设备确定是否已分配与所述页面对应的锁;

所述计算设备创建与所述页面对应的所述锁;

所述计算设备将所述基础版本号与所述页面的最新验证版本号进行比较;以及

所述计算设备在确定所述基础版本号等同于所述最新验证版本号之后,发送所述修改操作已通过验证的指示。

13.根据权利要求12所述的方法,其特征在于,还包括:基于执行所述修改操作的所述请求,向第二计算设备发送对所述页面的所述记录的更新。

14.根据权利要求12或13所述的方法,其特征在于,对所述数据库的所述页面的所述记录执行所述修改操作的所述请求包括对所述页面的所述记录的所述更新。

15.根据权利要求12至14中任一项所述的方法,其特征在于,还包括:

从第二计算设备接收对所述数据库的第二页面的请求;

确定所述第二页面的最新验证版本号;

检索所述第二页面的所述最新验证版本号对应的数据;以及

向所述第二计算设备发送所述数据。

16.根据权利要求12至15中任一项所述的方法,其特征在于,还包括:

所述计算设备接收对所述数据库的第二页面的记录执行第二修改操作的第二请求,其中,所述第二请求包括所述第二页面的基础版本号;

将所述第二页面的所述基础版本号与所述第二页面的最新验证版本号进行比较;以及

在确定所述第二页面的所述基础版本号与所述第二页面的所述最新验证版本号不同后,发送所述第二修改操作未通过验证的指示。

17.根据权利要求12至16中任一项所述的方法,其特征在于,还包括:

从第二计算设备接收对所述页面的所述记录执行第二修改操作的请求;

确定与所述页面对应的所述锁已分配给第三计算设备;以及

在确定所述锁已分配给所述第三计算设备之后,发送所述第二修改操作未通过验证的指示。

18.根据权利要求12至17中任一项所述的方法,其特征在于,还包括:将所述事务的所述标识添加到与所述锁对应的事务列表中。

19.根据权利要求18所述的方法,其特征在于,所述锁包括:

所述页面的所述标识;

所述事务列表;以及

第二计算设备的指示。

20.根据权利要求19所述的方法,其特征在于,还包括:

接收已提交所述事务的指示;

从所述事务列表中删除所述事务;

确定所述事务列表是否为空;以及

在确定所述事务列表为空之后,释放所述锁。

管理共享数据库的系统和方法

[0001] 相关申请案交叉申请

[0002] 本申请要求于2018年11月15日提交的、发明名称为“管理共享数据库的系统和方法(SYSTEMS AND METHODS FOR MANAGING A SHARED DATABASE)”的美国专利申请第16/192,514号的优先权,其内容通过引用并入本文,如全文再现一般。

背景技术

[0003] 数据库系统通常配置为存储可由一个或多个处理节点随时访问的数据。根据操作和条件,可以允许处理节点读取和/或写入数据库系统的数据库。然而,由于各种原因,例如,通信访问问题、通信网络拥塞问题、基于远程位置的延迟等,处理节点在完成其对数据库的页面的记录执行操作的请求时往往会遇到显著的延迟。处理节点可以位于各种不同的位置,例如,位于处理节点和实现数据库系统的存储节点之间存在明显通信延迟的位置。因此,当多个处理节点尝试以并发方式写入数据库的同一部分时,经常会发生冲突。为了防止和/或解决尝试并发访问数据库系统的处理节点之间发生的冲突,提出了一种配置,所述配置能有效地最小化处理延迟、协调冲突并有效地执行对数据库系统的读/写访问以成功执行所需的操作。

发明内容

[0004] 以下发明内容仅用于说明目的,并不旨在限制或约束详细的说明。以下发明内容仅以简化的形式呈现了各种描述的方面,作为下文提供的更详细描述的序言。

[0005] 关系数据库等数据库可用于存储数据。数据库可以是在由一个或多个处理节点访问的存储节点上实现的共享数据库,每个处理节点具有数据库的读和/或写权限。中间系统,以下称为“公共日志节点”,可以协调处理节点和实现数据库的存储节点之间的事务。每个事务可以包括一个或多个请求,以对数据库的页面的记录执行修改操作。中间系统可以包括配置为协调事务的软件和/或硬件。公共日志节点可以从处理节点接收对数据库的页面的记录执行修改操作的请求。这些页面可以存储于在存储节点上实现的数据库中。公共日志节点可以验证操作和/或使操作发送到存储节点以在数据库上执行。

[0006] 当多个处理节点并发请求对数据库的同一页面的记录执行修改操作时,可能会发生冲突。公共日志可以检测和/或防止这些冲突。如果公共日志检测到冲突,则处理节点和/或公共日志可以撤销或取消与冲突对应的事务。公共日志节点和/或处理节点可以维护与数据库的页面对应的锁。每个锁可以分配给单个处理节点。分配到锁的处理节点具有对该锁对应页面的记录执行操作的独占权限,而其他处理节点不允许对该锁对应页面的记录执行修改操作。可以允许其他处理节点读取与该锁对应的页面。在持有锁的事务已提交或撤销(即,中止)后,可以将该锁释放。

[0007] 根据本公开的一个方面,提供了一种方法,其可以包括接收对数据库的页面的记录执行修改操作的请求。该请求可以由公共日志节点等第一计算设备接收。该请求可以从处理节点等第二计算设备接收。该请求可以包括页面的标识。所述请求可以包括与所述页

面相对应的基础版本号。所述请求可以包括与所述修改操作相对应的事务的标识。该方法可以包括:确定是否已分配与所述页面对应的锁。该方法可以包括:在确定已分配所述锁之后,确定所述锁是否分配给了第二计算设备。该方法可以包括:将所述基础版本号与所述页面的最新验证版本号进行比较。该方法可以包括:在确定所述锁已分配给了所述第二计算设备,并且确定所述基础版本号等同于所述页面的所述最新验证版本号之后,发送所述修改操作已通过验证的指示。

[0008] 根据本公开的另一方面,提供了一种方法,其可以包括接收对数据库的页面的记录执行修改操作的请求。所述请求可以由计算设备接收。所述请求可以包括页面的标识。所述请求可以包括与所述页面相对应的基础版本号。所述请求可以包括与所述修改操作相对应的事务的标识。该方法可以包括:确定是否已分配与所述页面对应的锁。该方法可以包括创建与所述页面对应的锁。该方法可以包括:将所述基础版本号与所述页面的最新验证版本号进行比较。该方法可以包括:在确定所述基础版本号等同于所述页面的所述最新验证版本号之后,发送所述修改操作已通过验证的指示。

[0009] 这里的发明内容不是对本文描述的新特征的穷举,也不是对权利要求的限制。下面将更详细地描述这些功能和其他功能。

附图说明

[0010] 关于以下描述、权利要求和附图,本公开的这些和其他特征、方面和优点将得到更好的理解。本公开通过示例的方式来说明,但不限于附图,在附图中,相似的数字表示相似的元件。

[0011] 图1为可用于实现本文描述的代表性实施例提供的设备和方法的计算系统的框图;

[0012] 图2为本公开的一个或多个说明性方面提供的包括数据库系统的节点的图;

[0013] 图3为本公开的一个或多个说明性方面提供的用于检测冲突的方法的流程图;

[0014] 图4为本公开的一个或多个说明性方面提供的页面更新的调用流程图;

[0015] 图5为本公开的一个或多个说明性方面提供的用于释放锁的方法的流程图;

[0016] 图6为本公开的一个或多个说明性方面提供的用于提交事务的方法的流程图;

[0017] 图7为本公开的一个或多个说明性方面提供的用于中止事务的方法的流程图;以及

[0018] 图8A和图8B为本公开的一个或多个说明性方面提供的用于处理操作的方法的流程图。

具体实施方式

[0019] 在以下对各种说明性实施例的描述中,参考形成本文的一部分的附图,这些附图中,通过说明的方式示出了可以实践本公开的各方面的各种实施例。可以理解的是,在不脱离本发明范围的情况下,可以利用其他实施例,并可以做出结构上或功能修改。

[0020] 关系数据库等数据库可用于存储数据和/或访问存储的数据。数据库可以使用结构化查询语言(structured query language,SQL)数据库管理系统等关系型数据库管理系统(relational database management system,RDBMS)来管理和访问。该数据库可以是共

享数据库,其中多个处理节点可以访问该数据库。处理节点可以包括只读处理节点、只写处理节点、读/写处理节点或其任意组合。处理节点可以位于与实现数据库的存储节点不同的距离处,这可能导致处理节点和/或数据库之间的通信延迟。数据库本身可以物理分布。例如,数据库的不同部分可以位于不同的存储节点上。数据库的某些部分可能比数据库的其他部分更接近处理节点。

[0021] 数据库将数据存储在页面中,其中每个页面包括存储数据的一条或多条记录。每个处理节点可以共享数据库的页面的单个视图。换句话说,无论数据库的页面是从第一处理节点还是第二处理节点访问的,从数据库返回的页面都是相同的。例如,返回给请求数据库的页面的第一处理节点的页面可以与返回给请求数据库的页面的第二处理节点的页面相同。

[0022] 处理节点可以发送对数据库的页面的记录执行各种修改操作的请求。修改操作可以包括插入记录操作(在页面中插入新记录)、删除记录操作(删除页面的记录)和更新记录操作(更新页面的记录)。这些请求可以是事务的一部分。每个事务可以包括一个或多个请求,其中每个请求是对数据库的页面的记录执行修改操作的请求。每个事务可以通过例如事务编号进行标记。每个事务可以源自单个处理节点。换句话说,对于单个事务,对数据库的页面的记录执行修改操作的每个请求都可以源自同一处理节点。

[0023] 可以在每个处理节点上维护本地日志缓冲区。本地日志缓冲区可以包括日志记录,并且每个日志记录可以包括将要在数据库的页面的记录上执行的修改操作的信息(例如,插入、删除或更改数据库的页面的记录的操作的信息)。日志缓冲区可以包括针对处理节点所请求的操作的重做和/或撤销信息。日志记录可以是复合的,其中每个日志记录可以包括对同一页面的记录的多次更改的信息和/或对不同页面的记录的多次更改的信息。例如在InnoDB数据库存储引擎中,复合日志记录可以称为迷你事务(mini transaction,MTR)。

[0024] 对数据库的页面的记录执行各种修改操作的请求可以由处理节点发送到与数据库通信的公共日志节点等中间系统。公共日志节点包括公共日志,公共日志是驻留或实现在公共日志节点上的软件系统。请求可以包括日志记录。换句话说,在一些实施例中,处理节点可以将日志记录发送到公共日志节点。公共日志可以从一个或多个处理节点接收日志记录。公共日志可以检查日志记录中的冲突操作,或者换句话说,公共日志可以验证日志记录。让公共日志执行此冲突检查可能比让单独的系统或单独的设备执行冲突检查更有效。公共日志可以验证日志记录,以确保两个处理节点不会同时请求对数据库同一页面的记录执行操作。公共日志在验证日志记录后,可以将日志记录发送到数据库,和/或根据日志记录向数据库发送指令。

