



(12) 发明专利

(10) 授权公告号 CN 1716248 B

(45) 授权公告日 2010.06.23

(21) 申请号 200510075473.9

审查员 韩燕

(22) 申请日 2005.05.31

(30) 优先权数据

10/879,665 2004.06.29 US

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 D·J·内特尔顿 S·J·贝利

(74) 专利代理机构 上海专利商标事务所有限公司 31100

代理人 李玲

(51) Int. Cl.

G06F 17/30 (2006.01)

(56) 对比文件

US 5544345 A, 1996.08.06, 说明书第2栏第33-38行, 第8栏第1-12行, 第17段第6-13行, 第29栏第2-52行, 第30栏第34-54行、图2.

US 5088036 A, 1992.02.11, 全文 .

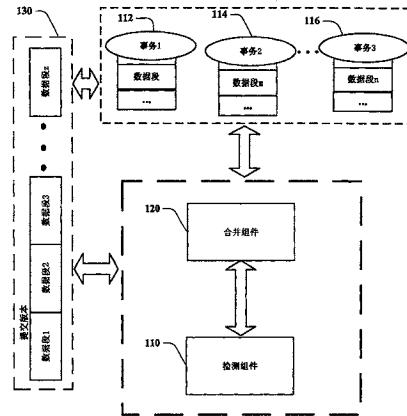
权利要求书 2 页 说明书 9 页 附图 11 页

(54) 发明名称

并发事务和页面同步

(57) 摘要

提供了由多个用户有效执行并发事务和在被修改的对象的物理层之下的逻辑层跟踪数据的系统和方法。每个事务可以将要修改的所述数据段的提交版本复制到其相应的空间，并在修改过程中更新这个副本。一检测组件，用于检测作为其他事务提交的结果，所操作的任何数据段是否需要更新，以及一合并组件，将所述数据段与其提交版本同步。在完成所述检测和合并进程后，各种优化过程可以被包含在内作为所述提交阶段的一部分。



1. 一种数据管理引擎系统,包括 :

一检测组件,所述检测组件被配置以 :

检测数据段的副本是否与所述数据段的提交版本不同;以及

确定是否有其它事务并发地在所述数据段的其它副本上操作;

一页面标记组件,所述页面标记组件被配置以在所述数据段的副本与所述提交版本不同的情况下,将所述数据段的其它副本标记为需要与所述提交版本合并;

一合并组件,所述合并组件被配置以在所述数据段的副本与所述提交版本不同的情况下,将所述数据段的副本与所述提交版本同步。

2. 如权利要求 1 所述的数据管理引擎系统,其特征在于,还包括锁管理器,所述锁管理器被配置以便于对所述数据段的子层锁定。

3. 如权利要求 2 所述的数据管理引擎系统,其特征在于,所述数据段是数据库的数据页面,且所述子层锁定是对于所述数据页面的行层次锁定。

4. 如权利要求 3 所述的数据管理引擎系统,其特征在于,还包括最优化特征,所述最优化特征被配置以对操作在所述数据段上的事务强制一个或多个条件,所述一个或多个条件包括确保在所述数据段上操作的事务没有消耗所述数据段的所有可用空间或确保这些事务没有在合并进程中重组数据或两者。

5. 一种数据库引擎系统,包括 :

一锁管理器,所述锁管理器被配置以在子页面层对数据页面进行锁定,以允许对于所述数据页面在所述子页面层的改变;

一检测组件,所述检测组件被配置以检测相比于所述数据页面的提交版本,是否已经有对于所述数据页面的副本的改变,并且,当检测到对于所述数据页面的副本的改变时,确定是否存在所述数据页面的其它副本,并且,如果存在所述数据页面的其它副本,就通过与其相关联的页面标记组件来将所述数据页面的其它副本标记为需要合并;以及

一合并组件,所述合并组件用所述数据页面的提交版本来更新所述数据页面的副本。

6. 一种执行并发事务的方法,所述方法包括 :

复制数据页面的提交版本,以创建所述数据页面的提交版本的相应副本;

接收对所述相应副本上的资源的一个或多个改变;

确定所述数据页面的提交版本的其它副本是否正在被其它事务同步地使用;

将所述其它副本标记为需要合并;

检测所述相应副本相对所述数据页面的提交版本是否保持最新;以及

将所述数据页面的相应副本与所述数据页面的提交版本合并。

7. 如权利要求 6 所述的方法,其特征在于,还包含锁定所述数据页面上的资源。

8. 如权利要求 7 所述的方法,其特征在于,还包含在合并时丢弃所述锁。

9. 一种便于并发事务中的同步的系统,所述系统包含 :

用于复制提交的数据页面的装置;

用于检测在所述提交的数据页面的副本内的更新的信息的装置;

用于确定是否存在所述提交的数据页面的其它副本的装置;

用于如果检测到更新的信息,将所述提交的数据页面的其它副本标记为需要与所述提交的数据页面合并的装置;以及

用于将所述提交的数据页面的副本上的所述更新的信息与所述提交的数据页面合并的装置。

10. 一种用于操作数据的方法,所述方法包含 :

将提交的数据页面复制到为事务所保留的空间;

检测所述提交的数据页面的副本上更新的信息;

确定在为其它事务所保留的其它空间内是否存在所述提交的数据页面的其它副本;

将所述其它副本标记为需要与所述提交的数据页面合并;以及

将所述提交的数据页面的副本内的更新的信息与所述提交的数据页面合并。

11. 如权利要求 10 所述的方法,其特征在于,还包含在所述提交的数据页面上工作的并发事务上跟踪所述提交的数据页面上的空间可用性。

12. 如权利要求 10 所述的方法,其特征在于,还包含在所述提交的数据页面上的行层次上锁定资源。

并发事务和页面同步

技术领域

[0001] 本发明涉及数据管理,尤其设备通过采用页面版本 / 复制方法对数据的物理段(例如在数据库管理系统的子页面层处)操作执行并发事务的系统和方法。

背景技术

[0002] 计算机技术(例如,微处理器速度、存储器容量、数据传输带宽、软件功能等等)增长的优势广泛地引起了各个行业中增长的计算机应用。例如,更为强大的服务器系统(通常被配置成服务器阵列)通常会被提供给源自诸如万维网等外部资源的服务请求。随着本地企业内部互联网系统变得更为复杂,由此要求对较大的网络负载和相关应用程序、内部系统要求的服务也相应地增长。

[0003] 一般地,计算机系统中的问题是处理增长的可用信息或数据量以及持续监视有多个用户输入的更新。以某种形式为数据库存储在磁盘或其它介质上的绝对信息量大量增长。虽然几十年前文件和磁盘以千字节来计量 - 那时是百万字节(兆字节),接着是亿万字节(千兆字节) - 现在创建出了百万兆字节(兆兆字节)甚至亿万兆字节的数据库并且使用在日常的活动中。

[0004] 这样,在数据管理系统(DBMS)的管理下,许多商业数据被存储在数据库中。对于这样的DBMS系统,在大型安装中对数据库事务处理容量的需求大量地增长。同时,所有新的数据库应用程序中的一大部分处于关系型数据库环境中。这样的关系型数据库可以进一步为支持对数据库各种形式的查询提供理想的环境。例如,在使用特设的未系统化伴随查询中的增长 - 是期望在这样的数据库环境中加速的趋势。这会导致对同时支持大量事务处理和对相同的数据库未系统化查询的增长的需求。因此,需要可以同时有效地支持大量事务和复杂查询且在两者之间干扰最小,而共享数据副本的系统和方法。

[0005] 一般地,商业条目通过大量相对简单的事务创建和维护它们的数据库,带有能够并发执行多个这样事务的数据引擎的基本功能。这样多个事务的每个可以被认为是一件作业,他可以又由一系列对数据库的操作,诸如读、些、更新和 / 或删除,完成。这样的事务可以表示很好理解的商业运作(例如,建立新的客户记录、标注一帐户付款或转帐等等)。公司越来越对运行更多对它们网上数据的特设的未系统化查询感兴趣。这样,纠正数据库状态或数据库完整性的方法就变得越来越重要。

[0006] 在一般的数据库中,完整性可以指保证两个或多个并发执行的事务不互相干扰,而导致在数据库中产生不正确状态的问题。一应用程序编制器一般应该保证每个事务产生一正确的状态,且当数据库处于正确状态时每个事务开始。这样的事务管理应该遵守ACID(原子性、一致性、隔离性和持久性)标准。然而,当创建了多个数据页面的副本,遵守这样的标准会是困难的。同时,数据库引擎越是有效地支持并发事务,终端用户在运行速度和降低数据库增长方面就能获得越好的体验。

[0007] 特别地,传统的影像页面技术一般不在数据修改的提交阶段提供有效的操作,并且对于并发业务的合并要求额外的开销。此外,采用日志记录一般会增加数据库管理和恢

复的复杂性。

[0008] 因此,需要克服上述与传统的系统和方法相关联的缺陷,所述的方法与数据库操作相关且一般与从多个用户接收数据的任何数据管理系统相关联。

发明内容

[0009] 以下提供了本发明的简述,以提供对本发明一个或多个方面的基本理解。该简述不是对本发明的扩展的概述。它并非旨在标识本发明的关键元素或描绘本发明的范围。本概述的唯一目标是以简化的形式提供本发明的一些概念,以作为后面提出的更为详细描述的序言。

[0010] 本发明提供了通过采用版本 / 复制方法执行并发事务的系统和方法,其中检测组件检测所操作的数据段是否由于其它处理委托需要更新,以及将数据段与其相关联的版本同步的合并组件。所述数据段可以是多个事务之间共享数据的物理单元和 / 或交换单元,诸如数据库中的数据页面。相应地,可以在数据页面的各种副本上进行并发事务操作,每个事务修改其相应的数据页面的副本,因而便利了多用户操作。应该理解虽然本发明主要在数据库事务的环境中描述,本发明不是这样地局限,可以被应用到任何数据管理系统,其中所期望的是在被修改对象的物理层之下的逻辑层跟踪数据。

