



(12) 发明专利

(10) 授权公告号 CN 103299267 B

(45) 授权公告日 2016. 05. 18

(21) 申请号 201180061441. 7

(22) 申请日 2011. 04. 15

(30) 优先权数据

12/973, 668 2010. 12. 20 US

(85) PCT国际申请进入国家阶段日

2013. 06. 20

(86) PCT国际申请的申请数据

PCT/US2011/032631 2011. 04. 15

(87) PCT国际申请的公布数据

W02012/087366 EN 2012. 06. 28

(73) 专利权人 销售力网络公司

地址 美国加利福尼亚州

(72) 发明人 比尔·C·艾德森 克雷格·韦斯曼

凯文·奥利弗 詹姆斯·泰勒

西蒙·Z·费尔 多诺万·A·谢奈德

(74) 专利代理机构 北京德琦知识产权代理有限

公司 11018

代理人 康泉 宋志强

(51) Int. Cl.

G06F 7/00(2006. 01)

G06F 17/00(2006. 01)

(56) 对比文件

CN 101430688 A, 2009. 05. 13,

US 2009/0282045 A1, 2009. 11. 12,

US 2010/0211619 A1, 2010. 08. 19,

US 2007/0130137 A1, 2007. 06. 07,

审查员 刘华楠

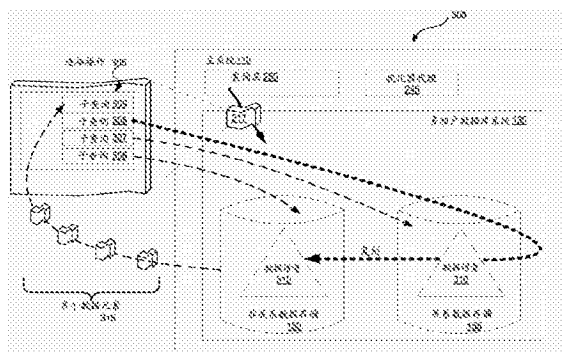
权利要求书5页 说明书14页 附图7页

(54) 发明名称

用于执行多租户存储中的交叉存储连接的方法和系统

(57) 摘要

描述了用于执行多租户存储中的交叉存储连接的方法和系统。在一个实施例中,这种方法包括:从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据;接收指定需从多租户数据库系统中检索的数据的请求;基于请求,检索需检索的数据的一个或多个位置;基于请求生成数据库查询,其中数据库查询指定需检索的多个数据元素,多个数据元素包括存在于非关系数据存储内的一个或多个数据元素和存在于关系数据存储内的一个或多个其它数据元素;以及对多租户数据库系统运行数据库查询,以检索数据。



1. 一种从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据的方法,所述方法包括:

在用于所述多租户数据库系统的主系统处接收请求,所述请求指定需从所述多租户数据库系统中检索的数据;

基于所述请求,经由主组织处存储的客户概要,确定需检索的数据元素的多个位置;

在所述主系统处,使用经由所述客户概要确定的需检索的数据元素的所述多个位置来基于所述请求生成数据库查询,其中所生成的数据库查询指定需检索的多个数据元素,所述多个数据元素包括存在于所述非关系数据存储内的一个或多个数据元素和存在于所述关系数据存储内的一个或多个其它数据元素;以及

对所述主组织的所述多租户数据库系统运行在所述主组织处生成的所述数据库查询,以检索被指定需经由所接收的所述请求而检索的所述数据。

2. 如权利要求 1 所述的方法,其中所述客户概要描述需检索的数据的所述一个或多个位置,所述客户概要将所述数据的所述多个数据元素中的每个数据元素指定为存在于所述非关系数据存储中或存在于所述关系数据存储中,或者指定为既能从所述非关系数据存储中又能从所述关系数据存储中获得;并且

其中所述方法进一步包括经由所述主系统响应于接收所述请求,检索所述客户概要。

3. 如权利要求 1 所述的方法,其中所述数据库查询包括多个子查询;

其中所述多个子查询中的至少一个子查询旨在从所述非关系数据存储中检索存在于所述非关系数据存储中的所述一个或多个数据元素;并且

其中所述多个子查询中的至少第二子查询旨在从所述关系数据存储中检索存在于所述关系数据存储中的所述一个或多个其它数据元素。

4. 如权利要求 1 所述的方法,其中对所述多租户数据库系统运行所述数据库查询包括:参考存储在所述关系数据存储和所述非关系数据存储中的数据元素。

5. 如权利要求 1 所述的方法,其中所述数据库查询指定连接操作。

6. 如权利要求 5 所述的方法,其中所述连接操作包括多个子查询,所述多个子查询基于经由所述主系统的优化器代理能获得的查询优化而生成,所述查询优化选自包括如下的组:

对所述多个子查询的指定排序;

用于运行相应子查询的目标数据存储;

基于需检索的数据的一个或多个预查询评估;

从所述关系数据存储至所述非关系数据存储的复制顺序;

存储器内连接操作,指定需从所述关系数据存储和所述非关系数据存储中的每个数据存储中检索的、且被置于所述优化器代理可存取的存储器中的所述多个数据元素中的至少一个或多个数据元素,并且指定相应的子查询,以在完成所述请求时从所述优化器代理可存取的所述存储器中检索所述多个数据元素中的所述至少一个或多个数据元素。

7. 如权利要求 5 所述的方法,其中所述连接操作包括:

需对所述非关系数据存储运行的第一子查询,其中所述第一子查询识别存在于所述非关系数据存储中的所述一个或多个数据元素;

需对所述关系数据存储运行的第二子查询,其中所述第二子查询确定所述第一子查询

识别的存在于所述非关系数据存储中的所述一个或多个数据元素和存在于所述关系数据存储中的所述一个或多个其它数据元素之间的数据增量；

需对所述关系数据存储和所述非关系数据存储运行的第三子查询，其中所述第三子查询将与所确定的数据增量对应的数据从所述关系数据存储复制至所述非关系数据存储；以及

需对所述非关系数据存储运行的第四子查询，其中所述第四子查询通过取得存在于所述非关系数据存储中的所述一个或多个数据元素和取得从所述关系数据存储复制至所述非关系数据存储并因此能从所述非关系数据存储中获得的所述一个或多个其它数据元素，取得需从所述非关系数据存储中检索的数据。