[0025] 图1是可用于实现根据本文描述的代表性实施例的方法的计算设备100的框图。计算设备100可用于实现处理节点210或220、公共日志节点235和/或存储节点240,如下文和图2中所述。特定节点可以利用所有所示的组件或所述组件的仅子集,并且集成级别可能因节点而异。此外,节点可以包含组件的多个实例,例如多个处理单元、处理器、存储器、发送器、接收器、存储设备等。计算设备100可以是任何类型的合适计算设备,例如数据中心内的计算机或服务器。计算设备100可以包括中央处理器(central processing unit,CPU)114、总线120和/或存储器108,并且可选地还可以包括大容量存储设备104、视频适配器110,和/

或输入/输出(input/output,I/O)接口112(虚线所示)。本领域技术人员将理解,CPU114代表了处理能力。在一些实施例中,可以提供专门的处理核心来代替传统的CPU。例如,可以在CPU 114之外或代替CPU 114提供图形处理单元(Graphics Processing Unit,GPU)、张量处理单元(Tensor Processing Unit,TPU)和/或其他的加速处理器(或处理加速器)。

[0026] 所述CPU 114可包括任何类型的电子数据处理器。存储器108可包括任意类型的非瞬时性系统存储器,例如静态随机存取存储器(static random access memory,SRAM)、动态随机存取存储器(dynamic random access memory,DRAM)、同步DRAM(synchronous DRAM,SDRAM)、只读存储器(read-only memory,ROM)或它们的组合。例如,存储器108可以包括用于启动时的ROM和用于在执行程序时使用的程序和数据存储的DRAM。总线120可以是包括存储器总线或存储器控制器、外围总线和/或视频总线的任何类型的几种总线体系结构中的一个或多个。

[0027] 大容量存储设备104可以包括任何类型的非瞬时或非易失性存储设备,用于持久地存储数据、程序和其他信息,并使数据、程序和其他信息可通过总线120访问。大容量存储设备104可以包括例如固态驱动器、硬盘驱动器、磁盘驱动器和/或光盘驱动器中的一个或多个。

[0028] 视频适配器110和I/O接口112可以提供将外部输入和输出设备耦合到计算设备100的接口。输入和输出设备的示例包括耦合到视频适配器110的显示器118和I/O设备116,例如耦合到I/O接口112的触摸屏。其他设备可以耦合到计算设备100,并且可以使用额外的或更少的接口。例如,可使用如通用串行总线(Universal Serial Bus,USB)(未示出)等串行接口将接口提供给外部设备。

[0029] 计算设备100还可以包括一个或多个网络接口106,其可以包括以太网线等有线链路和/或用于访问一个或多个网络122的无线链路中的至少一个。网络接口106可以使得计算设备100通过网络122与其他计算设备100通信。举例来说,网络接口106可以经由一个或多个发射器/发射天线以及一个或多个接收器/接收天线提供无线通信。计算设备100可以与局域网或广域网通信,用于数据处理以及与其他处理单元、互联网或远程存储设施等远程设备通信。

[0030] 图2示出了根据本公开的一个或多个说明性方面的包括数据库系统200的节点的图。数据库系统200包括两个处理节点210、220、公共日志节点235和存储节点240。处理节点210和220可以是对存储节点240的存储数据库250具有只读访问、只写访问或读/写访问的处理节点。处理节点210和220可以包括SQL节点,或用于访问存储数据库250的任何其他类型的处理节点。处理节点210和220可以位于相同的位置、不同的位置和/或其组合。例如,处理节点210和220可以位于同一数据中心。在另一个示例中,处理节点210和220可以位于彼此相距数百公里的不同数据中心中。数据中心可以采用高速、高带宽的数据连接方式连接在一起。数据中心可以构成为客户提供数据库系统服务的大规模网络的一部分。每个处理节点210和220可以包括一个或多个计算设备100。虽然图2示出了包括两个示例处理节点210和220的数据库系统200,但在替代方面,数据库系统200可以包括任何数量的处理节点。

[0031] 处理节点210和220可以分别从用户节点203和205接收对数据库的记录执行操作的请求。用户节点203和205可以包括用户界面,所述用户界面由用户和/或软件用于生成对数据库的记录执行操作的请求。用户节点203和205可以向处理节点210和220发送执行读取

记录操作的请求以读取存储数据库250的页面的记录。用户节点203和205还可以向处理节点210和220发送对存储数据库250的页面执行修改操作的请求。虽然图2示出了两个用户节点203、205分别向处理节点210、220发送操作，但在替代方面，用户节点203、205可以同时向两个处理节点210、220发送请求。此外，在替代方面，数据库系统可以包括任意数量的用户节点，并且每个用户节点可以向单个处理节点或多个处理节点发送请求。

[0032] 处理节点210和220可以包括任何类型的数据库管理系统(即，RDBMS)218和228。例如，处理节点210和220可以包括Oracle数据库管理系统、MySQL数据库管理系统、Microsoft SQL Server管理系统、PostgreSQL数据库管理系统、IBM DB2数据库管理系统、Microsoft Access数据库管理系统、SQLite数据库管理系统、或任何其他类型的数据库管理系统。由处理节点210和220操作的数据库管理系统218和228的类型可以对应于正在使用的存储数据库250的类型。例如，如果存储数据库250是MySQL数据库，则数据库管理系统218和228可以包括MySQL RDBMS系统。

[0033] 数据库管理系统218和228可以更新存储数据库250。数据库管理系统218和228可以创建在发送到公共日志节点235之前临时存储在本地日志缓冲器215或225中的日志记录。公共日志节点235可以验证日志记录中包括的操作，并将日志记录转发到存储节点240。每个日志记录可以包括将要对存储数据库250的页面的记录执行的操作的信息。存储节点240的日志应用245可以将日志记录中包括的操作应用到存储数据库250。

[0034] 处理节点210和220可以将存储在存储数据库250中的页面的一部分存储在其各自的本地缓冲池213和223中。本地缓冲池213和223可以分别是数据库管理系统218和228的一部分。本地缓冲池213和223可以包括先前由处理节点210和220访问的页面。例如，如果处理节点210先前请求一个页面的最新验证版本，则该页面的最新验证版本可以存储在本地缓冲池213中。在请求对一个页面的记录进行操作之前，处理节点210、220的数据库管理系统218、228可以确定其是否具有该页面的最新验证版本。如果处理节点210、220的数据库管理系统218、228不具有页面，或者如果处理节点210或220具有的页面的版本不是页面的最新验证版本，则处理节点210、220的数据库管理系统218、228可以向存储节点240请求页面的最新验证版本。响应于所述请求，存储节点240可以从存储数据库250检索页面的最新验证版本，并将页面的最新验证版本发送到请求所述页面的处理节点210、220的所述数据库管理系统218、228。

[0035] 存储节点240包括以任何合适格式存储数据的存储数据库250。存储数据库250将数据存储在多个页面中，其中每个页面将数据的一部分存储在存储数据库250中。每个页面都可以分配到一个页码，其中页码可以用于标识页面。当页面被修改时，页面的每个版本都可能分配到一个版本号。每次对页面的记录执行修改操作时，页面的版本号可以递增。当对页面的记录执行修改操作时，页面的版本号可能会线性增加，即，版本号可能会增加但不会减少。修改操作可以由日志应用245执行，日志应用245可以执行一个或多个请求以对存储数据库250中的页面的记录执行修改操作。存储节点240可以从公共日志节点235接收日志记录，并且日志应用245可以基于日志记录中包括的信息执行修改操作，从而将这些修改操作应用到存储数据库250。

[0036] 存储数据库250可以包括任何类型的数据库。例如，存储数据库250可以包括Oracle数据库、MySQL数据库、Microsoft SQL Server、PostgreSQL数据库、IBM DB2数据库、

Microsoft Access数据库、SQLite数据库或任何其他类型的数据库。存储节点240可以跨一个或多个计算设备100分布或在一个或多个计算设备100上分布，和/或可以使用一个或多个计算设备100实现。计算设备100可以位于不同的位置。例如，实现存储节点240的第一组计算设备100可以在第一数据中心中，实现存储节点240的第二组计算设备100可以在第二数据中心中。存储数据库250可以在任何合适的存储设备中实现，例如存储器中或磁盘上的数据、存储在单个存储节点240上的数据和/或分布在多个存储节点240上的数据。

[0037] 处理节点210和220的数据库管理系统218、228可以将对页面的请求发送到数据库存储组件240。请求可以包括一个或多个页码。响应于所述请求，存储节点240可以将与所请求的页码相对应的页面发送到处理节点210和220的数据库管理系统218、228。处理节点210和220的数据库管理系统218、228可以接收尚未提交到存储数据库250的页面。例如，处理节点210和220的数据库管理系统218、228可以接收页面的最新验证版本，该页面可能尚未提交到存储数据库250。不允许处理节点210和220写入尚未提交到存储数据库250的页面中的数据。

[0038] 如上所述，每个处理节点210和220的数据库管理系统218、228可以分别维护本地日志缓冲器215和225。本地日志缓冲器215和225可以描述将要对存储数据库250的页面的记录执行的一个或多个操作。例如，本地日志缓冲器215可以存储日志记录，其中包括处理节点210的数据库管理系统218请求的操作的信息。本地日志缓冲器215和225中的每个日志记录可以包括页码、基础页面版本号、下一个页面版本号、将要对页面的记录执行的操作的信息、旧值、新值和/或其任何组合。

[0039] 日志记录中的页码可以指示存储数据库250中将要对其执行操作的页面。每个页面可以与当前版本号关联，该版本号可以用于跟踪页面的版本和/或修改页面的最后一个处理节点。日志记录中的基础页面版本可以包括页面的版本，其中包括操作所基于的记录。当处理节点210或220的数据库管理系统218、228中的一个生成操作时，基础页面版本是相应的处理节点210或220所知悉的页面的最新验证版本。下一个页面版本可以包括在执行操作后页面将具有的版本号。该操作可以包括对页面的记录进行任何更改的指令。新值可以包括页面的全部或部分数据的新值。旧值可以包括存储在页面上的全部或部分数据的当前值。

[0040] 每个处理节点210和220的数据库管理系统218、228可以实现并发控制机制，以防止本地冲突的发生。处理节点210和220处的并发控制机制可以解决在单个处理节点210或220内发生的冲突。换句话说，并发控制机制可以解决处理节点210的操作之间发生的冲突，但通常不能解决处理节点210和处理节点220的操作之间发生的冲突。公共日志230可以解决处理节点210和处理节点220的操作之间发生的冲突，通常称为全局冲突。

[0041] 从处理节点210和220发送到公共日志节点235的执行修改操作的请求，可以由日志应用245应用到存储数据库250。处理节点210和220可以将其本地日志缓冲器215和225的日志记录发送到公共日志节点235。处理节点210或220中的一个可以发送其各自的本地日志缓冲器215或225中的所有日志记录、其各自的日志缓冲器215或225中的日志记录的一部分、基于其各自的本地日志缓冲器215或225中的日志记录生成的指令、或基于其各自的本地日志缓冲器215或225中的日志记录的任何其他数据。公共日志节点235可以基于本地日志缓冲区215或225或从处理节点210和220接收的其他数据，使存储节点240的日志应用245