[0011] 依照本发明的方法,事务 T1 要求在子页面层对数据页面的修改,开始将数据页面的最近提交 (committed) 版本 (“黄金复制”) 从数据库复制到其自己的空间中。在授权对修改所需子页面层 (例如,行层次) 的适当锁定后,事务 T1 开始要求修改所复制的副本。同时,可以提交修改相同页面的其他段的各种其它事务 (例如,不被 T1 锁定的其它行),因而改变由 T1 并发修改的数据页面的“黄金复制”。在 T1 提交之前,本发明检测到对“黄金复制”这样的改变,并将所述改变与 T1 自身的修改相合并,使得在提交阶段维持数据的一致性。相应地,更新黄金复制较早的副本并 / 或将其与较晚的修改同步。同样地,对于重读操作,(例如,必须读出 T1 正在修改的页面),对该页面的黄金复制的任何改变应该被合并到事务的空间中去,使得为重读过程采用这个页面合适的提交的读版本。此外,如果对事务 T1 作出决定不提交,那么就简单地丢弃由 T1 修改的页面的副本。

[0012] 在本发明的相关方面,可以提供页面标记组件用于通知对黄金复制的改变的事务。每个事务所操作的关于页面的信息可以被存储在数据库引擎中,并作为事务提交,关于提交版本的信息根据所需提供给其它事务。这样,如果对不同的页面进行并发事务的操作,那么一般不需要合并步骤,且页面标记组件不需要向其它事务提供信息,用于与页面最新提交的版本一致。相似地,如果事务回卷且不执行提交,一般无需对其它事务的页面标记。

[0013] 依照本发明的另一方面,可以引入各种最优化特征以有效地执行合并过程。例如,可以强制条件使得在事务的提交阶段之前 (例如,在“加固”事务作出的对数据库的修改之前),可以保证对特定页面的空间可用性 (例如,对页面各种版本操作的事务不消耗该页面所有的存储空间) 和减轻对页面上数据的重组 (例如,事务无需为了合并各种副本而将数据在页面上来回移动)。这种框架使得可以有效和简单地启用提交操作,这改进了多用户操作且保存了系统资源。

[0014] 在本发明的相关方面,可以提供旋转锁系统使得只有在一个时间只有一个用户可以提交的机制。一般在事务达到提交状态可以采用这种互斥组件,用于在数据库操作中提

供一致性。

[0015] 为了完成上述和相关的目标,本发明于是包含下面详细描述的特征。以下描述和附图详细提出了本发明的某些说明性方面。然而,这些方面只是指示了本发明的原理可以采用的各种方法中的一些方法。结合附图,本发明的其它方面、优势和新颖性特征将通过以下对本发明的详细描述变得显而易见。

附图说明

- [0016] 图 1 是依照本发明的一方面,数据管理引擎的部分的框图。
- [0017] 图 2 是依照本发明的一方面,作为具有锁管理器的数据库环境的部分的数据管理引擎的框图。
- [0018] 图 3 说明了依照本发明的另一方面,由锁管理器管理的示例性锁分层结构。
- [0019] 图 4 说明了依照本发明的一方面,采用并发事务的客户端服务器网络的框图。
- [0020] 图 5 是依照本发明的一方面,带有页面标记组件的数据库引擎的组件的框图。
- [0021] 图 6 说明了依照本发明的一个方面的特定方法。
- [0022] 图 7 是依照本发明的另一方面的示例性方法。
- [0023] 图 8 是依照本发明的一方面,用于检测和合并方法的示例性流程图。
- [0024] 图 9 是依照本发明的一方面,用于作为提交阶段的一部分的检测和合并过程的另一示例性方法。
- [0025] 图 10 是说明可以采用本发明各个方面的适当的计算环境的原理图。
- [0026] 图 11 说明了依照本发明的一方面,采用数据修改方法的客户端 - 服务器系统。

具体实施方式

[0027] 现在参考附图描述本发明,在整个附图中相同的参考编号指向相同的组件。在以下描述中,为了解释,提供对本发明的一个彻底的理解,许多特殊的细节将被阐述。然而,很显然,可能在没有特殊描述的情况下实现本发明。在另一个例证中,为了更容易的描述本发明,以框图形式示出了公知的结构和设备。

[0028] 如本发明中所使用的,术语“组件”、“处理程序”、“模型”、“系统”等等意在指计算机相关实体,可以是硬件、硬件和软件的组合、软件或执行中的软件。组件可以是但不限于运行在处理器上的进程、处理器、对象、可执行件、执行的线程、程序和 / 或计算机。为了说明,运行在服务器上的应用程序和服务器可以是组件。一个或多个组件可以驻留在执行的进程和 / 或线程中,组件可以局限在一个计算机上和 / 或分布在两个或多个计算机之间。同样的,可以从各种计算机可读介质执行这些组件,所述计算机可读介质具有存储于其上的各种数据结构。这些组件可以经由本地和 / 或远程进程通信,诸如依照具有一个或多个数据分组的信号(例如来自于一个组件的数据,所述组件通过信号与本地系统、分布式系统和 / 或通过诸如因特网等网络与其他系统中的另一个组件交互)。

[0029] 本发明提供了采用检测组件和合并组件,用于在预定义的数据段(例如在数据库中子页面层处)上执行并发事务的系统和方法。每个事务将数据段在其最近提交版本(“黄金复制”)处从数据存储副本到事务自身的空间中。检测组件可以基于由其他事务执行的修改检测所操作的数据段是否需要更新。应该理解虽然本发明主要在数据库事务的背景中

描述,本发明并不局限于此并可以被应用到任何数据管理系统,其中多个用户要求对数据存储空间的数据段的改变。

[0030] 首先参见图 1,说明了依照本发明的一个方面的数据段引擎的示例性部分,包含检测组件 110 和合并组件 120,带有运行于其上的事务 112 到 116。除了由事务所创建的任何数据段,事务 112 到 116 中的每个可以在其相应的数据段副本上运作(例如数据页面的副本形成数据库),所述的数据段副本是从数据段的提交版本复制而来的(i、m、n 和 z 是整数)。这些操作可以例如包含对数据库每个页面中数据的行层次操作。一般的本发明的数据管理引擎可以是多用户引擎,所述多用户引擎可以进一步使得多个用户在数据段的数据存储中(未示出)同时执行操作。当事务要求修改数据段中的数据时,首先将数据复制到为该事务所保留的空间。接着事务可以操作这样的副本数据段和修改所需数据。

[0031] 检测组件 110 可以在多个并发事务 112 到 116 上操作,以检测从共享数据段的提交版本(“黄金复制”)复制而来的数据段中的任何一个数据段是否需要更新,举例而言,在以下情况需要这种更新:当事务 T_1 从数据段最新的提交版本(黄金复制)复制了数据段,另一个事务处理了数据段并由此修改了黄金复制。为了维护数据段存储空间的一致性,现在需要更新事务 T_1 以反映出对黄金复制的这种改变,这是由于另一个事务的处理。这种更新可以由合并组件提供,所述合并组件可以对多个并发事务 112、114 和 116 进行操作,其中 N 是整数。

[0032] 例如,在任何给定的时间多个操作 1 到 N(N 是整数)可以期望对数据存储空间特定数据段的修改。这种修改可以包含在特定时间对数据的改变、插入数据字段、删除数据字段等等。接着,事务可以将数据段复制到所保留的空间,每个事务在其数据页面相应副本中插入行或删除行。相应地,检测组件 110 接着可以检测发生在特定数据段上的黄金复制上的改变,并结合合并组件 120 操作使得在事务提交阶段之前,要提交的页面被适当地更新了。

[0033] 图 2 依照本发明的特定方面,说明了数据库引擎的部分的示例性方面,所述数据库引擎与锁管理器 230 交互。锁管理器 230 可以确定是否可以授权特定资源上的锁,且一般很好地适于管理者子页面锁定(例如行层次锁定),使得事务 T_1 到 T_n 的每一个可以在数据页面的相应副本上操作。由于在任何给定的时间,可以存在数据页面的许多副本,锁管理器 230 一般可以确保并发时间可以修改相同数据页面但不同行中的信息。例如,锁管理器 230 可以为一个时间为特定行授权专用锁,其他时间于是被限制修改这行,即使仍然允许修改其他的行。因此,基于操作正常的并发性,锁管理器 230 可以授权或拒绝对特定资源的锁。

[0034] 依照本发明的特定方面,锁管理器 230 可以独立于数据库引擎的检测组件 210 和合并组件 220。在相关的方面,额外的信息存储在锁管理器中(例如,在特定位置授权专用锁、清空空间可用性等等)并可以由检测组件 210 采用以便跟踪哪个事务 212-216 正在数据库的哪个黄金复制上工作。相应地,在数据页面上各种事务的修改发生在数据页面的各自副本上,监测器组件 210 结合合并组件 220 可以保存事务 212-216 的 ACID 特性,其中锁管理器 230 一般可以保证在逻辑上允许事务修改特定资源。这便利了在多个用户修改数据库时,在次页面层(例如,行层)并发事务的操作。此外,在数据页面读取操作时(例如,不需要对数据页面的修改),一般不需要请求共享锁,因为可以从数据页面的提交状态请求读

取。此外,如果事务要求读操作,所述事务具有任何数据页面的副本,那么读操作也无需共享锁,因为可以使用提交页面和 / 或任何页面的事务版本来完成读取(一般可以优先于提交页面使用处理页面,假设没有在其他地方修改页面)。