8. 如权利要求 5 所述的方法，其中所述连接操作选自包括以下的组：

指定来自所述关系数据存储的两个或两个以上关系表格的连接操作；

指定来自所述关系数据存储的至少一个关系表格和存在于所述非关系数据存储中的至少一个或多个数据结构的连接操作；

指定存在于所述非关系数据存储中的两个或两个以上分离且不同的数据结构的连接操作，其中所述两个或两个以上分离且不同的数据结构中的每个数据结构均缺少重叠的共享键。

9. 如权利要求 5 所述的方法，其中所述连接操作包括：

需对所述非关系数据存储运行的第一子查询，其中所述第一子查询检索存在于所述非关系数据存储中的所述一个或多个数据元素；

需对所述关系数据存储运行的第二子查询，其中所述第二子查询确定存在于所述非关系数据存储中的所述一个或多个数据元素和存在于所述关系数据存储中的所述一个或多个其它数据元素之间的数据增量；以及

需对所述关系数据存储运行的第三子查询，其中所述第三子查询基于所确定的数据增量检索存在于所述关系数据存储中的所述一个或多个其它数据元素。

10. 如权利要求 9 所述的方法，其中所述第三子查询包括以下之一：

需对所述关系数据存储运行的基于时间的子查询，其中基于所述关系数据存储中具有比与存在于所述非关系数据存储中的所述一个或多个数据元素中的那些数据元素相对应的任何时间戳晚的时间戳的那些数据元素，所述基于时间的子查询指定需从所述关系数据存储中检索的所述一个或多个其它数据元素；以及

需对所述关系数据存储运行的基于记录标识符的子查询，其中基于所述关系数据存储中具有比与存在于所述非关系数据存储中的所述一个或多个数据元素中的那些数据元素相对应的任何记录标识符在数值上大的记录标识符的那些数据元素，所述基于记录标识符的子查询指定需从所述关系数据存储中检索的所述一个或多个其它数据元素。

11. 如权利要求 9 所述的方法，进一步包括：

在所述多租户数据库系统处接收新事务，每个新事务指定需写至所述非关系数据存储的新数据；

将所述新数据写至所述关系数据存储的附加日志；并且

其中所述第三子查询基于所确定的数据增量检索存在于所述关系数据存储中的所述一个或多个其它数据元素包括：所述第三子查询从所述关系数据存储的所述附加日志中检

索存在于所述关系数据存储中的所述一个或多个其它数据元素。

12. 如权利要求 11 所述的方法,进一步包括:

当所述附加日志达到刷新阈值时,将写至所述关系数据存储的所述附加日志的所述新数据刷新至所述非关系数据存储。

13. 如权利要求 1 所述的方法,其中存在于所述非关系数据存储中的所述一个或多个数据元素包括多个压缩的平面文件或多个二进制文件或所述压缩的平面文件和所述二进制文件的组合。

14. 如权利要求 1 所述的方法,其中所述关系数据存储包括根据关系数据库管理系统(RDBMS)实现的关系数据库,其中所述关系数据库的多个关系表格通过所述关系数据库中的两个或两个以上关系表格中每个关系表格的一个或多个重叠的公共特性彼此相互关联。

15. 如权利要求 1 所述的方法,其中所述非关系数据存储包括具有多个底层硬件存储装置的分布式结构化数据库,每个底层硬件存储装置均为所述非关系数据存储提供全部存储容量的至少一部分,并且其中所述非关系数据存储中的数据元素基于主键是能参考的,并且基于两个或两个以上关系表格之间的一个或多个重叠的公共特性是不能参考的。

16. 如权利要求 1 所述的方法,

其中所述关系数据存储包括从包括以下的组中选择的关系数据库实现:Oracle 兼容数据库实现、IBM DB2 企业服务器兼容关系数据库实现、MySQL 兼容关系数据库实现和微软 SQL Server 兼容关系数据库实现;并且

其中所述非关系数据存储包括从包括以下的组中选择的 NoSQL 非关系数据库实现:Vampire 兼容非关系数据库实现、Apache Cassandra 兼容非关系数据库实现、BigTable 兼容非关系数据库实现和 HBase 兼容非关系数据库实现。

17. 如权利要求 1 所述的方法,

其中经由所述多租户数据库系统的接口接收所述请求包括经由所述多租户数据库系统的网页服务器接收所述请求,其中所述网页服务器向发起所述请求的位于远程的终端用户客户机器提供基于网页的接口;并且

其中所述请求包括对来自所述多租户数据库系统的服务的在用于所述多租户数据库系统的主组织中操作的请求。

18. 如权利要求 1 所述的方法,其中所述多租户数据库系统进一步包括由多个分离且不同的客户组织共享的硬件和软件单元,所述分离且不同的客户组织中的每个客户组织均位于距离主组织远程的位置,所述主组织具有在其内运行的所述多租户数据库系统。

19. 如权利要求 1 所述的方法,

其中所述非关系数据存储包括多个分布式计算节点,每个所述分布式计算节点均至少包括存储器、一个或多个处理器以及一个或多个可通信地接合的硬盘驱动器,并且其中每个所述分布式计算节点均包括隔离的非关系数据库示例,所述隔离的非关系数据库示例具有在没有来自集中式事务许可的授权或控制的情况下读取、写入和更新非关系数据库事务的功能;并且

其中所述关系数据存储包括单体关系数据库实例,所述单体关系数据库实例包括与集中式事务许可协调计算资源的存储器和处理器,所述集中式事务许可控制是否将所述单体关系数据库实例的更新或改变交付给可通信地与所述单体关系数据库实例接合并由所述

单体关系数据库实例控制的持久存储装置上的持久存储器。

20. 一种从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据的装置,其中所述装置包括:

接收模块,用于接收指定需从所述多租户数据库系统中检索的数据的请求;

确定模块,用于基于所述请求,经由主组织处存储的客户概要,确定需检索的数据元素的多个位置;

生成模块,用于使用经由所述客户概要确定的需检索的数据元素的所述多个位置来基于所述请求生成数据库查询,其中所生成的数据库查询指定需检索的多个数据元素,所述多个数据元素包括存在于所述多租户数据库系统的非关系数据存储内的一个或多个数据元素和存在于所述多租户数据库系统的关系数据存储内的一个或多个其它数据元素;以及

运行模块,用于对所述主组织的所述多租户数据库系统运行在所述主组织处生成的所述数据库查询,以检索被指定需经由所接收的所述请求而检索的所述数据。

21. 如权利要求 20 所述的从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据的装置,其中所生成的数据库查询指定连接操作,所述连接操作包括:

需对所述非关系数据存储运行的第一子查询,其中所述第一子查询识别存在于所述非关系数据存储中的所述一个或多个数据元素;

需对所述关系数据存储运行的第二子查询,其中所述第二子查询确定所述第一子查询识别的存在于所述非关系数据存储中的所述一个或多个数据元素和存在于所述关系数据存储中的所述一个或多个其它数据元素之间的数据增量;