更新存储数据库250。公共日志节点235可以将日志记录发送到存储节点240，所述日志记录中包括将对数据库的页面的记录执行的修改操作的信息，并且日志应用245可以将这些操作应用到存储数据库250。公共日志节点235可以向存储节点240发送日志记录，例如从处理节点210或220接收的本地日志缓冲器215或225的日志记录。事务的所有操作可以在事务被提交之前应用于存储数据库250。但是，在实际提交事务之前，不会认为该事务已完成。

[0042] 处理节点210的数据库管理系统218和处理节点220的数据库管理系统228可以各自维护本地缓冲池213和223，本地缓冲池213和223可以包括从存储数据库250检索的页面。本地缓冲池213和223可以包括一个或多个不同的页面，其中每个页面具有不同的页码。本地缓冲池213和223可以包括页面的一个或多个版本，其中页面的每个版本具有不同的版本号。

[0043] 公共日志节点235的公共日志230可以检测不同处理节点210和220上的事务之间发生冲突的实例。例如，如果处理节点210和处理节点220都发送对同一页面的相同版本的记录执行修改操作的请求，则公共日志230可以确定是否发生了冲突。每个事务可以在单个处理节点210或220上运行，并且修改操作可以不在处理节点210和220之间交叉。换句话说，单个事务的每个修改操作可以由处理节点210请求，或者单个事务的每个修改操作也可以由处理节点220请求。下文和图3中描述的方法300是公共日志230可用于解决冲突的方法的示例。

[0044] 公共日志230可以确定修改操作是否导致冲突。公共日志节点235可以防止不同处理节点210和220上的两个事务同时更新同一页面。公共日志节点235可以向发送了执行修改操作请求的请求处理节点210或220发送修改操作是否导致冲突的指示，即修改操作已通过验证还是未通过验证的指示。

[0045] 处理节点210或220可以向公共日志节点235发送提交事务的请求。在处理节点210或220从公共日志节点235接收到事务的每个操作均已通过验证的指示后，可以发送提交事务的请求。然后，公共日志230可以确定是否将事务提交到存储数据库250。如果事务中的一个或多个操作未通过验证，公共日志230可以中止事务。公共日志节点235可以向处理节点210或220的请求数据库管理系统218、228发送事务正在中止的指示。

[0046] 在确定中止事务，或从公共日志节点235接收到事务正在中止的指示之后，发送提交事务请求的处理节点210或220可以在本地撤销事务。为了在本地撤销事务，发送了提交事务请求的处理节点210或220可以撤销事务中的请求对存储在其各自的本地缓冲池213或223中的页面的任何影响。公共日志节点235可以向存储节点240发送正在撤销事务的指示。存储节点240可以撤销事务中包括的修改操作对存储在存储数据库250中的页面的任何影响。

[0047] 当公共日志节点235接收到提交事务的请求并确定包括在事务中的每个操作已验证时，公共日志节点235可以向存储节点240发送提交事务的请求。存储节点240可以将事务应用于存储数据库250。公共日志节点235可以向处理节点210和220发送事务已提交的指示。在事务提交到存储数据库250之后，可能难以和/或不可能撤销包括在事务中执行的操作。

[0048] 公共日志节点235可以维护和授予一个或多个锁。处理节点210和220可以维护由锁管理器233分配的锁的全部或部分的本地副本。处理节点210和220可以不维护锁的本地

副本，而是可以通过公共日志节点235访问锁。锁可以是存储在存储器中的数据结构。锁可以由在公共日志节点235上执行的锁管理器233管理。锁可用于防止冲突的发生。锁指示了对特定页面的访问限于处理节点210或220中的一个。锁提供对单个处理节点210或220的独占访问，用于对页面的记录执行操作。

[0049] 不持有锁的处理节点210或220也许仍然能够读取页面，但无法对页面的记录执行修改操作。锁可以包括页码(或页面的其他标识)、持有锁的处理节点的标识以及一个或多个事务的事务列表。事务列表可以指示在持有锁的处理节点上的所有正在运行的事务，这些事务试图对页面的记录执行修改操作。正在运行的事务是在页面的记录上具有挂起的修改操作的事务，其中该事务尚未提交。

[0050] 在事务提交之后，公共日志230可以从锁中包括的事务列表中删除已提交的事务。如果锁中包括的事务列表为空，则公共日志230可以使锁管理器233释放(删除)该锁。在锁被释放(删除)之后，任何其他处理节点210或220都可以对页面的记录执行操作。当事务中止时，公共日志230可以从包括该中止的事务的每个锁的事务列表中删除该事务。公共日志230可以在删除中止的事务之后释放具有空事务列表的任何锁。

[0051] 公共日志230可以监控从处理节点210和220接收的日志记录。当例如从处理节点210接收到指示事务T已经更新页面P的日志记录时，公共日志230可以确定是否存在与页面P对应的锁。如果公共日志230确定不存在与页面对应的锁P，公共日志230可以将与页面P对应的锁授予处理节点210，并且如果事务T不存在，则可以将其添加到锁的事务列表中。另一方面，如果与页面P对应的锁由另一个处理节点，例如处理节点220持有，则可能存在冲突，事务T可能会被中止和/或撤销。

[0052] 公共日志节点235可以向处理节点210和220发送事务已提交和/或中止的记录。可以在预定义数量的事务已提交和/或中止后以预定间隔发送记录，或者可以在事务被提交和/或中止时实时发送记录，或者可以以任何其他间隔发送记录。对于处理节点210和220中的每一个，所发送的记录可以指示其他处理节点已经提交和/或中止的事务。

[0053] 图3是根据本公开的一个或多个说明性方面的用于检测冲突的方法300的流程图。在一个或多个实施例中，方法300或其一个或多个步骤可以由公共日志节点235的公共日志230执行。方法300或其一个或多个步骤可以体现在计算机可执行指令中，这些计算机可执行指令存储在非瞬时性大容量存储设备104等非瞬时性计算机可读介质中，加载到存储器108中，并由实现公共日志节点235的计算设备100的CPU 114执行。流程图中的某些步骤或步骤的某些部分可以按顺序省略或更改。

[0054] 在步骤305，接收对页面的记录执行修改操作的请求(“请求”)。该请求可以从处理节点接收，例如处理节点210或220中的一个。传输请求的处理节点可以称为请求处理节点。所述请求可以包括待执行修改操作的页面的页码、所述请求对应的事务、所述页面的基础版本号、所述页面的更新版本号、所述事务对应的请求处理节点、将要对页面的记录执行的修改操作、和/或用于对页面的记录执行修改操作的任何其他信息。修改操作可以包括任何修改页面的记录的操作，例如，在页面中插入新记录的插入记录操作、删除页面的记录的删除记录操作、更新页面的记录的更新记录操作。

[0055] 该请求可以单独接收，也可以与一个或多个其他请求一起接收。例如，处理节点210和220可以发送多个请求，其中每个请求是对页面的记录执行修改操作的请求。在一些

实施例中，由处理节点210、220发送的每个请求可以包括来自本地日志缓冲器215、225的日志记录，该日志记录包括将要在页面的记录上执行的修改操作的信息。

[0056] 在步骤310，可以确定是否存在与页面对应的活动锁(例如，锁已分配给与请求中包括的页面对应的页面)。锁可以存储在哈希表、搜索树或任何其他数据结构中。可以搜索锁以确定是否存在与步骤305中接收的请求中指示的页码相对应的锁。可以查询锁的集合以确定是否存在与页面相对应的活动锁。

[0057] 如果在步骤310确定不存在对应于页面的活动锁，则可以在步骤320创建锁。创建的锁可以包括页面、请求处理节点和请求中包括的事务。例如，创建的锁可以包括页码、事务编号和请求中包括的请求处理节点的名称或其他标识。在创建锁之后，可以在下文描述的步骤335将请求中包括的基础页面版本号与最新页面版本号进行比较。

[0058] 如果在步骤310确定存在对应于请求了修改操作的页面的活动锁，则可以在步骤315确定活动锁是否分配给了请求处理节点。检索与页面对应的活动锁，并确定分配到活动锁的处理节点。可以将已分配到与页面对应的活动锁的处理节点与请求处理节点进行比较。如果所述请求处理节点不是已分配到与页面对应的活动锁的处理节点，则可能发生冲突，并且可以在步骤330检测到冲突。在步骤330，检测到冲突，并且方法300进入步骤350，在步骤350中，将请求未通过验证的指示发送到请求处理节点和/或存储。所述指示可以包括页码、事务编号、请求的操作和/或与请求的操作相对应的任何其他细节。

[0059] 如果在步骤315确定活动锁已分配给请求处理节点，或在步骤320已创建锁之后，在步骤335，通过将请求中指示的页面的基础版本号与页面的最新验证版本号进行比较，确定请求中指示的页面的基础版本号是否为页面的最新验证版本。公共日志230可以维护存储数据库250的页面的最新验证版本号的记录。页面的最新验证版本号可以包括对应于已经通过验证和/或已写入公共日志230的页面的最新版本的版本号。页面的最新验证版本可能已经提交到存储数据库250，或者可能尚未提交到存储数据库250。

[0060] 如果页面的基础版本号与页面的最新验证版本号不相同，则在步骤330可以检测到冲突。当检测到冲突时，方法300进入步骤350，并且将请求中包括的修改操作未通过验证的指示发送到请求处理节点和/或存储。

[0061] 如果在步骤335确定基本版本是最新的验证版本，则在步骤340，可以将请求中包括的修改操作的事务编号添加到锁的事务列表中。事务列表可以包括有序列表，并且请求中包括的修改操作的事务编号可以添加在有序列表的末尾，即附加到有序列表后。在事务已经添加到事务列表之后，在步骤350，可以存储和/或发送请求中包括的修改操作已经通过验证的指示符。如果在步骤330检测到冲突，则请求中包括的修改操作可以视为未通过验证。否则，如果在步骤340没有检测到冲突，并且事务已添加到锁的事务列表中，则请求中包括的修改操作可以视为已经通过验证。

[0062] 布尔数组可用于存储和/或发送请求中包括的修改操作已通过或未通过验证的指示。布尔数组可以对应多个请求，每个请求对页面的记录执行修改操作。布尔数组的每个索引可以与请求相关联，并且可以包括关于该请求中包括的修改操作是通过验证还是未通过验证的指示。在预定数量的修改操作通过或未通过验证之后，可以将布尔数组发送到请求处理节点。布尔数组可以对应接收到的日志缓冲区，布尔数组的每个索引可以对应日志缓冲区的一条日志记录。如果修改操作通过了验证，则可以在步骤345将修改操作通过验证的

指示符写入布尔数组。如果修改操作未通过验证，则可以在步骤330将修改操作未通过验证的指示符写入布尔数组。

[0063] 在日志缓冲区中的每个日志记录已经通过或未通过验证之后，可以在步骤350将布尔数组或一个或多个修改操作已经通过或未通过验证的任何其他指示符发送到请求处理节点。对于接收到的日志缓冲区中的每个日志记录，布尔数组可以指示修改操作是通过还是未通过验证。如果任何修改操作未通过验证，则包括该修改操作的事务可能会被中止。事务可以由公共日志230中止。