[0035] 图 3 说明了由锁管理器 230 授权的示例性锁分层结构。如锁定粒度中所示的,为要求修改数据库 2 的表 2 的数据页面 2 的行 3 的事务 T 授权一专用锁。这样,数据页面 2 的副本可以被复制到事务 T 的保留空间(未示出),事务 T 可以接着继续进行其修改。授权给行 3 的这种专用锁对其他事务可以是可视的,一般其他事务不能在数据页面 2 的相应副本上修改这行,直至 T 提交(或回滚)其修改且锁被释放。

[0036] 图 4 说明了依照本发明的一方面,多个客户机要求修改数据库,其中在每个客户机 420 上运行的可以是客户机进程,例如 web 浏览器 410。同样的,在服务器 450 上运行的可以是相应的服务器进程,例如 web 服务器 460。此外,嵌入在 Web 浏览器 410 中的可以是脚本或应用程序 430 和运行在客户机计算机 420 的运行时间环境 440 中可能存在代理 415, 用户打包和解包格式化的数据分组。与服务器 450 通信的是数据库管理系统(DBMS)480, 其管理到数据库(未示出)的接入。DBMS480 和数据库(未示出)可以位于服务器自身中, 或可以远程地位于远程数据库服务器(未示出)中。运行在 Web 服务器 460 上的可以是数据库接口应用程序编程接口(API)470, 它提供到 DBMS 480 的接入。客户机计算机 420 和服务器计算机 450 可以通过网络 490 互相通信。应该理解其他安排也是可能的, 例如客户端计算机和服务器计算机是相同的计算机。当客户端进程, 例如 Web 浏览器 410, 从数据库请求数据, 脚本或应用程序 390 发出查询, 所述查询通过网络(例如, 因特网)490 发送到服务器计算机 450, 它在服务器计算机 450 中由服务器进程翻译, 例如 Web 服务器 460。客户机 420 向服务器 450 的请求可以包含多个命令, 来自服务器 450 的响应可以返回多个结果集合。

[0037] 在这种通信中, 可以由表格数据流(TDS)来提供会话、表示和应用服务元件。由于 TDS 无需任何特定的传输提供者, 它可以在多个传输协议和网络 490 上实现。返回的对客户机命令的响应可以是自我描述且是面向记录的;(例如, 数据流可以描述名称、类型和被返回的行的可任选描述)。

[0038] 在客户机 420, 数据可以是结构化查询语言(SQL)命令, 它是服务器端 450 可以接收的语言, 后面是其相关联的二进制数据的 SQL 命令(例如, 用于大量副本的命令的数据)或提醒信号。当期望一连接时, 客户机 420 可以发送连接信号到服务器。即使客户机 420 可以有多于一个到服务器 450 的连接, 可以分别地且以相同方式建立每个连接路径。

[0039] 一旦服务器 450 接收到来自客户机 420 的连接信号, 它会通知客户机它接收或拒绝了连接请求。如同发送一个 SQL 命令或一批 SQL 命令; 接着 SQL 命令(例如由统一码(Unicode)格式表示)可以被复制到缓冲器的数据部分并接着发送到 SQL 服务器侧 450。此外, 各种开放式数据库连接(ODBC)例行程序会造成 SQL 命令被置入客户机消息缓存或会造成消息缓存被发送到服务器。一旦期望修改并先于事务的提交阶段, 一般可以确保特定页面的空间可用性(例如, 在页面各种副本上操作的事务不会消耗该页面所有的存储空间), 可以减缓对页面周围数据的重组(例如, 事务无需为了合并各种版本而将数据在页面上到处移动)。这种框架使得可以有效和简单地发生提交操作, 由此改进了多用户操作和保存系统资源。

[0040] 图 5 说明了依照本发明的一个方面数据库引擎的部分的示意性框图, 还进一步包

括页面标记组件 530。一般地,在事务 T1 对特定页面操作时,事务 T1 必须能够检测在复制 T1 后是否对该页面的黄金复制有改变。依照本发明的一个方面,通过页面标记组件 530 可以便利这种检测,所述页面标记组件 530 与检测组件 510 结合运作。页面标记组件 530 可以将对黄金复制的改变通知事务。每个事务所操作的页面的信息可以存储在数据库引擎中并作为事务提交,关于提交版本的信息根据需要被提供给其他事务。因此,如果并发事务对不同的页面进行操作,那么无需合并步骤,且页面标记组件一般无需向其他事务提供信息,用于与最近提交的页面版本保持一致。相似地,如果事务回卷且不执行提交,一般无需其他事件的页面标记。

[0041] 相应地,当事务 T1 提交时,就作出确定对其他事务有什么影响 - 例如,另一个事务 T2 可以对 T1 修改的相同页面的另一个副本操作,如果这样,会分配标记给 T2 所操作的副本。关于每个事务 512-516 操作的页面的信息可以被存储在数据库引擎中。在事务 512 提交其更新前,会对其检查以确定是否需要合并。这也可以通过检查由页面标记组件分配的相关标记和 / 或通过检查其他事务是否修改了相同的页面来完成。这样,如果并发事务 512-516 对不同的页面进行处理,那么一般无需合并步骤。同样地,如果要求事务 512-516 的任何一个回卷,那么简单地丢弃相应的页面副本,而其他由该事务标记的事务页面(现在被回卷)可以仍旧执行与提交版本的合并。

[0042] 图 6 说明了可以与本发明的合并过程结合使用的最优化特征。并发事务 T1 可以对数据页面 610 的副本进行操作。数据页面 610(黄金页面) 开始包含行 A。在 611 处,T1 开始对数据页面 610 的副本的操作,以在 612 处插入行 B 和 C。在 612 处的这个插入操作之前,可以引入最优化特征以检查和 / 或强制两个条件 660, 即 : 由于其他事务修改这个页面,页面上空间的可用性(页面未满) 和页面无需重组。这样,可以在插入操作和 / 或选择在其上操作的页面之前确定数据页面上消耗的空间和期望的空间可用性(例如,甚至在复制这样的页面之前,可以作出确定是否由足够的空间可用于执行修改)。因此,除了对子页面修改的逻辑考虑之外(例如,是否逻辑上允许在特定位置插入行),本发明的优化特征也可以考虑该页的物理特性(例如,在插入时,在特定页面上是否有空间可用)。随后,在 612 处插入行 B 和 C。这种对行 B 和 C 的插入会占据这样的数据页面副本上的可用空间,如 620 处所描述的。由此有效而简单地发生 T1' 提交操作 616, 而同时提高多用户操作和保存系统资源。然而上述的优化特征无需用于实践本发明的其他方面,它是用来提高性能。特别地,不符合“页面未满”条件一般要求数据库引擎找到另一个页面以存储数据。在相关方面,T1 应回卷其修改并释放该数据页面上的空间,这种附加空间的可用性可以在所有事务的过程中示出。回卷也可以丢弃相关联的锁(例如专用行锁),这样使得较早存储在空间上的、要被现在回卷事务消费的信息被丢弃同时消除相关联的锁。

[0043] 图 7 说明了依照本发明的一个方面的示例性方法。方法 700 从请求修改来自数据存储区域的数据的事务 T1 开始,例如在数据库中数据页面“A”的子页面层处修改。在数据库环境的特定例子中,事务 T1 可以是多个并发事务的一部分,所述多个并发事务作为多用户数据库引擎的部分运作。在 710 处事务 T1 对数据操作的请求可以包含在特定数据页面上在特定时隙处对数据的改变、用更新的行替换行、插入行、删除行等等。事务 T1 接着可以在 720 复制数据页面“A”以保存空间,如副本 A1。在 730 处,作出的检测是数据页面“A”的黄金副本已经由其他事务修改。随着合并需求,这种检测可以由例如检测组件作出,页面 A1

可以由页面标记组件标记。随后，在 740 处，用新的改变将页面 A1 合并和更新到所提交的副本。事务 T1 于是可以提交，而维持数据库的一致性，在那里没有数据丢失。

[0044] 图 8 说明了依照本发明的一个方面的相关方法，其中依照本发明的一个方面在 810 处要求事务重读数据页面。在事务的生存期内，这种重读会在例如事务对相同数据执行多于一次的“SELECT”操作时发生 - 事务在读取提交隔离层运作。在 820 处，取得旋转锁，允许在任一时间只有一个事务合并的机制。这种互斥组件可以一般被用于确保数据操作中的一致性。在 830 处，检测方法确定页面是否需要与对数据页面的“黄金复制”作出的新改变合并，这可以在多个并发事务之间共享。如果不需要，在 840 处释放旋转锁，操作在 890 处结束。另一方面，如果确定需要合并页面，那么方法继续到合并进程，页面的提交版本于是在 860 处被复制到事务的相应空间。随后，作为合并进程的附加部分，对于每个在页面上的改变，在 870 处更新相关的提交页面。接着在 880 处用提交页面的更新副本替换事务页面。

[0045] 现在参见图 9，说明了依照本发明的一个方面的相关方法，其中概述了作为提交程序 910 的部分检测和合并改变。在 920 处，获得旋转锁，允许在同一时间只有一个用户提交的机制，且确保了数据库操作中的一致性。在 925 处，为检测和合并进程选择已由事务修改的下一数据页面。随后，开始检测进程，在 927 处确定页面是否被其他事务使用。如果是，方法继续到 929，其中被其他页面使用的该页面的副本被标记用于合并，例如通过如上所述的页面标记组件。否则，在 930 处，通过检测方法确定页面是否需要和对可以在多个并发事务之间共享的数据页面的“黄金复制”的新改变合并。如果需要合并该页，那么方法继续到合并过程，且页面的提交版本于是在 960 处被复制到事务的相应空间。随后，作为合并进程的附加部分，对于每个在页面上执行的改变，在 970 处更新相关的提交页面。接着在 990 处用提交页面的更新版本代替事务的页面，随后是在 940 处释放锁。另一方面，如果在 930 处，确定无需合并页面，方法直接继续到在 940 处为页面释放锁。随后，在 945 确定是否达到了由事务操作的最后页面。如果否，操作返回到步骤 925，否则操作在 947 处释放旋转锁并在 955 处结束。