需对所述关系数据存储和所述非关系数据存储运行的第三子查询,其中所述第三子查询将与所确定的数据增量对应的数据从所述关系数据存储复制至所述非关系数据存储;以及

需对所述非关系数据存储运行的第四子查询,其中所述第四子查询通过取得存在于所述非关系数据存储中的所述一个或多个数据元素和取得从所述关系数据存储复制至所述非关系数据存储并因此能从所述非关系数据存储中获得的所述一个或多个其它数据元素,取得需从所述非关系数据存储中检索的数据。

22. 一种用于执行多租户存储中的交叉存储连接的系统,其中所述系统包括:

处理器和存储器;

与多租户数据库系统的通信接口,所述多租户数据库系统具有在其内实现的关系数据存储和非关系数据存储;

请求处理器,接收指定需从所述多租户数据库系统中检索的数据的请求;

客户概要处理器,基于所述请求,经由主组织处存储的客户概要,确定需检索的数据元素的多个位置;

子查询生成器,使用经由所述客户概要确定的需检索的数据元素的所述多个位置来基于所述请求生成数据库查询,其中所生成的数据库查询指定需检索的多个数据元素,所述多个数据元素包括存在于所述非关系数据存储内的一个或多个数据元素和存在于所述关系数据存储内的一个或多个其它数据元素;以及

查询运行器,对所述主组织的所述多租户数据库系统运行在所述主组织处生成的所述数据库查询,以检索被指定需经由所接收的所述请求而检索的所述数据。

23. 如权利要求 22 所述的系统,其中所述子查询生成器基于所述请求生成数据库查询包括所述子查询生成器基于所述请求生成多个子查询来完成所述请求,所述子查询包括:

需对所述非关系数据存储运行的第一子查询,其中所述第一子查询识别存在于所述非关系数据存储中的所述一个或多个数据元素;

需对所述关系数据存储运行的第二子查询,其中所述第二子查询确定所述第一子查询识别的存在于所述非关系数据存储中的所述一个或多个数据元素和存在于所述关系数据存储中的所述一个或多个其它数据元素之间的数据增量;

需对所述关系数据存储和所述非关系数据存储运行的第三子查询,其中所述第三子查询将与所确定的数据增量对应的数据从所述关系数据存储复制至所述非关系数据存储;以及

需对所述非关系数据存储运行的第四子查询,其中所述第四子查询通过取得存在于所述非关系数据存储中的所述一个或多个数据元素和取得从所述关系数据存储复制至所述非关系数据存储并因此能从所述非关系数据存储中获得的所述一个或多个其它数据元素,取得需从所述非关系数据存储中检索的所述数据。

用于执行多租户存储中的交叉存储连接的方法和系统

[0001] 优先权声明

[0002] 本申请涉及 2010 年 12 月 20 日提交的、标题为“METHODS AND SYSTEMS FOR PERFORMING CROSS STORE JOINS IN A MULTI-TENANT STORE(用于执行多租户存储中的交叉存储连接的方法和系统)”的、具有申请号 12/973,668 和代理人案号 No. 8956P006/320US 的美国专利申请,以及 2010 年 4 月 19 日提交的、标题为“METHODS AND SYSTEMS FOR PERFORMING CROSS STORE JOINS IN A MULTI-TENANT STORE(用于执行多租户存储中的交叉存储连接的方法和系统)”的、具有申请号 61/325,709 和代理人案号 No. 8956P006Z/320PROV 的临时实用申请,并且要求这两个申请的优先权,这两个申请的全部内容通过引用并入本申请;并且本申请进一步涉及 2010 年 4 月 20 日提交的、标题为“METHODS AND SYSTEMS FOR OPTIMIZING QUERIES IN A MULTI-TENANT STORE(用于优化多租户存储中的查询的方法和系统)”的、具有申请号 61/325,951 和代理人案号 No. 8956P007Z/321PROV 的临时实用申请,并且要求这个临时实用申请的优先权,该申请的全部内容通过引用并入本申请。

[0003] 著作权通告

[0004] 本专利文档的公开内容的一部分包含受到著作权保护的材料。当本专利文档或本专利公开内容在专利和商标局的专利文件或记录中出现时,著作权的所有者不反对任何人复制本专利文档或本专利公开内容,但除此以外,著作权的所有者无论怎样都保留所有著作权的权利。

技术领域

[0005] 本申请中描述的主题大体涉及计算领域,更具体地,涉及用于执行多租户存储中的交叉存储连接的方法和系统。

背景技术

[0006] 在本背景技术部分中讨论的主题不应该只因其在在本背景技术部分中提及而被臆断为现有技术。类似地,在本背景技术部分中提及的问题或者与本背景技术部分的主题相关的问题不应当被臆断为以前已经在现有技术中被认知。本背景技术部分中的主题仅仅代表不同方法,这些不同的方法本身还可以对应于要求保护的主题的实施例。

[0007] 在计算环境中,可以选择各种数据存储环境来持久地存储数据。例如,可以将数据存储在与由操作系统管理的文件系统中,操作系统将文件系统数据持久地存储在硬驱上,或者可以将数据持久地存储在数据库中。各种类型的数据库是可利用的,每种数据库均具有其自身特定的益处和缺陷。例如,所谓的关系数据库利用由该数据库中每个数据表共享的公共特性,提供使该数据库中的各个数据表彼此“关联”的能力。例如,在关系数据库中,职工标识符可以被用作关联一个以上的表的公共特性。然而,这种数据库结构具有特定缺陷,缺陷之一是这些关系使高的计算开支花费水平和高的计算复杂程度成为必需,这限制了关系数据库能够扩展的范围。

[0008] 还存在非关系数据库模型和实现,并且非关系数据库模型和实现共同地表现出更好的可扩展性,但是也表现出与关系数据库模型和实现不相关联的不同缺陷。例如,非关系数据库实现经常对存储大文件或大对象表现出改善的可扩展性,但是非关系数据库实现可能不太适于其它方面,例如对选择性的数据集进行分类或者为快速改变数据集实现数据保障。

[0009] 不幸的是,对来自多个数据存储的信息同时进行参考的数据库查询是非常低效的,并且从否则可能从多个数据存储的实现中得到的优点中减损优点。此外,对多样数据库模型的不同实现进行同时参考的数据库查询,可能是利用从前的数据库查询机制完全无法实现的。

发明内容

[0010] 本发明的第一方面是一种从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据的方法,所述方法包括:在用于所述多租户数据库系统的主系统处接收请求,所述请求指定需从所述多租户数据库系统中检索的数据;基于所述请求,经由主组织处存储的客户概要,确定需检索的数据元素的多个位置;在所述主系统处,使用经由所述客户概要确定的需检索的数据元素的所述多个位置来基于所述请求生成数据库查询,其中所生成的数据库查询指定需检索的多个数据元素,所述多个数据元素包括存在于所述非关系数据存储内的一个或多个数据元素和存在于所述关系数据存储内的一个或多个其它数据元素;以及对所述主组织的所述多租户数据库系统运行在所述主组织处生成的所述数据库查询,以检索被指定需经由所接收的所述请求而检索的所述数据。