[0064] 图4示出了根据本公开的一个或多个说明性方面的页面更新的调用流程图。处理节点210和220可以与存储节点240通信，以存储和访问存储数据库250中的数据。尽管图4中未示出，但是处理节点210和220可以通过公共日志节点235与存储节点240通信。图4是通过方法300的步骤验证修改操作的示例。

[0065] 在时间450，存储数据库250可以包括存储页面405。在图4所示的示例中，存储页面405具有页码“1”、版本号“220N66”，并包括值为“1,2,3”的数据的数组。页面的版本号可以指示处理节点220为上一个更新页面的处理节点，并且页面的版本是“66”。虽然以示例性格式示出，但版本号、页码和/或数据可以以任何可用格式排列。

[0066] 在时间460，处理节点210可以发送对存储页面405的记录执行修改操作410的请求(“请求410”)，以更新页面405中存储的数据。请求410包括修改操作，以将页面405从基础版本号“220N66”更新为版本号“210N77”，以及将页面中存储的记录的数据更新为“1,22,3”。请求410可以由处理节点210通过公共日志节点235发送到存储节点240。

[0067] 在请求410中包括的修改操作提交到存储数据库250之前，处理节点220发送对存储数据库250的记录执行修改操作420的请求(“请求420”)，该请求包括对存储页面405的记录的更新。请求420中包括的基础版本号可以是“220N66”，其可以是存储在存储数据库250和/或处理节点220的本地缓冲池223中的存储页面405的当前版本号，并且可以是处理节点220知悉的页面405的最新版本号。

[0068] 在时间480，存储页面405的记录可以通过请求410中包括的修改操作进行更新，得到页面415。页面415包括请求410中包括的版本号和数据。包括请求410的事务可以提交到存储数据库250。存储页面405的页码“1”在页面415中可以保持相同。

[0069] 在时间490，公共日志230可以确定请求420中包括的修改操作导致冲突。公共日志230可以将请求420中包括的基础版本号与页面415的最新验证版本号进行比较，例如上文描述的图3的步骤335。由于请求420中包括的基础版本号不同于页面415的最新验证版本号，因此，检测到冲突，公共日志发送指示，指示修改操作未通过验证，并且请求420中包括的修改操作未在记录上执行(例如，未将对记录的更新写入存储数据库250)。

[0070] 图5是根据本公开的一个或多个说明性方面的用于释放锁的方法500的流程图。在一个或多个方面中，方法500或其一个或多个步骤可以由公共日志节点235的公共日志230和/或锁定管理器233执行。方法500或其一个或多个步骤可以体现在计算机可执行指令中，这些计算机可执行指令存储在大容量存储设备104等计算机可读介质中，加载到存储器108中，并由实现公共日志节点235的计算设备100的CPU 114执行。流程图中的某些步骤或步骤的某些部分可以按顺序省略或更改。

[0071] 在步骤505，例如由公共日志节点235接收到事务已经提交或中止的指示。该指示

可以从处理节点210或220等处理节点接收。该指示可以包括事务编号或事务的任何其他指示。

[0072] 在步骤510,可以确定与提交或中止的事务相对应的所有锁。可以分析由锁管理器233维护的所有锁,并且可以找到在其事务列表中具有事务的每个锁。可以生成锁列表,其中该列表包括在其事务列表中具有事务的每个锁。

[0073] 在步骤515,可以迭代地遍历锁列表,从锁列表中的第一个锁开始作为当前锁。锁列表可以以任何顺序遍历。可以选择锁列表中的任何锁作为当前锁,无论其在锁列表中的位置如何。

[0074] 在步骤520,可以将事务从当前锁的事务列表中删除。在步骤525,可以将当前锁从锁列表中删除。从锁列表中删除锁可以表明事务不再在锁的事务列表中。

[0075] 在步骤530,可以分析当前锁的事务列表以确定该列表是否为空。如果事务列表确定为空,则在步骤535,可以释放当前锁。如果事务列表包含其他事务,则可以不释放锁。在步骤535释放锁之后,或在步骤530确定事务列表包括至少一个其他事务之后,在步骤540,可以将锁列表中的下一个锁设置为当前锁。如果没有可用的锁可选择为当前锁,或者换句话说,如果在步骤540中没有锁保留在锁列表中,则方法500可以结束。在步骤540将下一个锁设置为当前锁之后,然后可以在步骤520将事务从新的当前锁的锁列表中删除。

[0076] 图6是根据本公开的一个或多个说明性方面的用于提交事务的方法600的流程图。在一个或多个方面中,方法600或其一个或多个步骤可以由处理节点210或220的数据库管理系统218、228或公共日志节点235的公共日志230中的一个执行。方法600或其一个或多个步骤可以体现在计算机可执行指令中,这些计算机可执行指令存储在非瞬时性大容量存储设备104等计算机可读介质中,加载到存储器108中,并由实现处理节点210或220或公共日志节点235的计算设备100的CPU 114执行。流程图中的某些步骤或步骤的某些部分可以按顺序省略或更改。

[0077] 方法600可用于确定是否提交事务或撤销事务。如果事务包含未通过验证的修改操作,则事务可能会中止和/或事务对每个页面做出的更改可能会撤销。事务的修改操作已更改的页面可以从锁中确定,例如由公共日志节点235的锁管理器233和/或处理节点210的数据库管理系统218或处理节点220的数据库管理系统228维护的锁。可以搜索由公共日志节点235的锁管理器233或处理节点210和220中的一个维护的所有锁,以确定在其事务列表中包括事务的所有锁的列表。锁可以由公共日志节点235的锁管理器233和/或处理节点210和220中的一个进行搜索。可以确定和/或将列表中锁的页码存储在列表中。处理节点210或220中的一个可以向公共日志节点235发送指令,以返回事务已修改的所有页码的列表。页码列表可以指示由事务的修改操作修改的每个页面。

[0078] 处理节点210和220等处理节点可以接收指示修改操作通过或未通过验证的一个或多个记录。所述一个或多个记录可以存储在布尔数组等数组或任何其他合适的数据结构中。方法600可以迭代遍历数组,处理数组中的每个操作。

[0079] 在步骤605,可以从数组中选择第一个修改操作作为当前修改操作。可以从数组中选择任何修改操作,或者换句话说,可以以任何顺序遍历数组中的修改操作。

[0080] 在步骤610,可以确定当前修改操作已经通过或未通过验证。数组可以指示当前修改操作是通过还是未通过验证。例如,如果数组是布尔数组,则在与修改操作对应的条目

中，“1”可以指示修改操作通过了验证，“0”可以指示修改操作未通过验证。

[0081] 如果在步骤610确定修改操作已通过验证，则在步骤620可以确定是否存在与在步骤610待检查的事务相对应的任何修改操作。数组可以包括用于事务中包括的所有修改操作的指示，在这种情况下，可以在步骤620确定是否已经完全遍历该数组。

[0082] 如果在步骤620没有剩余操作，由于在步骤610已检查了事务中包括的所有修改操作并确定均已通过验证，则可以在步骤630提交事务。可以发送应提交事务的指示。所述指示可以包括所述事务的事务编号和提交所述事务的指令。执行事务的处理节点210或220可以将指示发送到公共日志节点235。公共日志节点235和/或处理节点210或220可以向存储节点240发送指令以提交事务。公共日志节点235可以向所有处理节点或除与事务对应的处理节点之外的所有处理节点发送事务已提交的指示。

[0083] 如果在步骤620仍有操作，则可以在步骤625从数组中选择数组中的下一个修改操作作为当前操作。然后，可以在步骤610检查当前修改操作，以确定修改操作是否通过或未通过验证。

[0084] 如果在步骤610，确定当前事务的任何修改操作未通过验证，则可以在步骤615中止事务。下文和图7中描述的方法700是一种用于中止事务的方法。

[0085] 图7是根据本公开的一个或多个说明性方面的用于中止事务的方法700的流程图。在一个或多个方面中，方法700或其一个或多个步骤可以由一个或多个计算设备或实体执行。例如，方法700的部分可以由计算设备100的节点执行。方法700的全部或部分可以由处理节点210和220、公共日志节点235和/或存储节点240执行。方法700或其一个或多个步骤可以体现在计算机可执行指令中，这些计算机可执行指令存储在非瞬时性计算机可读介质等计算机可读介质中。流程图中的某些步骤或步骤的某些部分可以按顺序省略或更改。

[0086] 在步骤705，中止事务的过程开始于由事务中包括的修改操作修改的第一个页面。在步骤705，可以选择由事务中包括的修改操作修改的页面中的任何页面。

[0087] 在步骤710，确定修改操作是否可撤销。某些修改操作可能被撤销，而其他修改操作可能是不可撤销的。哪些修改操作可能是可撤销的，哪些修改操作可能是不可撤销的，可能取决于所使用的RDBMS的类型。预定的列表可以指示哪些类型的修改操作是可撤销的，哪些类型的修改操作是不可撤销的。

[0088] 如果修改操作是可撤销的，则在步骤720撤销修改操作。修改操作可以使用RDBMS的事务回滚机制撤销。修改操作对页面所做的更改可以使用存储的信息撤消。例如，回滚段可以存储用于撤销修改操作的撤消信息。可撤销修改操作可以在不影响任何其他事务的情况下撤销。

[0089] 如果修改操作是不可撤销的，则可以在步骤715从请求处理节点210、220的本地缓冲池213、223中删除页面。删除的页面可以用页面的最新验证版本替换，该版本可以从存储节点240检索。

[0090] 在步骤715从缓冲池中删除页面之后，或者在步骤720撤销修改操作之后，可以在步骤725确定是否有更多与事务相对应的页面要处理。如果还有更多的页面要处理，则可以从由事务的修改操作修改的页面中选择与事务对应的下一个页面。对于由事务中包括的修改操作修改的任何页面，如果其在步骤715或720中的任一步骤中尚未经过处理，则可以选择作为下一个页面。在步骤730选择下一个页面之后，可以在步骤710检查与该页面相对应

的修改操作，以确定修改操作是否可撤销。

[0091] 如果在步骤725确定所有页面在步骤715或720中的任一步骤中均已经过处理，则在步骤735中完成撤销。在步骤735，可以释放对应于事务的锁。可以向公共日志节点235发送事务已中止的指示。所述指示可以包括所述事务的事务编号。向公共日志节点235发送指示可以使公共日志230从对应于已撤销事务的所有锁的事务列表中删除事务，并释放在其事务列表中没有其他事务的任何锁。

[0092] 图8A和8B是根据本公开的一个或多个说明性方面的用于处理修改操作的方法800的流程图。在一个或多个实施例中，方法800或其一个或多个步骤可以由处理节点210的数据库管理系统218或处理节点220的数据库管理系统228执行。方法800或其一个或多个步骤可以体现在计算机可执行指令中，这些计算机可执行指令存储在大容量存储设备104等非瞬时性计算机可读介质中，加载到存储器108中，并由实现处理节点210或220的计算设备100的CPU 114执行。流程图中的某些步骤或步骤的某些部分可以按顺序省略或更改。