[0046] 现在参见图 10，说明了合适的计算环境的客户机和服务器端的简要、一般描述，其中可以实现本发明的各个方面。虽然以上在运行在一台计算机和 / 或一些计算机上计算机程序的计算机可执行指令的一般环境中描述了本发明，本发明的技术人员应该认识到也可以结合其它程序模块实现本发明。通常，程序模块包含执行特定任务和 / 或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等。此外，本领域的技术人员应该理解本发明可以用其它的计算机系统配置实现，包含单处理器或多处理器计算机系统、小型机、大型计算机以及个人电脑、手提计算设备、基于微处理器或可编程消费电子产品等等。如上所述，本发明的说明方面也可以在分布式计算环境中实现，其中在分布式计算环境中，任务通过一个通信网络连接的远程处理设备执行。然而，本发明的一些（如果不是全部的话）方面可以在单机上实现。在分布式计算环境中，程序模块可以位于本地和远程存储器存储设备中。该示例包含计算机 1020，包括处理单元 1021、系统存储器 1022 和将包含系统存储器的各种系统组件耦合到处理单元 1021 的系统总线 1023。处理单元 1021 可以是各种商业上可用的处理器中的任何一种。双微处理器或其它多处理器构架也可以被用作处理单元 1021。

[0047] 系统总线可以是多种类型的总线结构中的任何一种，包含存储器总线或存储器控

制器、外设总线和使用多种商业上可用总线结构中任何一种的本地总线。系统存储器可以包含只读存储器 (ROM) 1024 和随机存取存储器 (RAM) 1025。包含帮助在诸如启动期间在计算机 1020 内部的组件之间传送信息的基本例行程序的基本输入 / 输出系统 (BIOS) 存储在 ROM 1024 中。

[0048] 计算机 1020 还包含硬盘驱动器 1027、磁盘驱动器 1028, 例如从可移动磁盘 1029 读出和写入, 和光盘驱动器 1030, 例如, 从 CD-ROM 磁盘 1031 读出和写入或从其它光介质读出或写入。硬盘驱动器 1027、磁盘驱动器 1028 和光盘驱动器 1030 分别通过硬盘驱动器接口 1032、磁盘驱动器接口 1033 和光盘驱动器接口 1034 连接到系统总线 1023。这些驱动器和它们相关联的计算机可读介质为计算机 1020 提供对数据、数据结构、计算机可执行指令等的非易失性存储。虽然上述对计算机可读介质的描述指向硬盘、可移动磁盘和 CD, 本领域的技术人员应该理解其他类型的计算机可读介质, 诸如磁带盒、闪存卡、数字视频盘、Bernoulli 盒式磁盘等等, 也可以被用于示例性操作环境, 所述的介质还可以是任何包含用于执行本发明的方法的计算机可执行指令的介质。

[0049] 多个程序模块可以被存储在驱动器和 RAM 1025 中, 包含操作系统 1035、一个或多个应用程序 1036、其他程序模块 1037 和程序数据 1038。所述计算机中的操作系统 1035 实质上可以是任何商业上可用的操作系统。

[0050] 用户可以通过键盘 1040 和诸如鼠标 1042 等指点设备将命令和信息输入到计算机 1020 中。其他输入设备 (未示出) 可以包括麦克风、操纵杆、游戏手柄、卫星天线、扫描仪或类似的其他设备。这些和其他输入设备通常通过耦合到系统总线的串行端口 1046 连接到处理单元 1021, 但可以通过其他接口连接, 诸如并行端口、游戏端口或通用串行总线 (USB)。监视器 1047 或其他类型的显示设备也可以通过诸如视频适配器 1048 等接口连接到系统总线。除了监视器, 计算机一般包含其他外围输出设备 (未示出), 诸如扬声器和打印机。

[0051] 计算机 1020 可以在使用到一个或多个远程计算机 (诸如远程计算机 1049) 的网络环境中运作。远程计算机 1049 可以是工作站、服务器计算机、路由器、对等设备或其他通用网络节点, 且一般包含许多或全部相对计算机 1020 描述的元素, 虽然图 10 中只示出了存储器存储设备 1050。图 10 中所述的逻辑联接, 可以包含局域网 (LAN 1051) 和广域网 (WAN 1052)。这些网络环境在办公室、企业范围的计算机网络、内部互联网和因特网是很平常的。

[0052] 当在 LAN 网络环境中实现时, 计算机 1020 可以通过网络接口或适配器 1053 连接到局域网 1051。当在 WAN 网络环境中实现时, 计算机 1020 一般包括调制解调器 1054, 和 / 或, 被连接到 LAN 上的通信服务器, 和 / 或具有其他在广域网 1052, 诸如因特网上建立通信的装置。调制解调器 1054, 可以是内置的或外置的, 可以经由串行端口接口 1046 连接到系统总线。在网络环境中对于计算机 1020 描述的程序模块或其部分可以被存储在远程存储器存储设备中。应该理解所示的网络连接是示例性的, 可以采用其他在计算机之间建立通信链接的装置。

[0053] 依照计算机编程领域技术人员的实现, 本发明参考由计算机 (诸如计算机 1020) 执行的操作的动作和符号表示来描述的, 除非另外指出。这些动作和操作有时被称为是计算机可执行的。应该理解所述动作和用符号表示的操作包含电子信号的处理单元 1020 的

操作,所述的电子信号表示造成电子信号表示的转变或减少的数据比特,和在存储器系统(包含系统存储器 1022、硬盘驱动器 1027、软盘 1029 和 CD-ROM1031)中的存储位置上对数据比特的维护,由此重新配置或改变计算机系统的操作和其他信号的处理。维护这些数据比特所在的存储器位置是对应于这些数据比特有特定电子、磁性或光特性的物理位置。

[0054] 现在参见图 11,说明了依照本发明的一个方面,采用数据操作方法的客户机 - 服务器系统 1100。客户机 1120 可以是硬件和 / 或软件(例如线程、进程、计算设备)。系统 1100 也包含一个或多个服务器 1140。服务器 1140 也可以是硬件和 / 或软件(例如线程、进程、计算设备)。例如这些服务器 1140 可以是通过采用本发明执行这些转换的内部线程。客户机 1120 和服务器 1140 可以在两个或多个计算机进程之间依照本发明发送的数据分组的形式通信。客户机 / 服务器也可以共享同一进程。如所示的,系统 1100 包含可以便利客户机 1120 和服务器 1140 之间通信的通信构架 1180。客户机 1120 在操作上被连接到一个或多个客户机数据存储 1110,所述客户机数据存储 1110 可以存储对客户机 1120 而言为本地的信息。此外客户机 1120 可以接入和更新位于运行服务器进程的服务器计算机 1140 上的数据库 1160。在本发明的一个方面,通信框架可以是因特网,其中客户机进程为 Web 浏览器而服务器进程为 Web 服务器。这样,一般的客户机 1120 可以是通用计算机,诸如具有中央处理器(CPU)、系统存储器、调制解调器或用于将个人计算机连接到因特网的网卡和显示器以及诸如键盘、鼠标等其他组件的个人计算机。同样的一般的服务器 1140 可以是大学或公司的大型计算机或专用工作站等等。

[0055] 虽然关于某些所示的方面示出和描述了本发明,应该理解在阅读和理解了本规范和附图后,本领域的技术人员会作出等价的变化和修改。尤其对于上述组件(部件、设备、电路、系统等等),用于描述这些组件的这些术语(包括“装置”)意在对应于(除非特别指出)执行上述组件指定功能的任何组件(例如在功能上等价),即使在机构上不等价于所揭示的结构,所述组件执行这里所示的本发明示例性方面的功能。在这点上,也应该认识到本发明包含具有用于执行本发明的各种方法的动作和 / 或事务的计算机可执行指令的系统和计算机可读介质。此外,就具体实施方式或所赋权利要求所使用的术语“包含”、“具有”及其变形,这些术语只在以类似于术语“包括”的方式表示包含在内。