[0011] 本发明的第二方面是一种从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据的装置,其中所述装置包括:接收模块,用于接收指定需从多租户数据库系统中检索的数据的请求;确定模块,用于基于所述请求,经由主组织处存储的客户概要,确定需检索的数据元素的多个位置;生成模块,用于使用经由所述客户概要确定的需检索的数据元素的所述多个位置来基于所述请求生成数据库查询,其中所生成的数据库查询指定需检索的多个数据元素,所述多个数据元素包括存在于所述多租户数据库系统的非关系数据存储内的一个或多个数据元素和存在于所述多租户数据库系统的关系数据存储内的一个或多个其它数据元素;以及运行模块,用于对所述主组织的所述多租户数据库系统运行在所述主组织处生成的所述数据库查询,以检索被指定需经由所接收的所述请求而检索的所述数据。

[0012] 本发明的第三方面是一种用于执行多租户存储中的交叉存储连接的系统,其中所述系统包括:处理器和存储器;与多租户数据库系统的通信接口,所述多租户数据库系统具有在其内实现的关系数据存储和非关系数据存储;请求处理器,接收指定需从所述多租户数据库系统中检索的数据的请求;客户概要处理器,基于所述请求,经由主组织处存储的客户概要,确定需检索的数据元素的多个位置;子查询生成器,使用经由所述客户概要确定的需检索的数据元素的所述多个位置来基于所述请求生成数据库查询,其中所生成的数据库查询指定需检索的多个数据元素,所述多个数据元素包括存在于所述非关系数据存储内的一个或多个数据元素和存在于所述关系数据存储内的一个或多个其它数据元素;以及查询运行器,对所述主组织的所述多租户数据库系统运行在所述主组织处生成的所述数据库

查询,以检索被指定需经由所接收的所述请求而检索的所述数据。

附图说明

[0013] 实施例被图示作为示例,而不作为限制,并且当结合附图思考下面的具体实施方式时,对照下面的具体实施方式能够更全面地理解实施例,其中:

[0014] 图 1 图示在其中实施例可以操作的示例性体系结构;

[0015] 图 2 图示在其中实施例可以操作的可替代示例性体系结构;

[0016] 图 3 图示在其中实施例可以操作的可替代示例性体系结构;

[0017] 图 4 图示在其中实施例可以操作的可替代示例性体系结构;

[0018] 图 5 图示在其中实施例可以操作、被安装、被集成或被配置的系统图解表示;

[0019] 图 6 是图示根据一个实施例的用于执行多租户存储中的交叉存储连接的方法的流程图;以及

[0020] 图 7 图示根据一个实施例的呈计算机系统的示例性形式的机器的图解表示。

具体实施方式

[0021] 本申请中描述的是用于执行多租户存储中的交叉存储连接的系统、装置和方法。在一个实施例中,这种方法包括从具有关系数据存储和非关系数据存储的多租户数据库系统中检索数据。例如,在这种方法中,用于多租户数据库系统的主系统接收指定需从多租户数据库系统中检索的数据的请求;经由主系统基于请求,检索需检索的数据的一个或多个位置;在主系统处,基于请求生成数据库查询,其中数据库查询指定需检索的多个数据元素,多个数据元素包括存在于非关系数据存储中的一个或多个数据元素和存在于关系数据存储中的一个或多个其它数据元素;以及对多租户数据库系统运行数据库查询,以检索数据,。

[0022] 联邦查询(其是搜索或参考一个以上数据库的查询)可能是很低效的,尤其当在数据库的最低行层次参考数据(例如通过请求在不同数据存储中存储的表格之间的连接操作)时,由于操作耗用如此多网络带宽,以至于这样的连接操作不能很好地扩展,因此不能在较大的数据库实现上可行地实现。当在对多样模型进行操作的数据库实现之间请求数据连接时(如在例如关系数据库实现和非关系数据库实现之间的连接),对这样的连接操作的挑战进一步加重。本申请中描述的方法论以在更大的数据库系统上能够可行地实现的方式以及特别地以在影响对多样操作模型(如关系模型和非关系模型)进行操作的多种数据存储实现的系统上能够可行地实现的方式,促进执行这样的连接操作的能力。

[0023] 例如,利用本申请中描述的方法论,在一个或多个外键父(foreign key parent)是存储在关系型数据库实现(例如Oracle™)中的对象的情况下,可以通过对非关系数据库存储的对象发起非关系数据库查询来执行连接操作。举例来说,存储在非关系数据库中的子表可以具有存储在Oracle™中的“帐户”表作为其主表,尽管非关系数据库存储的对象持续在非关系数据库实现中且Oracle™存储的对象持续在关系数据库实现中。

[0024] 在下面的描述中,提出许多特定细节,诸如特定系统、语言、组件等的示例,以便提供对各实施例的充分理解。然而,将对本领域技术人员清楚的是,这些特定细节不需要被用来实践所公开的实施例。在其它情况下,为了避免不必要地模糊所公开的实施例,尚未详细

描述众所周知的材料或方法。

[0025] 除在附图中绘出的和在本申请中描述的各硬件组件以外,实施例进一步包括下面描述的各操作。根据这样的实施例描述的操作可以通过硬件组件执行,或者可以包含在机器可运行指令中,这可以用来引起用这些指令编程的通用或专用处理器执行这些操作。可替代地,操作可以由硬件和软件的结合执行。

[0026] 实施例还涉及用于执行本申请中的操作的系统或设备。所公开的系统或设备置可以是为所需用途而特别地构建的,或者其可以包括通用计算机,该通用计算机由存储在该通用计算机中的计算机程序选择性地激活或重新配置。这种计算机程序可以存储在非瞬态计算机可读存储介质中,例如但不限于包括软盘、光盘、CD-ROM 和磁光盘在内的任何类型的磁盘、只读存储器 (ROM)、随机存取存储器 (RAM)、EPROM、EEPROM、磁卡或光学卡或者适于存储非瞬态电子指令的任何类型的介质,它们每个均联接至计算机系统总线。在一个实施例中,其上存储有指令的计算机可读存储介质引起多租户数据库环境中的一个或多个处理器执行本申请中描述的方法和操作。在另一实施例中,执行这样的方法和操作的指令被存储于非瞬态计算机可读介质中以供后来运行。

[0027] 本申请中呈现的算法和显示内在不涉及任何特定的计算机或其它设备,关于任何特定编程语言描述的实施例也不涉及。将理解,各种编程语言可以用来实现本申请中描述的实施例的教导。