[0093] 方法800可以由处理节点210和220之一的数据库管理系统218、228执行，以便对存储数据库250的页面的记录执行修改操作。如上所述，处理节点210和220可以分别包括本地缓冲池213和223。当对存储在存储数据库250中的页面的记录执行修改操作时，处理节点210和220对存储在本地缓冲池213和223中的记录执行修改操作，可能会比处理节点210和220每次对页面的记录执行修改操作时从存储节点240检索记录更高效。

[0094] 在步骤805，可以接收执行修改操作的请求。还可以接收并将正在对其执行修改操作的页面存储在本地日志缓冲器213、225中。每个请求都可以对应于一个事务。从存储数据库250检索正在对其执行修改操作的页面。可以从用户节点203或205中的一个接收请求。

[0095] 在步骤810，可以接收全局页面更改列表。全局页面更改列表可以从公共日志节点235接收。全局页面更改列表可以包括由公共日志230验证的一个或多个页面更新的指示。全局页面更改列表可以指示最近更新的页面等各种页面的最新验证版本号。例如，当处理节点220向公共日志节点235发送修改操作时，公共日志230可以验证修改操作，如果公共日志230确定修改操作不引起任何冲突，公共日志节点235可以在全局页面更改列表中发送与修改操作对应的页面的最新验证版本号。全局页面更改列表可以周期性地发送，也可以以任何间隔发送。在执行方法800时，可以接收多个全局页面更改列表。

[0096] 在步骤815，可以确定存储数据库250中所有页面的单个最高版本号，可以称为“事务开始版本”。可以向公共日志节点235请求事务开始版本。公共日志230可以确定事务开始版本，并将事务开始版本发送到处理节点210或220。事务开始版本可以是存储数据库250中的页面的版本号，该存储数据库250中具有任何页面的最高版本号。

[0097] 在步骤820，可以确定与在步骤805接收的修改操作相对应的页面。与操作相对应的页面可以包括在处理或执行修改操作时被修改的页面。与修改操作相对应的页面可以包括处理节点210或220在执行修改操作期间使用的页面。可以生成页面列表，其中页面列表包括在执行修改操作时将被修改的所有页面的页码。

[0098] 在步骤825，可以迭代地遍历页面列表，从列表中的第一个页面开始作为当前页。可以以任何顺序遍历页面列表。可以选择列表中的任何页面作为当前页面，无论其在列表中的位置如何。从页面列表中选择当前页面后，可以从页面列表中删除该当前页面。

[0099] 在步骤830，可以确定修改操作是否正在修改当前页面。如果当前页面正在被修

改，则方法800可以使用页面的最新验证版本作为修改操作的基本版本，否则可能会发生如上文所述的关于图3的步骤330和335的冲突。如果当前页面没有被修改，例如正在读取当前页面，但在当前页面上没有执行修改操作，则可以使用当前页面的旧版本。例如，如果存储节点240使用数据库快照，则当前页面的旧版本可以用于页面读取。如果可以使用当前页面的旧版本，而不是当前页面的最新验证版本，则可以更快地执行页面读取。

[0100] 如果在步骤830确定当前页面正在被修改，则可以在步骤840确定当前页面的最新验证版本是否存储在正在执行修改操作的处理节点210或220的本地缓冲池213或223中。如上文在步骤810所述，全局页面更改列表可以指示一个或多个页面的最新验证版本号。处理节点210和220可以周期性地接收全局页面更改列表。当前页面的最新验证版本号可以基于全局页面更改列表确定。如果在步骤840，确定当前页面的最新验证版本存储在正在执行修改操作的节点的本地缓冲池213或223中，则在步骤860，将从本地缓冲池读取或更新当前页面的最新验证版本。

[0101] 如果在步骤840，确定当前页面的本地缓冲池版本号与当前页面的最新验证版本号不匹配，则可以在步骤850将所需版本号设置为页面的最新验证版本号。在步骤855，可以从存储节点240检索与当前页面的所需版本号相对应的页面。为了检索当前页面的所需版本号，处理节点210或220可以向存储节点240发送请求。数据库存储组件240可以从存储数据库250检索当前页面的所需版本。存储节点240可以将当前页面的所需版本发送到请求处理节点210或220。

[0102] 如果在步骤830确定修改操作没在修改当前页面，则在步骤835，可以将存储在相关本地缓冲池213或223中的当前页面的版本号与在步骤815确定的事务开始版本进行比较。如果在步骤835，存储在相关本地缓冲池中的当前页面的版本号对于事务开始版本无效，则在步骤860将从本地缓冲池中读取当前页面。

[0103] 如果在步骤835中，当前页面未存储在本地缓冲池中，或者存储在本地缓冲池中的当前页面的版本号大于事务开始版本，在步骤845，将页面的所需版本号设置为在步骤815确定的事务开始版本。

[0104] 在步骤855，基于所需的版本号检索当前页面。在步骤855，可以检索版本号小于或等于所需版本号的页面。可以检索当前页面的最高现有版本号，该版本号小于或等于所需版本号。例如，如果不存在与所需版本号匹配的当前页面的版本，则可以检索小于所需版本号的当前页面的最高版本号。

[0105] 在步骤855，可以将针对当前页面的所需版本号的请求发送到存储节点240。该请求可以不包括所需版本号，或是除了包括所需版本号之外，该请求还可以指示请求当前页面的最新验证版本。例如，请求中的标志可以指示请求当前页面的最新验证版本。存储节点240可以确定当前页面的最高验证版本号，该版本号小于或等于所需版本号，并且可以返回当前页面的该版本。如果设置了指示请求当前页面的最新验证版本的标志，则存储节点240可以返回当前页面的最新验证版本。在步骤855检索的页面可以存储在本地缓冲池213或223中。

[0106] 在步骤860，可以从本地缓冲池读取和/或修改当前页面。可以基于在步骤805接收的页面执行页面修改。在步骤865，可以确定页面列表中是否有更多的页面要处理以进行修改操作。如果页面列表中没有剩余的页面，则在步骤875可以认为修改操作完成。如果页面

列表中还有更多的页面，则在步骤870处可以选择下一个页面作为当前页面。可以选择页面列表中的任何页面作为下一个页面。在选择下一个页面作为当前页面后，可以将该页面的页码从页面列表中删除。然后，该方法可以进入步骤830，以确定是否正在修改当前页面。

[0107] 尽管图8中未示出，但是在某些情况下，页面读取操作可以使用当前页面的最新验证版本，而非使用当前页面的旧版本。例如，如果存储节点240不支持和/或实现数据库快照，则页面读取操作可以使用当前页面的最新验证版本。在这种情况下，在步骤830之后，无论是否正在修改页面，方法800都可以在步骤840继续，确定最新验证版本是否在本地缓冲池中。

[0108] 如上所述，当多个处理节点访问同一数据库系统时，经常会发生冲突。当处理节点和存储节点之间存在通信延迟时，数据库系统中特别容易发生冲突。本文描述的公共日志节点可以防止当多个处理节点尝试并发写入数据库系统中的页面的同一记录时，发生冲突。本文描述的公共日志节点可以最小化由于冲突检测而发生的任何处理延迟。

[0109] 尽管上文描述了示例性实施例，但是根据具体的结果或应用，各种特征和步骤可以以任何所需的方式组合、划分、省略、重新排列、修订或增强。本文中的各种元素已描述为“A和/或B”，其意指以下任何一种：“A或B”、“A和B”、“A中的一个或多个和B中的一个或多个”。本领域技术人员可以很容易地想到各种变更、修改和改进。尽管在本文中没有明确说明，但是本公开所明显说明的这些变更、修改和改进旨在成为本描述的一部分，且不脱离本公开的精神和范围。因此，上述描述仅作为示例而非限制。本专利仅由以下权利要求及其等效物所限定。

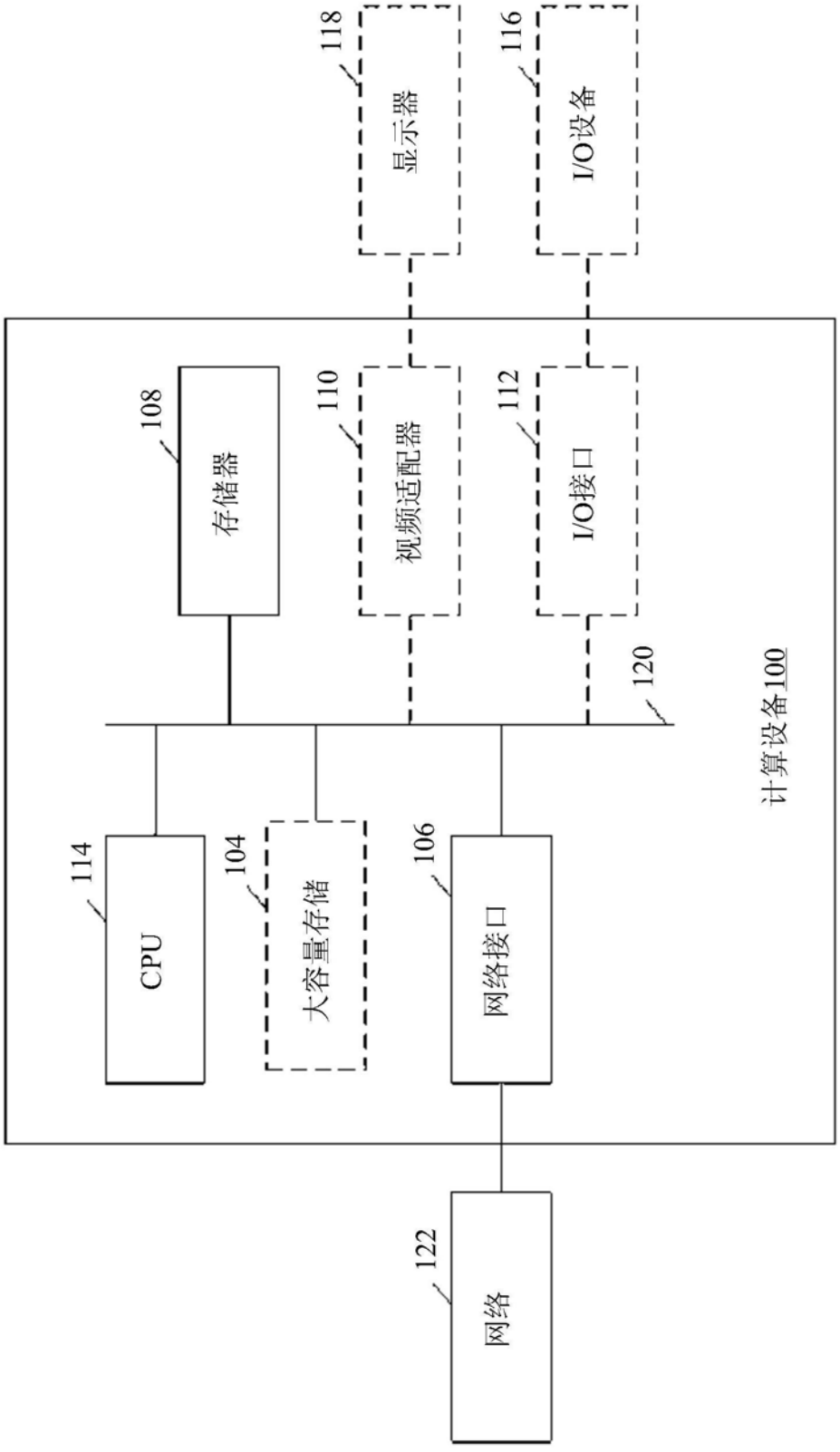

118
显示器
116
I/O设备
108
存储器
110
视频适配器
112
I/O接口
120
计算设备100
114
CPU
104
大容量存储
106
网络接口
122
网络