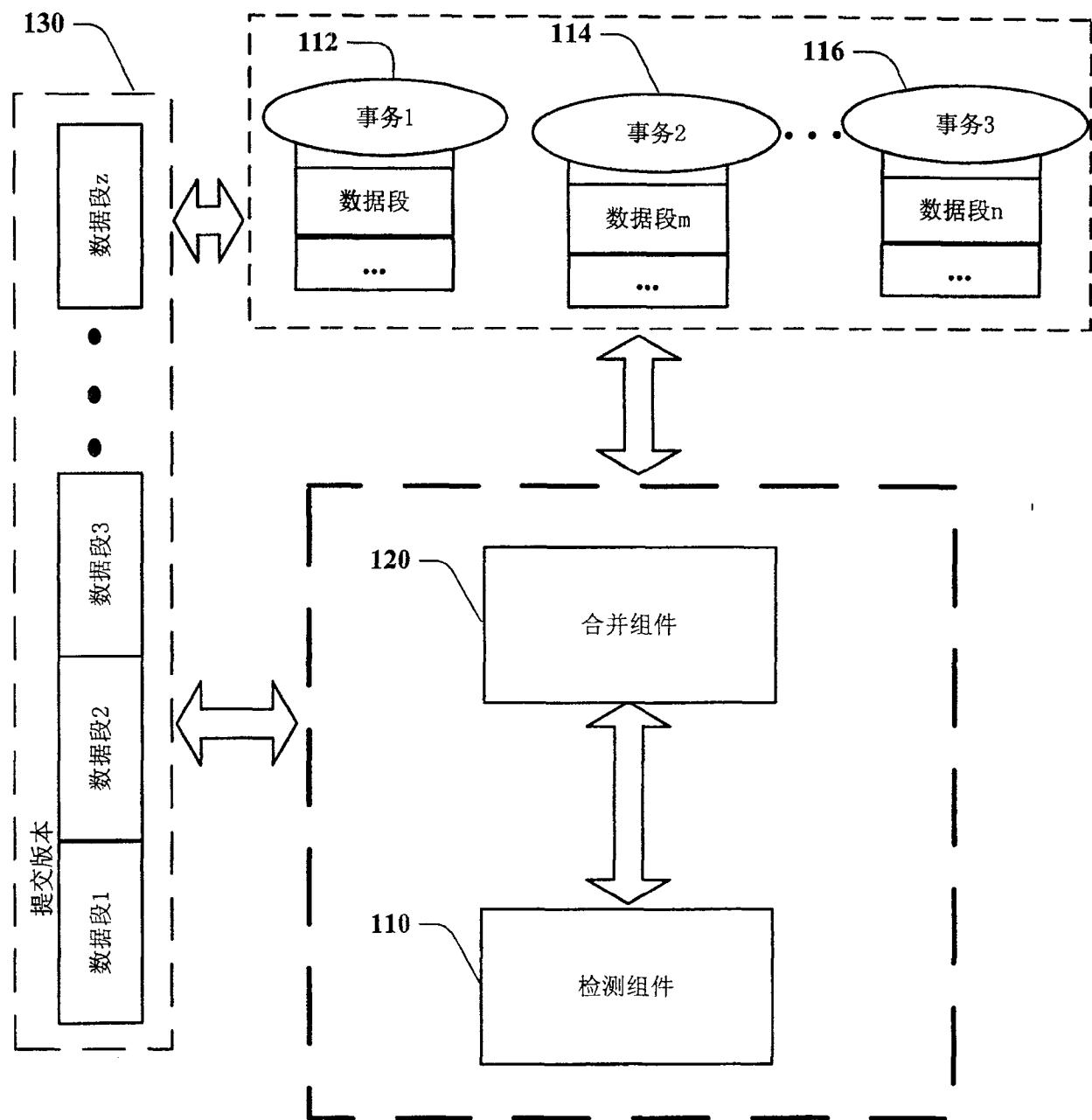


图 1

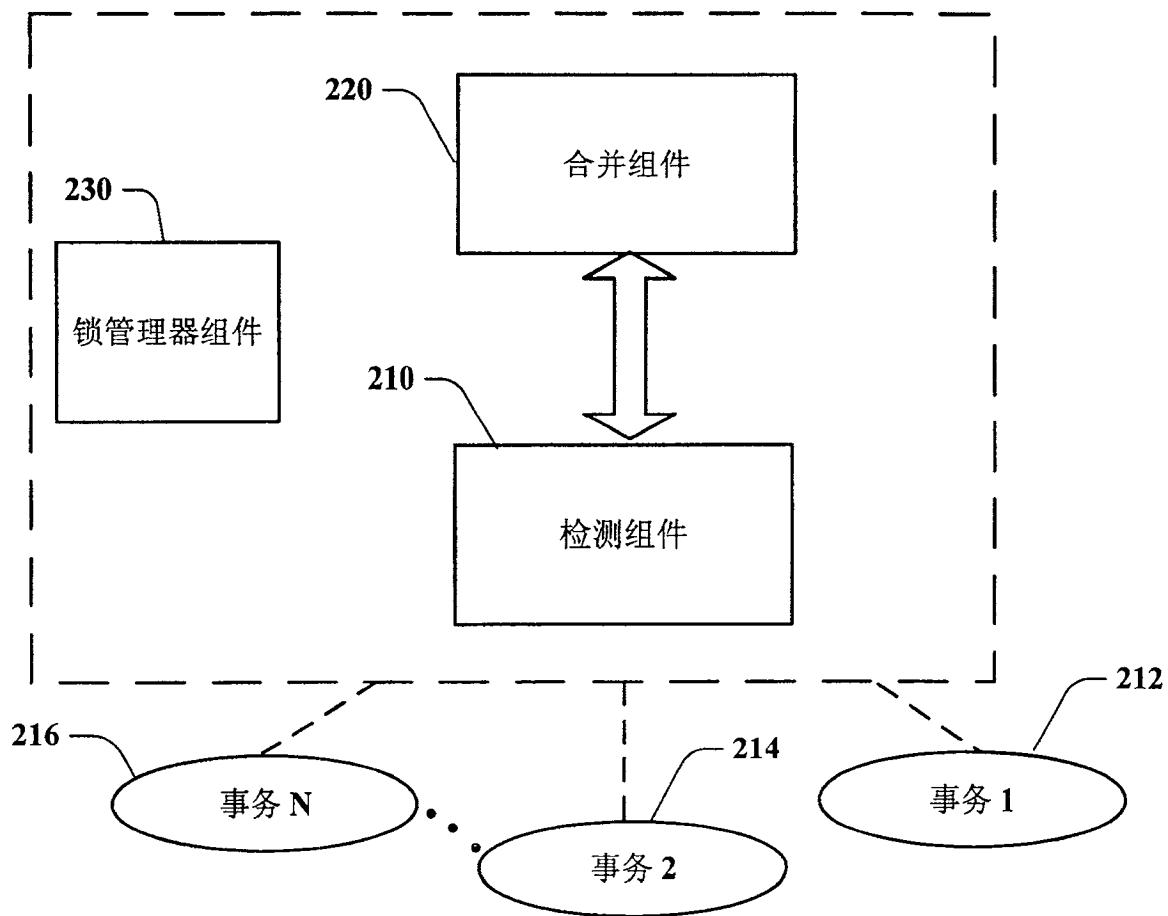


图 2

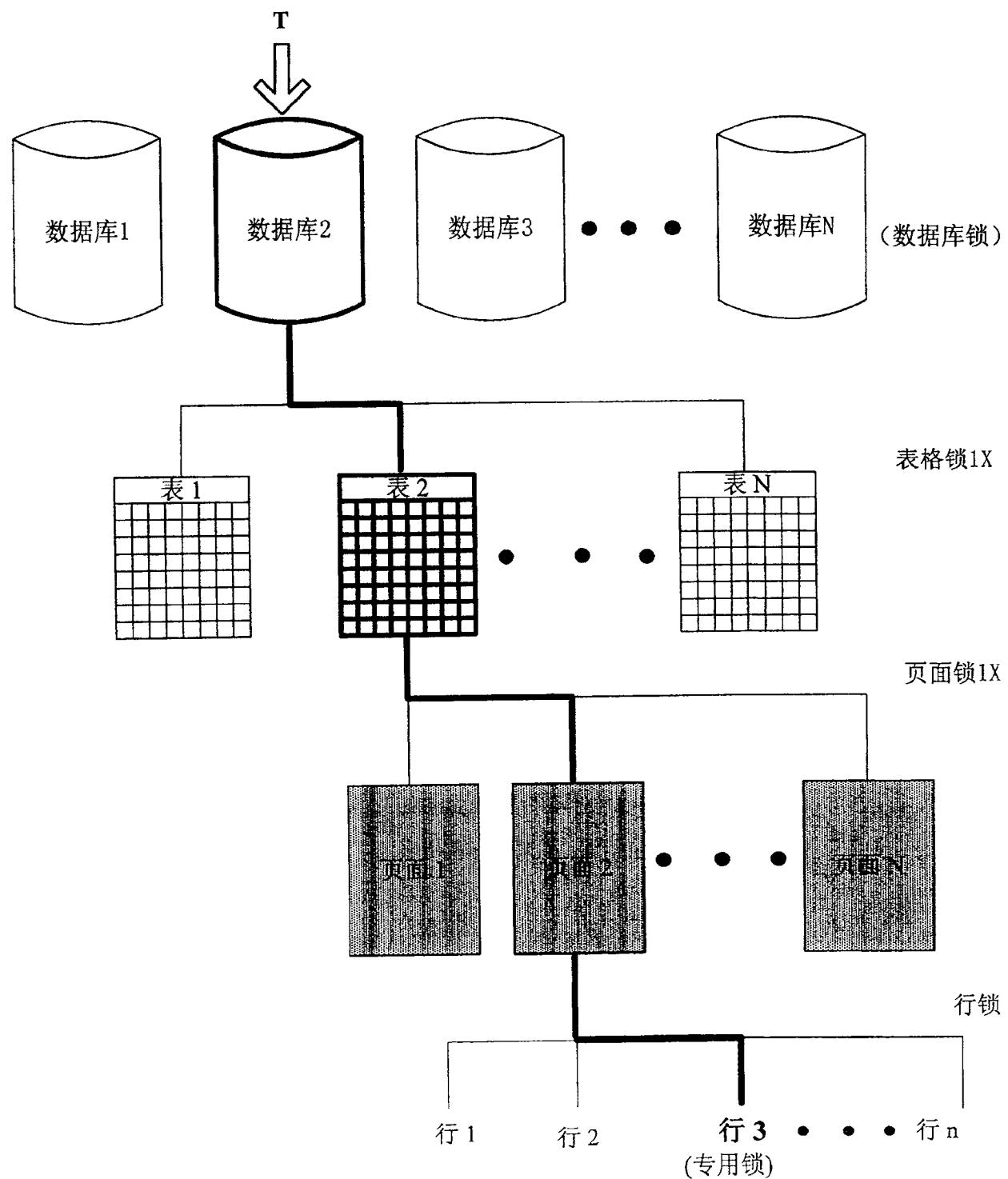