[0028] 图 1 图示在其中实施例可以操作的示例性体系结构 100。体系结构 100 描绘出主系统 110,主系统 110 经由网络 125 与若干个客户组织 (105A、105B 和 105C) 可通信地接合。在主系统 110 中是多租户数据库系统 130,在多租户数据库系统 130 内具有实现主系统 110 中的数据库功能和代码执行环境的多个底层硬件、软件和逻辑元件 120,并且其中,多租户数据库系统 130 的硬件、软件和逻辑元件 120 与多个客户组织 (105A、105B 和 105C) 分离且不同,多个客户组织 (105A、105B 和 105C) 通过经由网络 125 与主系统 110 可通信地接合来利用由主系统 110 提供的服务。在这种实施例中,每个分离且不同的客户组织 (105A-105C) 可以位于远离主组织的位置,该主组织经由在内部运行多租户数据库系统 130 的主系统 110 向客户组织 (105A-105C) 提供服务。可替代地,客户组织 105A-105C 中的一个或多个客户组织可以与主系统 110 共处一地,例如在提供持久地存储底层数据的多租户数据库系统 130 的同一个主组织中。

[0029] 在一个实施例中,多租户数据库系统 130 的硬件、软件和逻辑元件 120 至少包括非关系数据存储 150 和关系数据存储 155,非关系数据存储 150 和关系数据存储 155 根据实现主系统 110 中的数据库功能和代码运行环境的硬件、软件和逻辑元件 120 而操作。主系统 110 可以进一步经由网络接收来自多个客户组织 105A-105C 中的一个或多个客户组织的请求 115。例如,进入的请求 115 可以对应于对服务的请求,或对应于在多租户数据库系统 130 中检索或存储代表客户组织 105A-105C 之一的数据的请求。

[0030] 图 2 图示在其中实施例可以操作的可替代示例性体系结构 200。在一个实施例中,主系统 110 实现从具有关系数据存储 155 和非关系数据存储 150 的多租户数据库系统 130 中检索数据的方法。

[0031] 例如,在这种实施例中,在用于多租户数据库系统 130 的主系统 110 处接收请求 115,请求 115 指定需从多租户数据库系统 130 中检索的数据 218。在一些实施例中,在主系

统 110 中操作的不同的网页 (web) 服务器 210 经由网络 125 接收进入的请求 115。例如, 网页服务器 210 可以负责经由网络 125 接收来自各客户组织 105A-105C 的请求 115。网页服务器 210 可以向发起请求 115 (例如位于客户组织 105A-105C 中的终端用户客户机装置) 的终端用户客户机提供基于网页 (web) 的接口, 请求 115 构成对来自在像主系统 110 这样的提供例如远程实现的云计算服务的主组织内操作的多租户数据库系统 130 的服务的请求。根据特定的实施例, 优化器代理 245 还可以提供额外功能, 例如开发预查询和优化数据查询。

[0032] 在一个实施例中, 主系统 110 基于请求 115 检索需检索的数据 218 的一个或多个位置 216。在一个实施例中, 客户概要 240 描述需检索的数据 218 的一个或多个位置 216, 其中客户概要 240 将需检索的数据 218 的多个数据元素中的每个数据元素指定为存在于非关系数据存储 150 内或存在于关系数据存储 155 内, 或者指定为从非关系数据存储 150 和关系数据存储 155 中均可获得。在一个实施例中, 主系统 110 响应于接收请求 115, 检索客户概要 240。可替代地, 主系统 110 从客户概要 240 中检索需检索的数据 218 的一个或多个位置 216。

[0033] 例如, 在特定的实施例中, 将需检索的数据 218 的一个或多个位置 216 存储在客户概要 240 中, 并且从客户概要 240 中检索需检索的数据 218 的一个或多个位置 216, 客户概要 240 指定构成需检索的数据 218 的多个数据元素中的每个数据元素在多租户数据库系统 130 中位于哪里。这样的客户概要 240 可以经由例如全局缓存层访问, 全局缓存层向主系统 110 中实现或提供所描述的多租户存储能力的各元件提供快速高效的访问。在可替代实施例中, 可以通过主系统 110、通过优化器代理 245、通过主系统 110 的查询层 260 或通过主系统 110 中负责确定需从多租户数据库系统 130 中检索的跨多样数据库实现散布的数据 218 (例如如图 2 所示跨非关系数据存储 150 和关系数据存储 155 散布的具有多个数据元素的数据 218) 的位置 216 的其它元件, 从客户概要 240 中检索需检索数据 218 的一个或多个位置 216。

[0034] 在一个实施例中, 主系统 110 基于请求 115 生成数据库查询 217, 其中数据库查询 217 指定需检索的多个数据元素, 多个数据元素包括存在于非关系数据存储 150 中的一个或多个数据元素和存在于关系数据存储 155 中的一个或多个其它数据元素。在特定实施例中, 数据库查询 217 进一步基于需检索的数据 218 的被检索的一个或多个位置 216。这样的数据库查询 217 可以进一步由主系统 110 生成或委派, 去由主系统 110 的子系统 (例如查询层 260 或优化器代理 245) 生成。

[0035] 在一个实施例中, 主系统 110 对多租户数据库系统 130 运行所生成的数据库查询 217 来检索数据 218, 例如图 2 所示, 指向多租户数据库系统 130 的多个下层硬件、软件和逻辑元件 120 的朝向下的箭头表示数据库查询 217 被传递给多租户数据库系统 130 的实现功能, 并且数据 218 响应性地被多租户数据库系统 130 返回, 这由向上弯曲的箭头表示, 这样将从多样的数据存储 (非关系数据存储 150 和关系数据存储 155) 中的每个数据存储发出的多个数据元素送回至主系统 110。

[0036] 在一个实施例中, 数据库查询 217 包括多个子查询。在这样的实施例中, 多个子查询中的至少一个子查询旨在从非关系数据存储 150 中检索存在于非关系数据存储 150 内的一个或多个数据元素, 并且多个子查询中的至少第二子查询旨在从关系数据存储 155 中检

素存在于关系数据存储 155 内的一个或多个其它数据元素。例如图 2 所示的,在数据库查询 217 的放大图中的是数个子查询串,例如“从非关系数据存储 (例如 150) 中检索数据元素 ‘a’”、“从关系数据存储 (例如 155) 中检索数据元素 ‘b’”以及另一子查询串,该另一子查询串表述反映一般结构化查询语言 (SQL) 型查询的“select ‘x’ from ‘y’ where ‘z’ (在 ‘z’ 条件下,从 ‘y’ 中选择 ‘x’)”。依赖所选择的实现查询语言或句法,这种查询可能适合于或者可能不适合于查询底层数据存储 (例如 150 和 155)。