图1

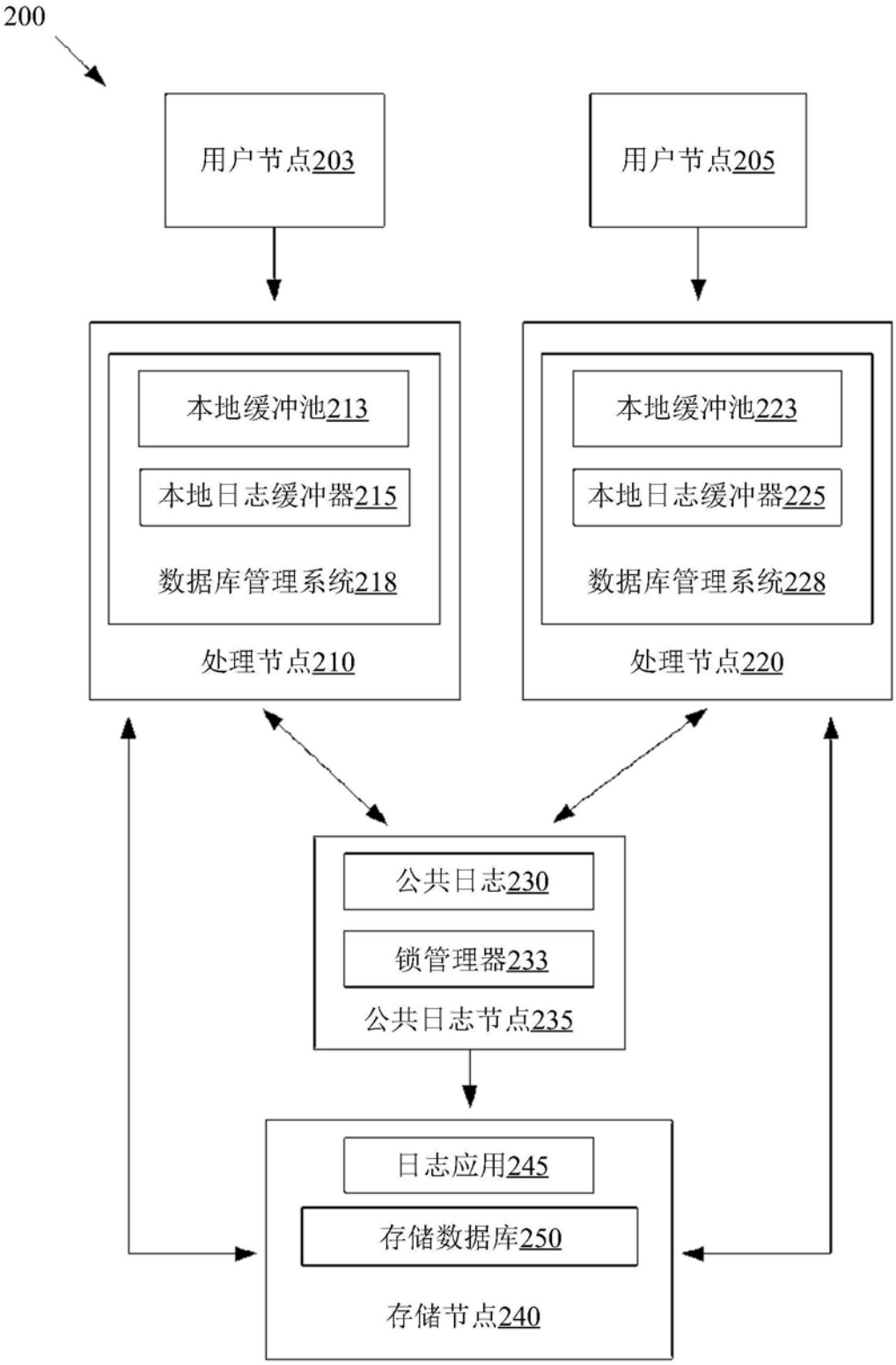

200
用户节点203
用户节点205
本地缓冲池213
本地日志缓冲器215
数据库管理系统218
处理节点210
本地缓冲池223
本地日志缓冲器225
数据库管理系统228
处理节点220
公共日志230
锁管理器233
公共日志节点235
日志应用245
存储数据库250
存储节点240

图2

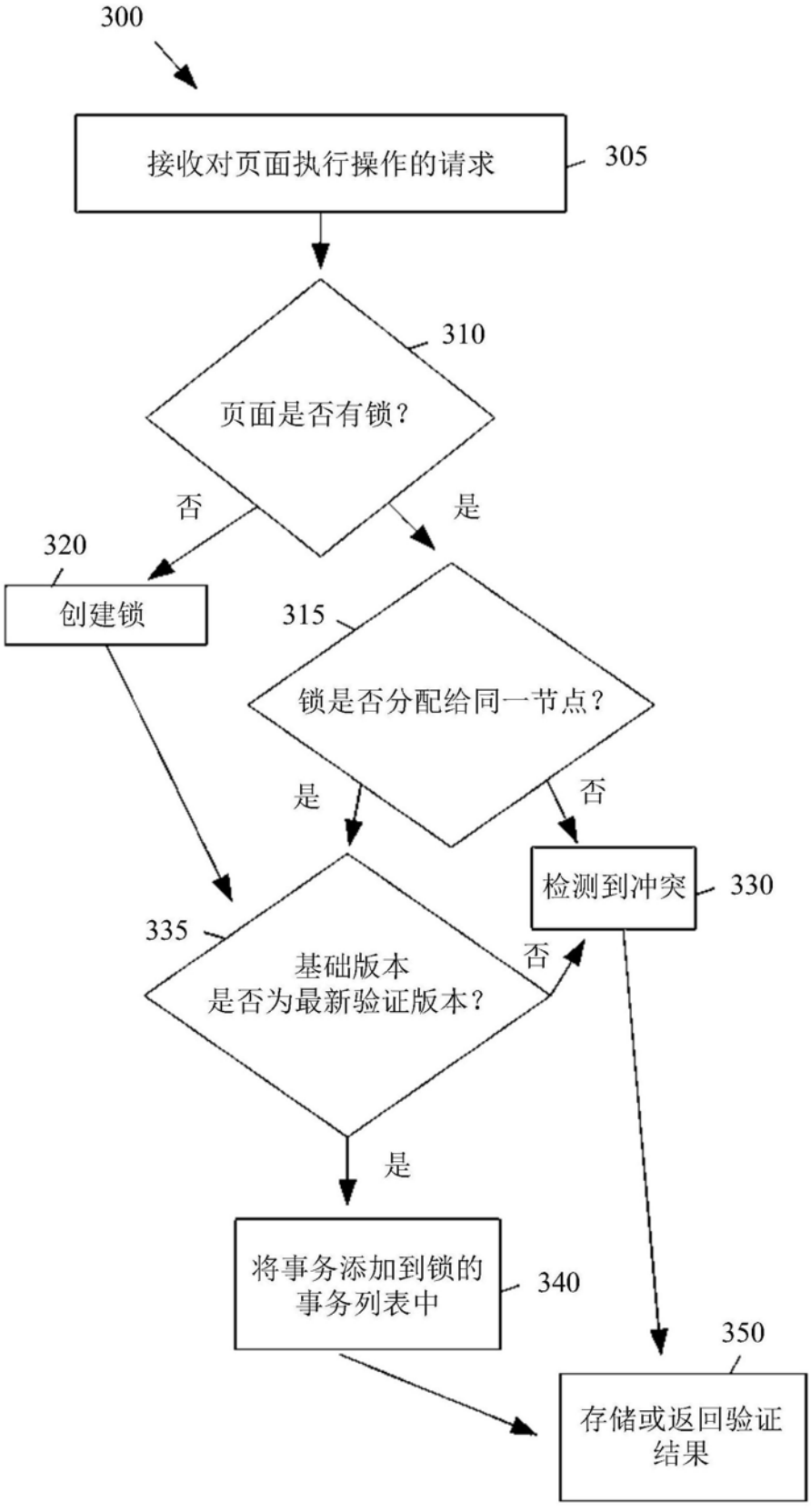

300
接收对页面执行操作的请求
305
310
页面是否有锁？
否
是
320
创建锁
315
锁是否分配给同一节点？
是
否
检测到冲突
330
335
基础版本
是否为最新验证版本？
否
是
将事务添加到锁的
事务列表中
340
350
存储或返回验证
结果