图 3

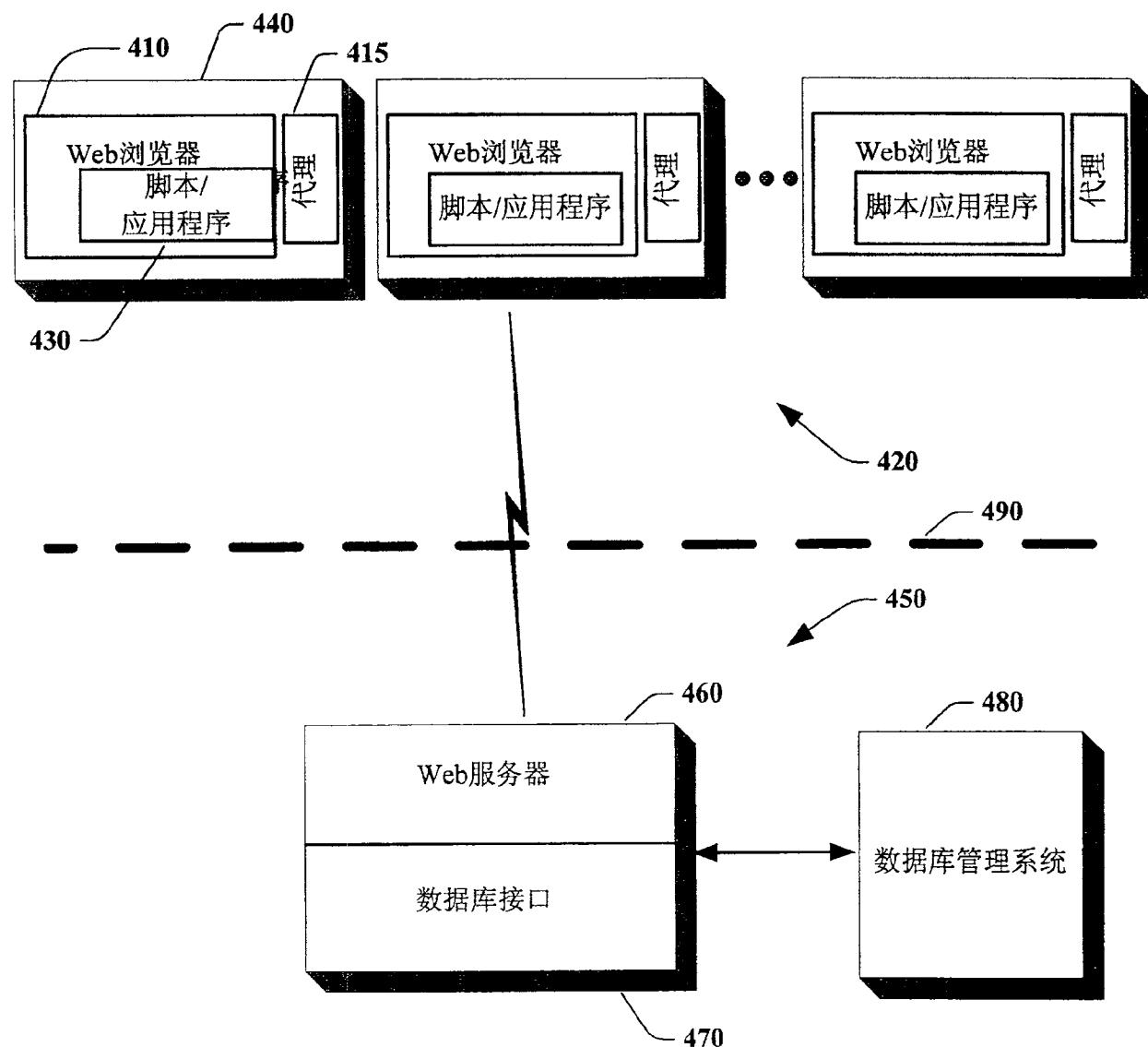


图 4

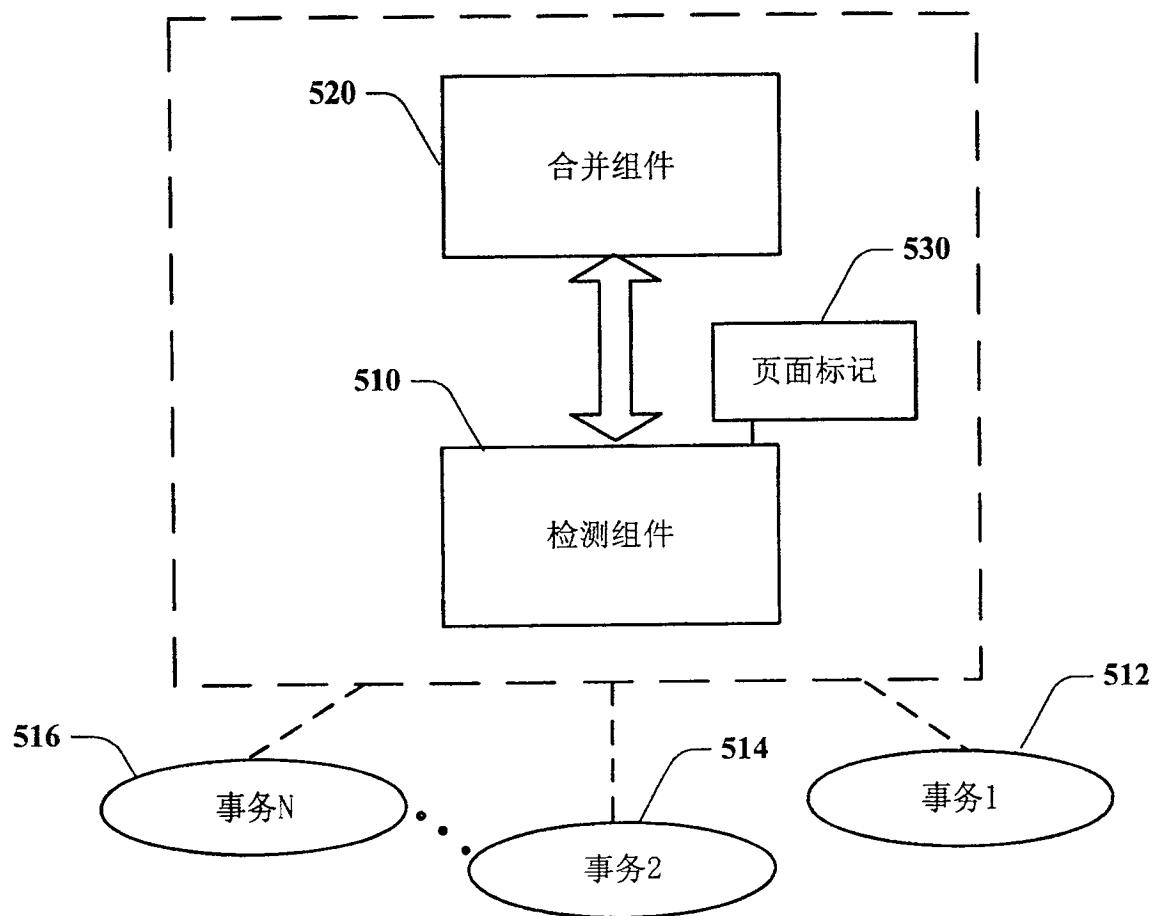


图 5

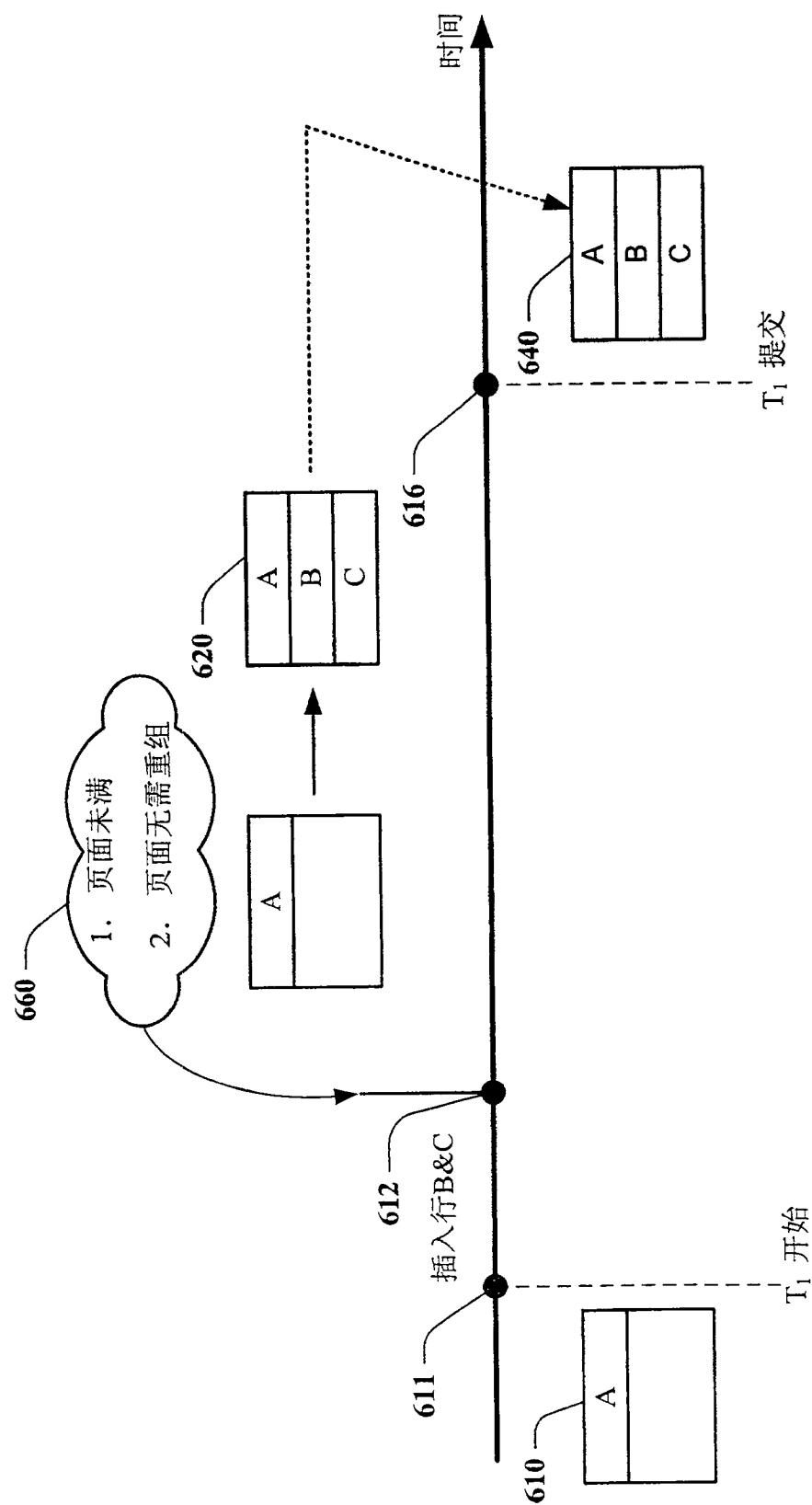