[0037] 因此,根据这样的实施例,对多租户数据库系统 130 运行数据库查询 217 包括对存储在关系数据存储 155 中和存储在非关系数据存储 150 中的数据元素进行参考,以便检索所需的数据 218。

[0038] 图 3 图示在其中实施例可以操作的可替代示例性体系结构 300。特别地,更加详细示出的是根据特定实施例的由数据库查询 217 指定的连接操作。

[0039] 例如,根据一个实施例,连接操作 305 由数据库查询 217 指定。

[0040] 在特定实施例中,连接操作 305 包括多个子查询。例如,在这样的实施例中,将对非关系数据存储 150 运行第一子查询 306,并且第一子查询 306 识别存在于非关系数据存储 150 中的一个或多个数据元素,这由指向非关系数据存储 150 的虚线弯曲箭头表示。

[0041] 在这样的实施例中,将对关系数据存储 155 运行第二子查询 307,并且第二子查询 307 确定第一子查询 306 识别的存在于非关系数据存储 150 内的一个或多个数据元素和存在于关系数据存储 155 内的一个或多个其它数据元素之间的数据增量 (data delta) 310。

[0042] 在该实施例中,将对关系数据存储 155 和非关系数据存储 150 运行第三子查询 308,其中第三子查询 308 把与所确定的数据增量 310 相对应的数据从关系数据存储 155 复制至非关系数据存储 150。例如,这样的第三子查询 308 可以检索存在于关系数据存储 155 中的一个或多个其它数据元素,将它们拖入例如临时表格、文件、临时缓存的数据等中,然后,这样的第三子查询 308 可以对非关系数据存储 150 发出对应于数据增量 310 的被检索的数据的插入或写入命令,导致将数据增量 310 的数据写入、存入或插入非关系数据存储 150 中,从而完成复制且因此进一步导致之前不获得的存在于关系数据存储 155 中的数据元素现在可从非关系数据存储 150 中获得。参照图 3 的虚线,图 3 的虚线示出对两个数据存储 (关系数据存储 155 和非关系数据存储 150) 运行的第三子查询 308,以将所识别的数据增量 310 从关系数据存储 155 复制至非关系数据存储 150。

[0043] 从一个数据存储向另一数据存储复制或同步数据的确定可以基于各种考虑。例如,从关系数据存储 155 向非关系数据存储 150 复制的决定可以基于将较小的数据集从其原始位置复制至具有较大数据集的位置的确定或策略。例如,从网络带宽角度看进行从关系数据存储 155 向非关系数据存储 150 的复制时,复制属于所请求的数据的一部分的一个或多个数据元素可以比在进行从非关系数据存储 150 向关系数据存储 155 的复制时更高效。

[0044] 在一些实施例中,相反情况可以同样是实际的,因此,数据的复制可以在从非关系数据存储 150 向关系数据存储 155 的另一方向上进行。这样的确定可以由例如优化器代理 245 进行或操纵。在利用非关系数据库实现 (例如 150) 的特定实施例中,在关系数据存储 155 (例如 Oracle™) 中存储的关系数据库型对象经由指定连接操作 305 的一个子查询被复制至非关系数据存储 150,然后因为复制导致所有所需数据都变成可从非关系数据存

储 150 中获得,所以指定数据检索操作的另一子查询可以从非关系数据存储 150 拖拽所有需求数据,尽管在这种示例中至少一些数据被持久地存储在关系数据库存储器 155(例如 Oracle™)中且最初仅可从关系数据存储 155 中获得。

[0045] 其它复制决定和考虑可以类似地考虑并通过优化器代理 245 实现。例如,一个复制策略可以基于复制操作提供的一致性保证,例如被复制的数据是否总是同步或者总是保证同步,或者某种偏差是否是可接受的风险。

[0046] 例如,拿需要从关系数据存储 155 向非关系数据存储 150 复制一千万个“帐户”表的复制操作使得能够从非关系数据存储 150 中查询十亿个子行为例。在这样的示例的情况下,非关系数据库查询引擎可以被用来经由 JDBC(Java 数据库连接)进行批量调出(bulk callout),以检索 Oracle™数据(例如关系)数据。在这样的示例中,Oracle™ RAC(Oracle 真正应用集群)可以用来或者基于 Oracle™11g 数据保护的保证机制可以用来提供必要的一致性,这里一致性保证是重要的考虑并且正在发起非常大量的数据复制。这样的大量复制可以最适当地预先执行,因为在查询时如此做,对数据的进入请求的响应性运行期间(on-the-fly)可能在对进入请求做出响应或在完成进入请求时需要不可接受的长延迟。

[0047] 相反地,拿小表格和相应地少的数据转移为例。在这样的示例中,例如响应于对数据的进入请求,存在于具有当前数据的关系数据存储 155(例如 Oracle™)中的小表格或数个小表格的全部内容可以在查询时间被复制至非关系数据存储 150,在数据中至少一部分数据存在于小表格中并由关系数据存储 155(例如 Oracle™)持久地存储。这种策略可以适于大的分析,这里如由例如优化器代理 245 确定的,这种查询和复制的固定成本不显著。

[0048] 依赖于例如与每个特定的组织相关联的表格或对象的尺寸,另一考虑可以基于逐个 OrgID。例如,可以采用下面的策略,即与大组织的介质对应的数据(例如基于预定的尺寸阈值)被预先复制。在一个实施例中,预先复制可以利用表面表格复制码(skinny table replication code),该表面表格复制码捕获关系数据存储 155 中的指定对象的改变并将这些改变推送至另一表格,例如非关系数据存储 150 中预复制的表格(例如在选择预先复制的情况下)。

[0049] 在特定的实施例中,特定的组织可以触发大量改变,因此,实时同步可能并不是必须的或者并不适于每个改变,而仍允许对非关系数据存储 150 执行所请求的分析。因此,在这样的实施例中,提供了以特定间隔同步更新的选项(例如经由优化器代理 245 和下面介绍的基于硬件的查询层代理 501 和 734)。在这样的实施例中,允许以特定间隔更新的策略提供更高效率的写入和更新,虽然有可能悬空参考和其它“杂乱数据(sloppy data)”的结果,该“杂乱数据”可以是可接受的偏差或者可以依赖于数据库查询的所采用的复制策略或底层对象而需要后续的数据检查和验证。

[0050] 关于复制的另一考虑可以是数据存储统计,例如表格的基数。可以由优化器代理 245 产生统计,通过优化器代理 245 使统计可以是可用的,或者可以由优化器代理 245 收集统计。例如,在需要相对小的一组行的特定实施例中,当确定正在被处理的全体数据库查询 217 大且发送整组所查询的行构成合理的传输成本(例如基于预定的大小比例或固定阈值等)时,子查询(例如 306-309)可以从关系数据存储 155 中查询所需的整组行并将整组所查询的行发送至非关系数据存储 150。这种策略可以避免在还没有接收到对数据的实际需求的情况下预复制或进行预先复制的必要,并且这种策略可以进一步避免上面描述的数据