图3

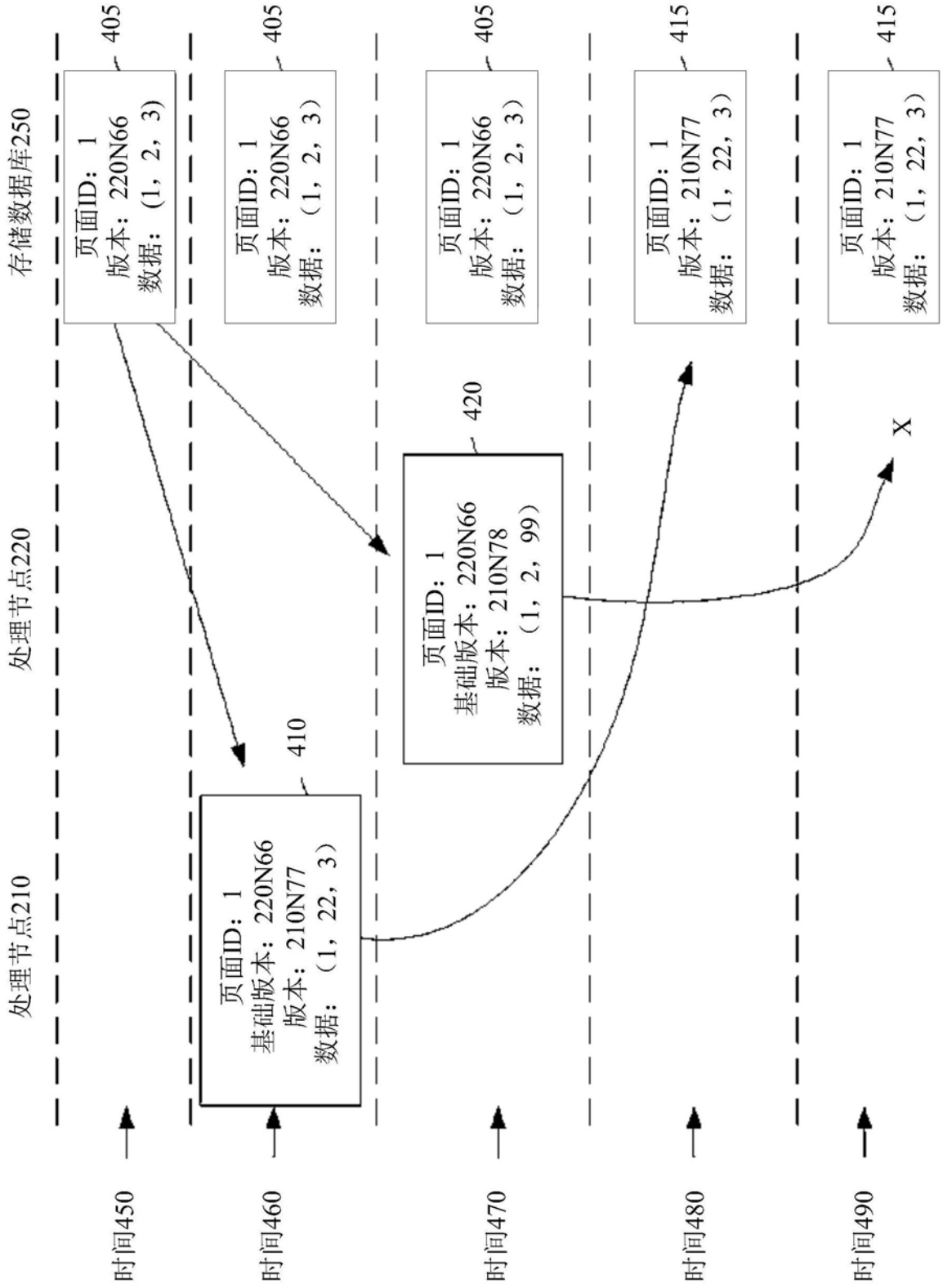

存储数据库250
处理节点220
处理节点210
405
页面ID：1
版本：220N66
数据：(1，2，3)
405
页面ID：1
版本：220N66
数据：（1，2，3）
405
页面ID：1
版本：220N66
数据：（1，2，3）
415
页面ID：1
版本：210N77
数据：（1，22，3）
415
页面ID：1
版本：210N77
数据：（1，22，3）
410
页面ID：1
基础版本：220N66
版本：210N77
数据：（1，22，3）
420
页面ID：1
基础版本：220N66
版本：210N78
数据：（1，2，99）
X
时间450
时间460
时间470
时间480
时间490

图4

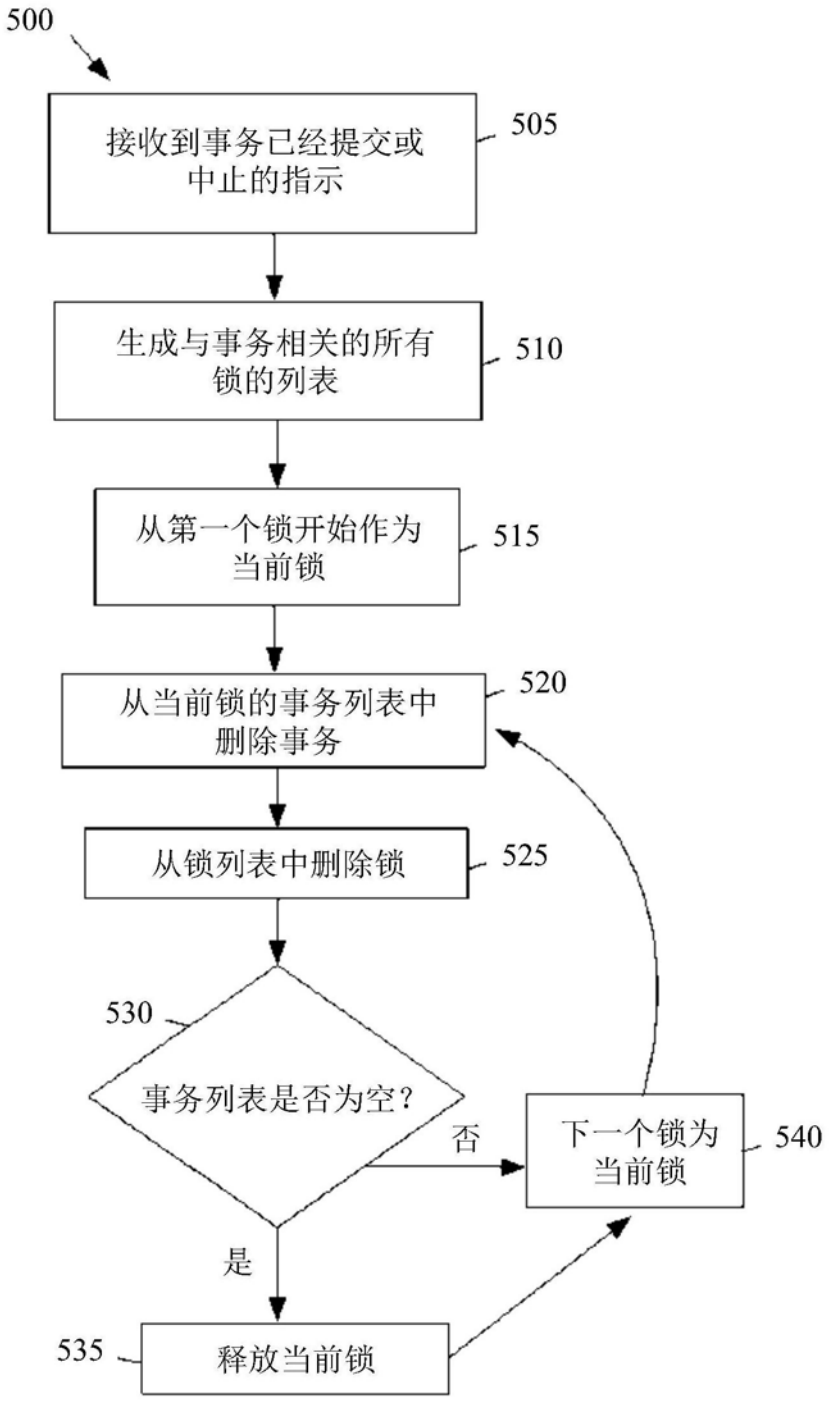

500
接收到事务已经提交或中止的指示
505
生成与事务相关的所有锁的列表
510
从第一个锁开始作为当前锁
515
从当前锁的事务列表中删除事务
520
从锁列表中删除锁
525
530
事务列表是否为空？
否
下一个锁为当前锁
540
是
535
释放当前锁

图5

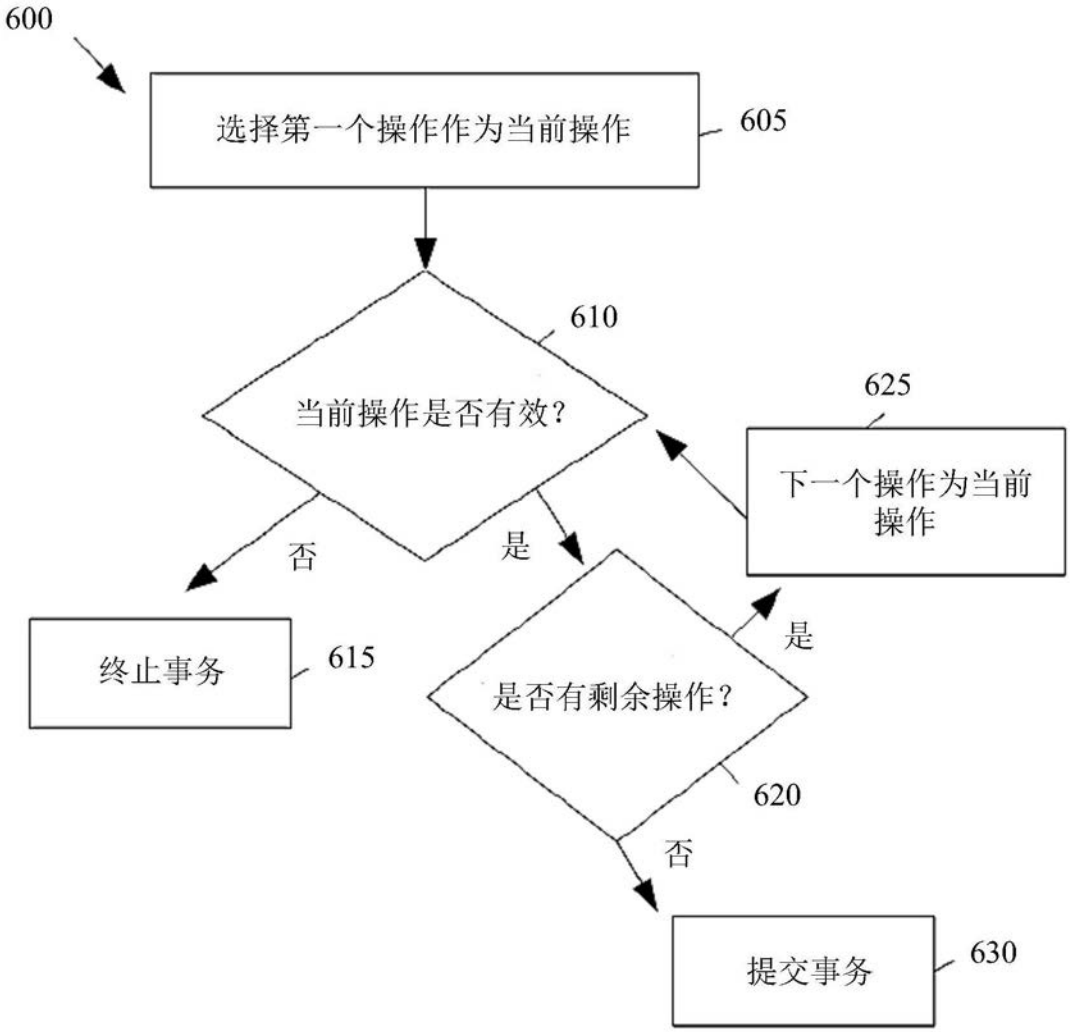

600
选择第一个操作作为当前操作
605
610
当前操作是否有效?
625
下一个操作为当前操作
否
是
终止事务
615
是
是否有剩余操作?
620
否
提交事务
630