图 6

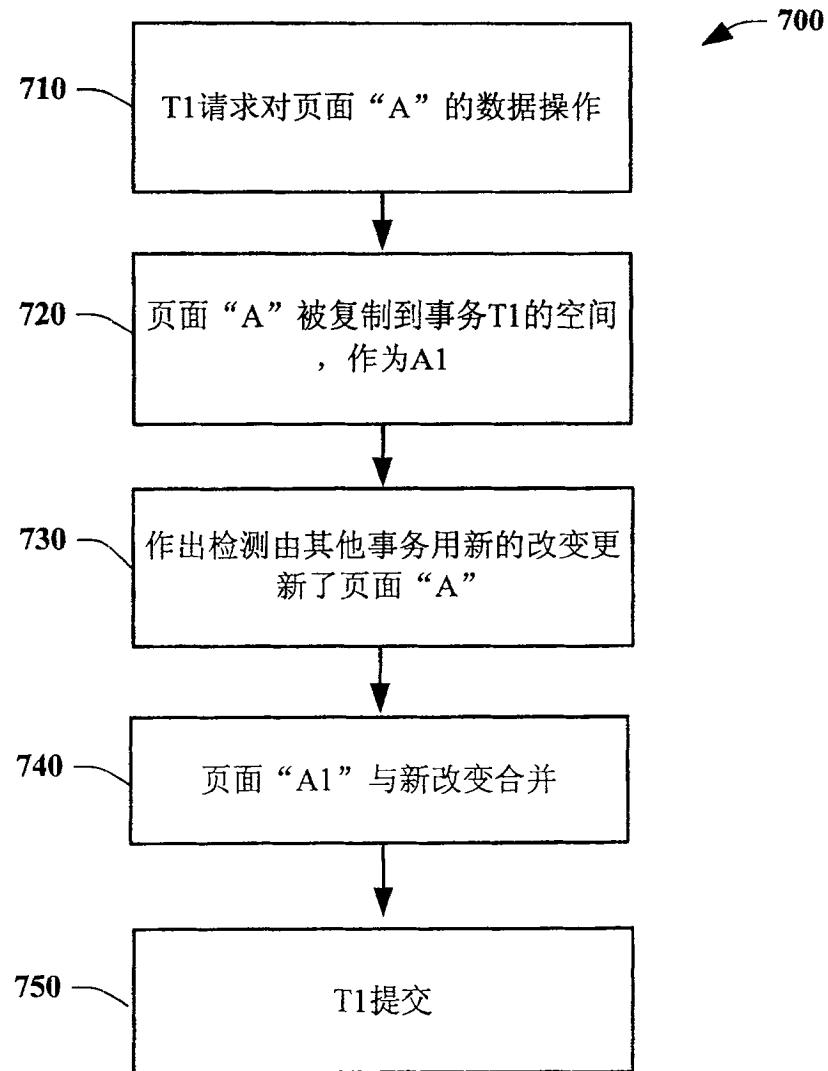


图 7

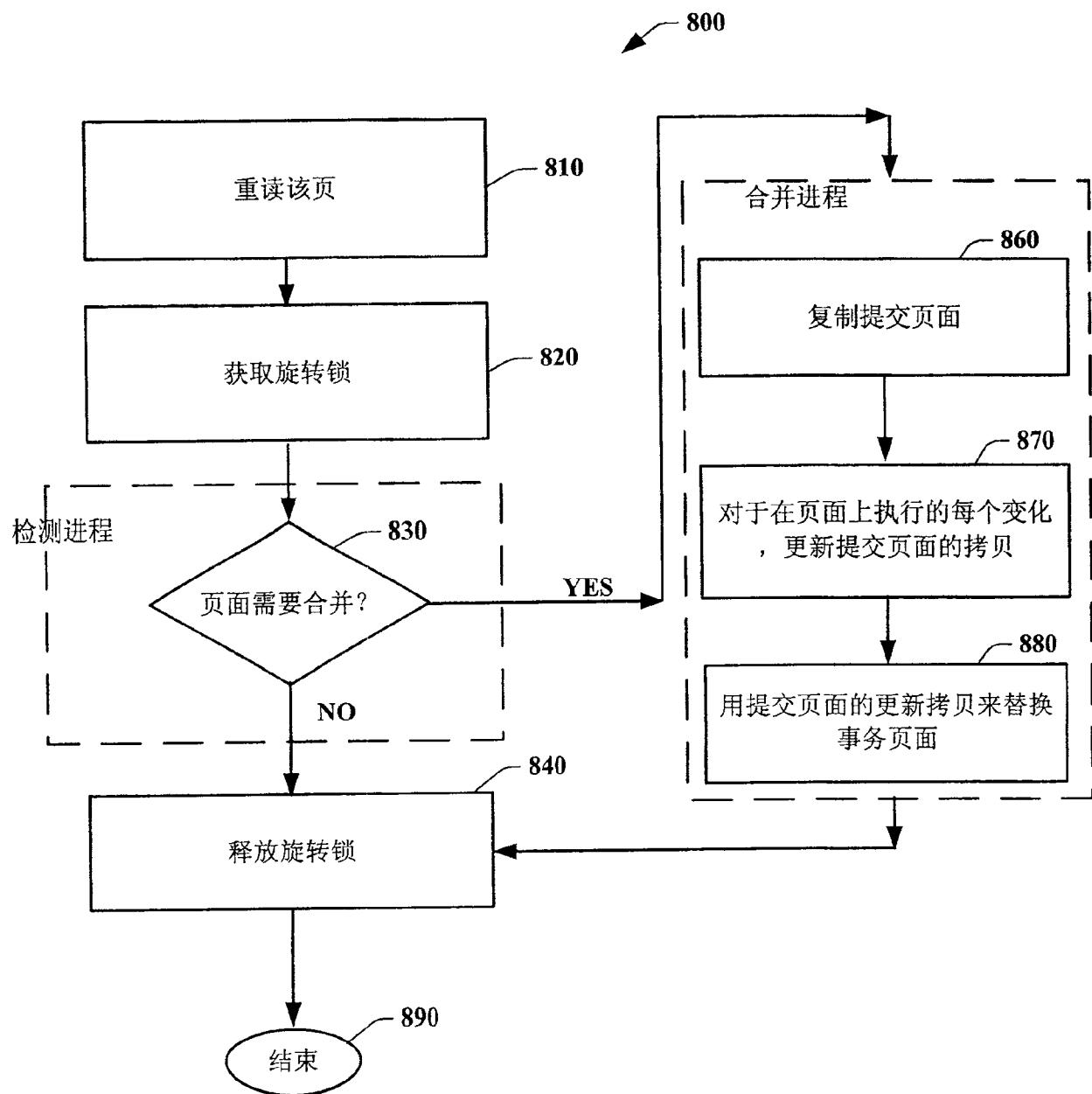


图 8