不一致性问题。

[0051] 将数据持久地存储至哪里的考虑和由此从哪里检索数据的考虑可以进一步基于特定数据存储的底层实现硬件。例如,针对大的平面文件和二进制文件在例如每千兆字节的成本方面廉价的存储器上的存储,可以优化非关系数据存储 150。这样的底层硬件实现可能可行的,因为非关系数据存储(例如 150)是针对压缩的平面文件和二进制文件的批量读取而优化的,因此非关系数据存储(例如 150)相对于像 Oracle™这样的关系模型数据存储(例如 155)对每千兆字节处理请求而言,在计算上更廉价。由于关系数据存储 155 实现例如对所有存储的数据的强制性分析、像事务处理这样的企业级数据保护和还原能力的要求,像 Oracle™这样的关系数据存储 155 可能需要在每千兆字节上更昂贵的存储器硬件。如果在最终完成以前数据库事务失效或被中断,则相比于有可能将缺少事务处理的数据存储处在不一致状态下的直接插入模式,这样的企业级数据保护进一步在失效的情况下通过事务处理具有“重做”特定事务的能力。

[0052] 因此,在特定的实施例,对所存储的数据的近期编辑被实现分析、事务处理和还原的关系数据存储 155 进行事务处理和存储,并且向关系数据存储 155 写入的更新中的至少一部分更新随后被复制、移动、事务处理或迁移至非关系数据存储 150,以便降低持久地存储相应数据的每千兆字节开支的成本。以这种方式,主系统 110 可以影响与特定关系数据存储 155 相关联的企业级数据保护,并且同时受益于经由一些非关系数据存储 150 可用的较廉价的持久存储。

[0053] 在一个实施例中,由于优化器代理 245 既具有来自非关系数据存储 150 又具有来自关系数据存储 155 的可用数据的“视野”,所以当与对数据 218 的进入请求 115 相对应的特定数据元素可以以一种以上的方式从一个以上的源中得到时,优化器代理 245 能够生成“选择性查询”。优化器代理 245 能够进一步生成需发出的多个子查询的改进序列或排序,以便完成这样的请求 115。

[0054] 因此,根据特定实施例并鉴于各种可用的考虑,在连接操作 305 中进一步包括第四子查询 309 并且将对非关系数据存储 150 运行第四子查询 309,其中第四子查询 309 通过既取得存在于非关系数据存储 150 中的一个或多个数据元素又取得从关系数据存储 155 复制至非关系数据存储 150 并因此可从非关系数据存储 150 中获得的一个或多个其它数据元素,从非关系数据存储 150 中取得需检索的数据。以这种方式,尽管在由连接操作 305 触发数据复制以前多个数据元素 315 中的一些数据元素最初不可从非关系数据存储 150 中获得,但是可以完全从数据存储之一(例如从非关系数据存储 150)中检索多个数据元素 315。

[0055] 在一些实施例中,作为在持久地存储被请求数据的一个或多个数据元素的数据存储之间复制数据的替代方式,可以采用一策略来将所有数据元素检索到与持久地存储数据的正被查询的两个或两个以上数据存储(例如 150 和 155)中的每个数据存储分离的位置中。例如,可以利用存储器内连接操作将数据检索至查询层 260 中,或者经由存储器内连接操作将数据检索至全局缓存层(例如如图 5 的要素 550)。这种存储器内连接操作可以基于可从优化器代理 245 中获得的已知统计或基于指定大小阈值(例如行数、数据在大小(例如数据的兆字节)上的数量、所请求的数据的基数等)来选择。

[0056] 基于例如主系统 110 中已知的统计和分析可以获得的其它考虑可以包括对例如从用于特定查询的已知挑选列表量中导出的特定数据库查询 217 或子查询(例如 306-309)

的查询成本,在该特定查询中最大数量的元素是已知的并且因此最大查询成本或估计查询成本是可确定的或者是从优化器代理 245 中已知的和可获得的。可以进一步基于已经进行的分析获知或者可以经由优化器代理 245 确定(例如通过一个或多个预查询),多个可用数据存储(例如 150 和 155)中哪些数据存储能够在最少量的时间内或利用/消耗最少的计算资源生成具有最少数量的行的结果。例如,在大数据库查询 217 情况下,可能可取的是,对来自每个数据存储(例如 150 和 155)的所需数据的一小部分进行预查询来确定哪些预查询导致更高效的结果,然后基于这样的确定,生成实现以更高效的数据存储(例如依赖于预查询的结果是 150 或 155)为目标的主数据库查询 217 所需的各子查询(306-309)。在优化器代理 245 提前进行适当分析的情况下,可以简单地请求查询策略,而不需要发出预查询,例如可以经由客户概要 240 对数据的一个或多个位置做出这样的分析确定并且随后存储和指定这样的分析确定。

[0057] 在可替代实施例中,可以执行不同的或附加的连接操作 305。例如,在一个实施例中,经由数据库查询 217 对多租户数据库系统 130 运行的连接操作 305 可以包括从由以下连接操作构成的一组连接操作 305 中选择的连接操作 305:指定来自关系数据存储 155 的两个或两个以上关系表格的连接操作 305;指定来自关系数据存储 155 的至少一个关系表格和存在于非关系数据存储 150 中的至少一个或多个数据结构的连接操作 305;以及指定存在于非关系数据存储 150 中的两个或两个以上分离且不同的数据结构的连接操作 305,其中两个或两个以上分离且不同的数据结构中的每个数据结构均缺乏重叠的共享键,例如共享特征、字符串、二进制或字母数字键,用这些关联或者以其它方式联系非关系数据存储 150 中的两个不同的数据结构。

[0058] 例如,在一个实施例中,非关系数据存储 150 提供存储许多数据结构、文件、对象和其它这种信息的能力,但是不实现“联系”这样的数据结构、文件、对象和其它信息的功能。然而,指定两个不同的数据结构中每个数据结构的连接操作 305 能够,通过依赖于在所实现的非关系数据存储 150 外部的功能和逻辑(例如多租户数据库 130 中的多个底层硬件、软件和逻辑元件 120),识别和链接或关联每个这样的数据结构,多个底层硬件、软件和逻辑元件 120 能够对非关系数据存储 150 运行适当形成的连接操作 305 以形成例如单个数据结构,该单个数据结构具有所有期望的但之前未关联的信息或者将由这种连接操作指定的期望信息检索和临时缓存在可替代的位置。