图6

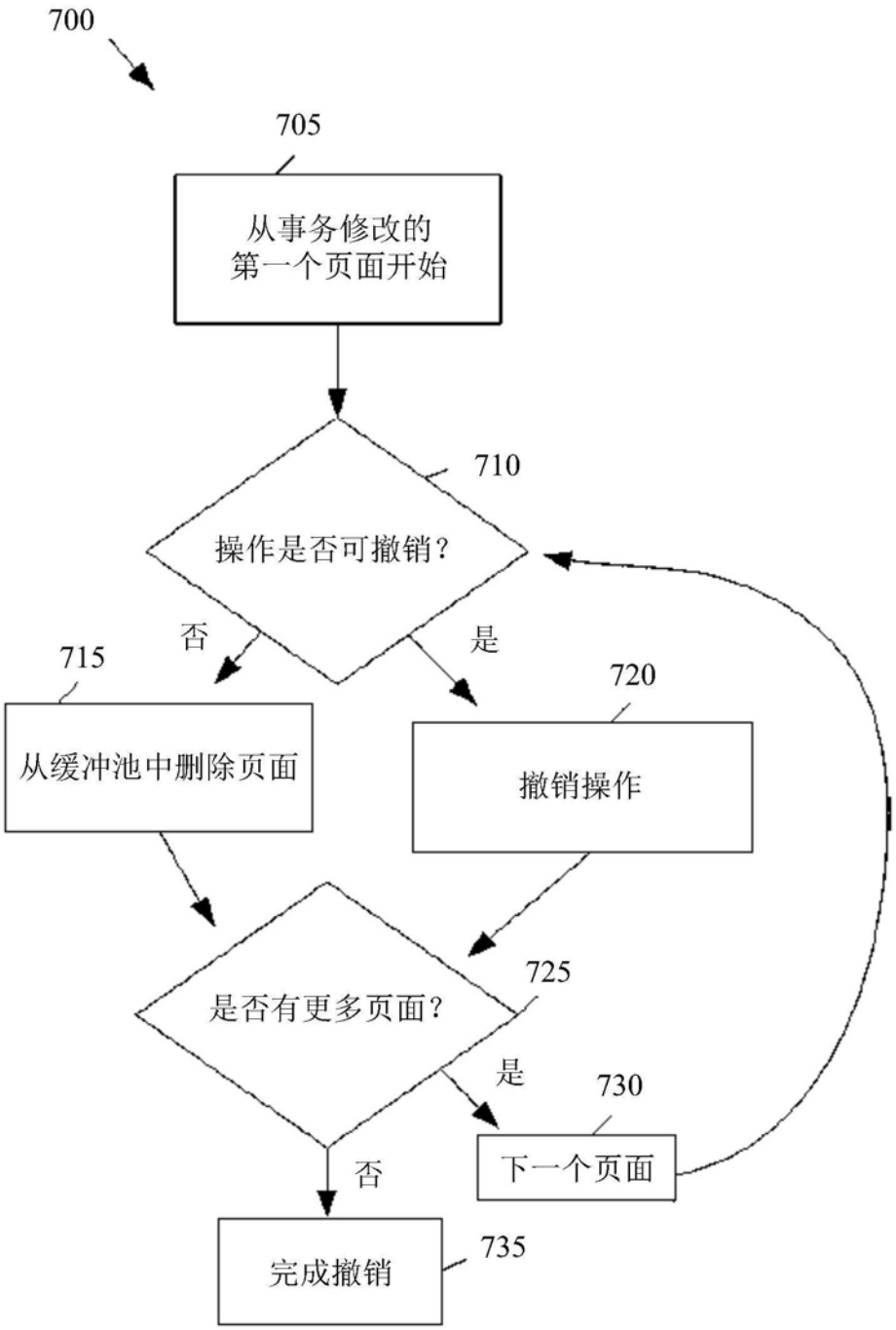

700
705
从事务修改的
第一个页面开始
710
操作是否可撤销?
否
是
715
从缓冲池中删除页面
720
撤销操作
725
是否有更多页面?
是
730
下一个页面
否
735
完成撤销

图7

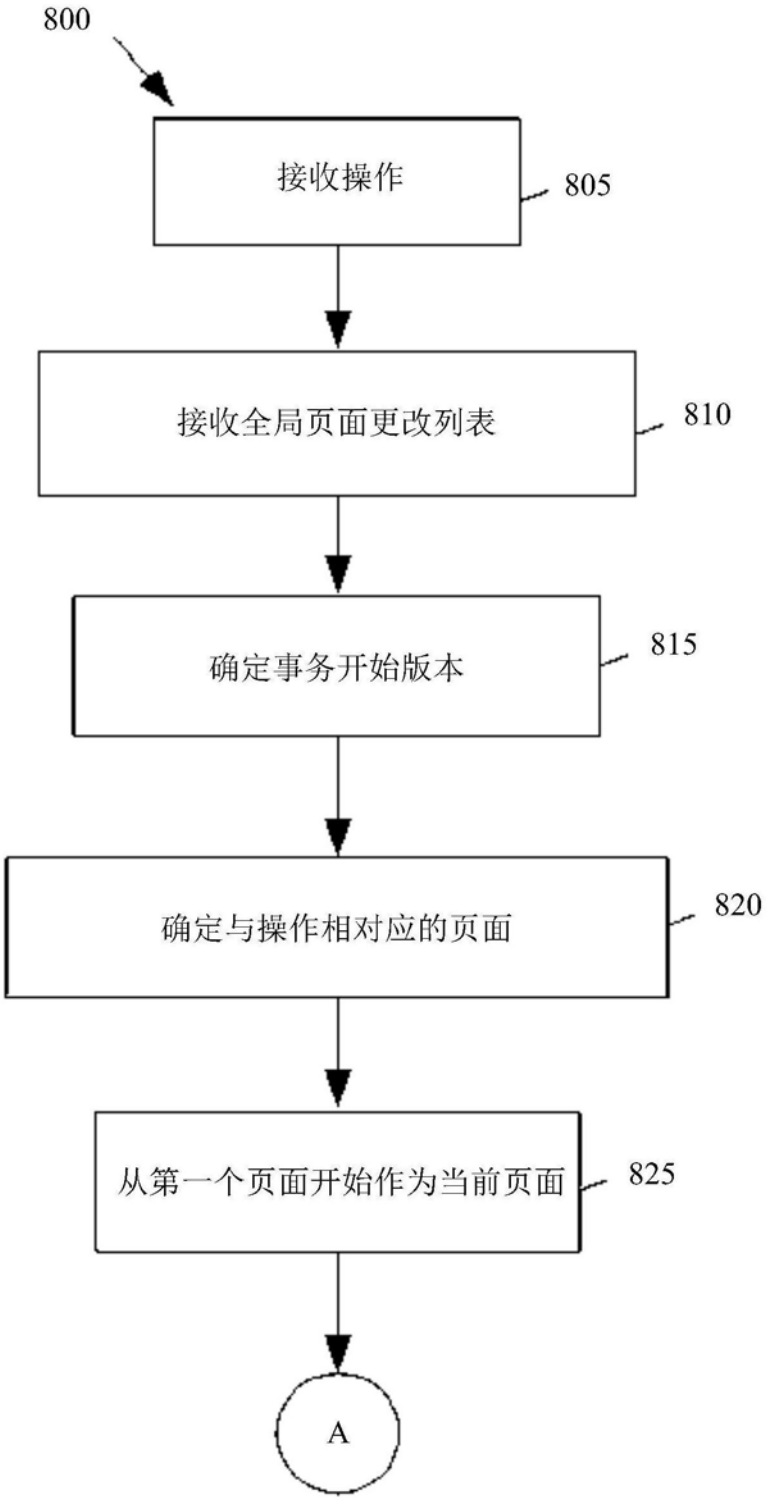

800
接收操作
805
接收全局页面更改列表
810
确定事务开始版本
815
确定与操作相对应的页面
820
从第一个页面开始作为当前页面
825
A

图8A

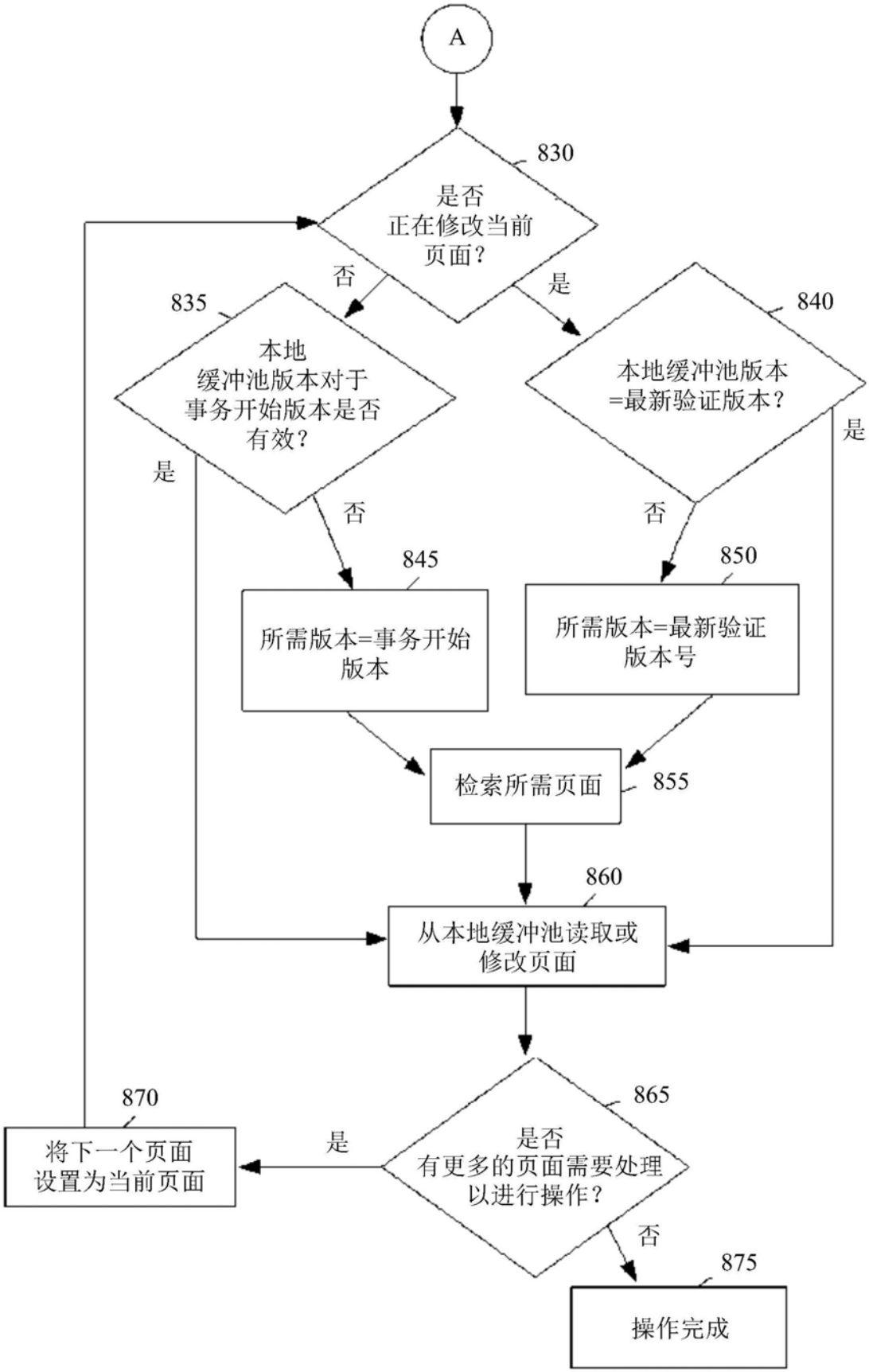

A
830
是否
正在修改当前
页面？
否
是
835
本地
缓冲池版本对于
事务开始版本是否
有效？
840
本地缓冲池版本
=最新验证版本？
是
是
否
否
845
所需版本=事务开始
版本
850
所需版本=最新验证
版本号
检索所需页面
855
860
从本地缓冲池读取或
修改页面
870
将下一个页面
设置为当前页面
是
865
是否
有更多的页面需要处理
以进行操作？
否
875
操作完成

图8B