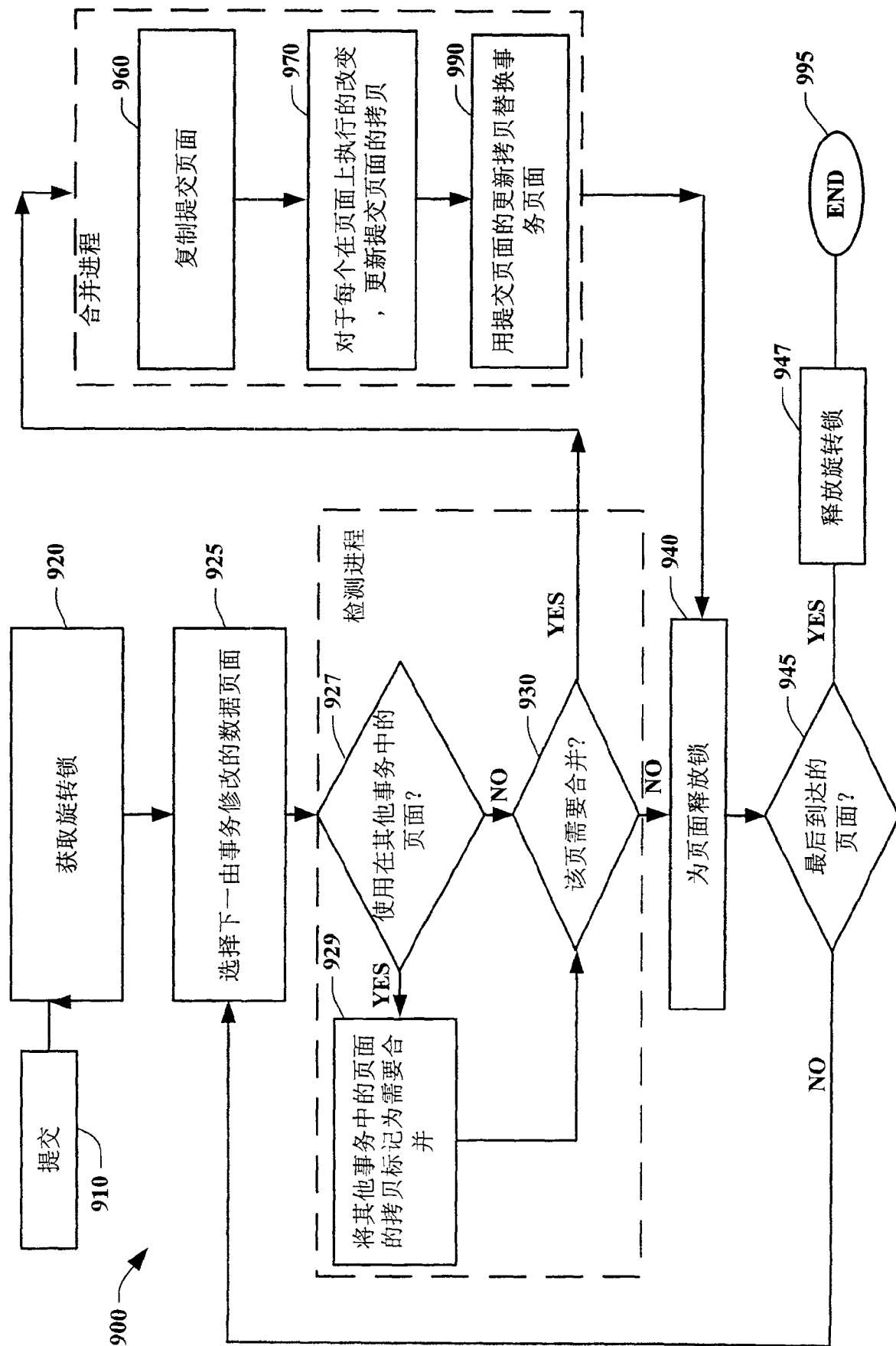


图 9

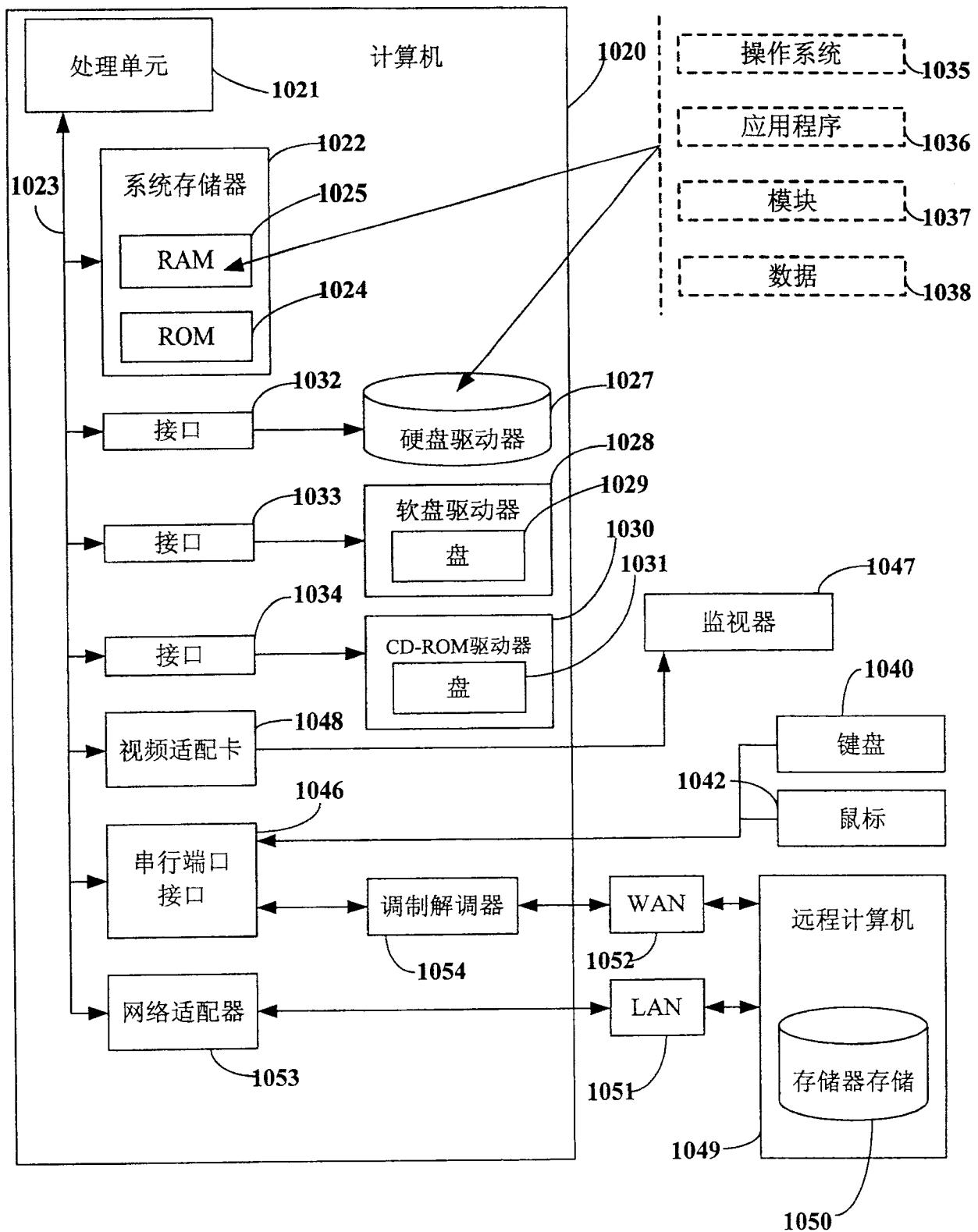


图 10

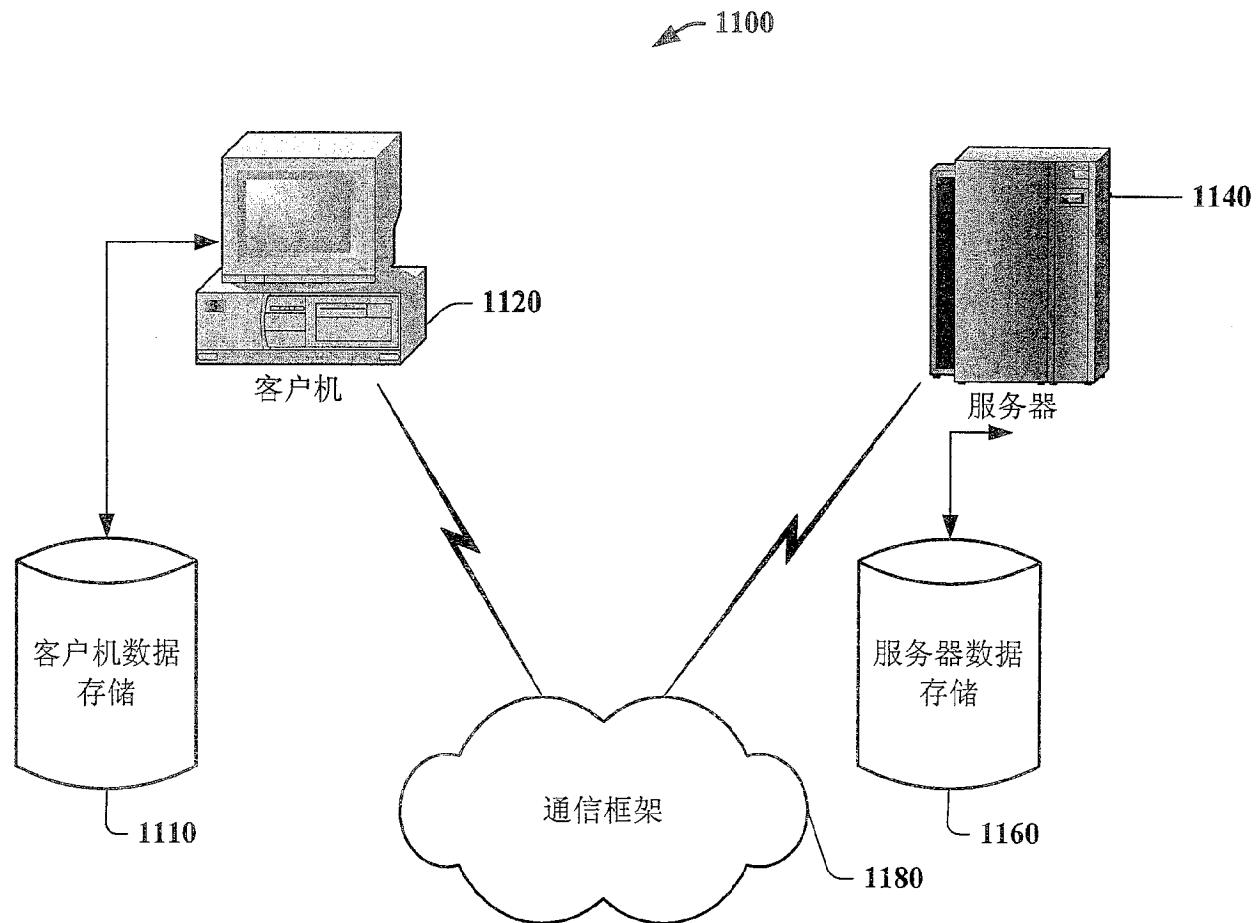


图 11