[0059] 在可替代实施例中,指定的连接操作 305 包括:需对非关系数据存储 150 运行的第一子查询(例如 306),其中第一子查询(例如 306)需检索存在于非关系数据存储 150 中的一个或多个数据元素;需对关系数据存储 155 运行的第二子查询(例如 307),其中第二子查询(例如 307)确定存在于非关系数据存储 150 中的一个或多个数据元素和存在于关系数据存储 155 中的一个或多个其它数据元素之间的数据增量 310;以及需对关系数据存储 155 运行的第三子查询(例如 308),其中第三子查询(例如 308)需基于所确定的数据增量 310 检索存在于关系数据存储 155 中的一个或多个其它数据元素。

[0060] 在这样的实施例中,基于所确定的数据增量 310 检索存在于关系数据存储 155 中的一个或多个其它数据元素的第三子查询(例如 308)可以包括基于时间的查询过滤机制或基于记录的过滤机制。

[0061] 例如,在一个实施例中,基于时间的子查询需对关系数据存储 155 运行,其中基于

关系数据存储 155 中具有比与存在于非关系数据存储 150 中的一个或多个数据元素中的那些数据元素相对应的任何时间戳晚的时间戳的那些数据元素,基于时间的子查询指定需从关系数据存储 155 中检索的一个或多个其它数据元素。

[0062] 在可替代实施例中,基于记录标识符的子查询需对关系数据存储 155 运行,其中基于关系数据存储 155 中具有在数值上比与存在于非关系数据存储 150 中的一个或多个数据元素中的那些数据元素相对应的任何记录标识符大的记录标识符的那些数据元素,基于记录标识符的子查询指定需从关系数据存储 155 中检索的一个或多个其它数据元素。

[0063] 图 4 图示在其中实施例可以操作的可替代示例性体系结构 400。特别地,以附加细节示出的是根据特定实施例的由多租户数据库系统 130 接收的新事务的处理。

[0064] 例如,在特定实施例中,被写入或插入多租户数据库系统 130 内去持久存储的新信息,可以被指定为当前长期存储在非关系数据存储 150 中,但仍可以临时被写至关系数据存储 155。例如,对临时性地将数据写至一个数据存储(例如关系数据存储 155)以及然后将该数据迁移至另一数据存储(例如非关系数据存储 150)的考虑可以包括例如对一个数据存储相对于另一数据存储的改善的写入响应时间以及然而来自交替数据存储的改善的检索时间。一个数据存储可以与较低的计算成本或较低的操作成本相关联。像非关系数据存储 150 这样的特定数据存储可以对很少更新但经常被检索的数据更高效地操作。可替代地,相对于前面的具有很少被更新的数据的示例,像关系数据存储 155 这样的另一数据存储可能对很零碎的数据或很频繁被更新或被添加的数据呈现出更好的操作效率。

[0065] 因此,根据特定实施例,在多租户数据库系统 130 处接收的新事务 415(例如在之前示出的请求 115 中)包括或指定需被写至非关系数据存储 150 的新数据 416。在一些实施例中,尽管有新数据 416 需被写至非关系数据存储 150 的指示,但是新数据 416 被写至关系数据存储 155 的附加日志 410。将新数据 416 写至哪里的这种指示可以由新事务 415 指定,例如由新事务 415 中的目标 419 属性指定。可替代地,可以基于由例如优化器代理 245 确定的新数据 416 的特性或基于与对应于新事务 415 的 OrgID 相关联的标记或存储偏好,由主系统 110 做出确定。

[0066] 在一些实施例中,包括基于确定的数据增量(例如 310)检索存在于关系数据存储 155 中的一个或多个其它数据元素的子查询的连接操作(例如 305)包括:从关系数据存储 155 的附加日志 410 中检索存在于关系数据存储 155 中的一个或多个其它数据元素的子查询。例如,可以检索被写至附加日志的新数据 416,或者可以检索存储在附加日志中的新数据 416 的元素。

[0067] 在一个实施例中,主系统 110 触发附加日志 410 的刷新,从而在附加日志 410 到达刷新阈值时把被写至关系数据存储 155 的附加日志 410 的新数据 416 刷新至非关系数据存储 150,这导致例如新数据然后存在于非关系数据存储 150 中作为被刷新的数据 417 并且对应于之前存在于关系数据存储 155 的附加日志 410 中的新数据 416。

[0068] 不同类型的数据可以由多租户数据库系统 130 存储。例如,在一个实施例中,存在于非关系数据存储 150 中的一个或多个数据元素对应于多个压缩的平面文件或多个二进制文件或压缩的平面文件和二进制文件的组合。这样的文件可以经由非关系数据库体系结构(例如 150)更高效地存储。

[0069] 在另一实施例中,关系数据存储 155 根据关系数据库管理系统(RDBMS)实现关系

数据库,其中,关系数据库的多个关系表格通过关系数据库中的两个或两个以上的关系表格中的每个关系表格的一个或多个重叠的公共特性相互关联,从而形成与关系型数据存储 155 一般地关联的“关系”。

[0070] 在一个实施例中,非关系数据存储 150 实现具有多个底层硬件存储装置的分布式结构化数据库,每个底层硬件存储装置均为非关系数据存储 150 提供总存储容量的至少一部分。在这样的实施例中,非关系数据存储 150 中的数据元素是在主键(primary key)基础上可参考的,但在两个或两个以上关系表格之间的一个或多个重叠的公共特性基础上是不可参考的,例如关系数据存储 155 中的数据元素的情况。

[0071] 在一个实施例中,关系数据存储 155 实现从以下数据库实现中选择的关系数据库模型:Oracle 兼容数据库实现、IBM DB2 企业服务器兼容关系数据库实现、MySQL 兼容关系数据库实现和微软 SQL Server 兼容关系数据库实现。

[0072] 在一个实施例中,非关系数据存储 150 实现从以下数据库实现中选择的 NoSQL 非关系数据库实现:Vampire 兼容非关系数据库实现、Apache Cassandra 兼容非关系数据库实现、BigTable 兼容非关系数据库实现和 HBase 兼容非关系数据库实现。

[0073] 在一个实施例中,非关系数据存储 150 包括多个分布式计算节点,每个计算节点均包括至少一个存储器、一个或多个处理器以及一个或多个可通信地接合的硬盘驱动器。在这样的实施例中,每个分布式计算节点可以进一步包括隔离的非关系数据库示例,该隔离的非关系数据库示例具有在没有来自集中式事务许可的授权或控制情况下读取、写入和更新非关系数据库事务的功能。

[0074] 在特定的实施例中,关系数据存储 155 实现包括存储器和处理器的单体关系数据库实例,存储器和处理器与集中式事务许可协调计算资源,集中式事务许可控制是否将单体关系数据库实例的更新或改变交付给可通信地与单体关系数据库实例接合并由单体关系数据库实例控制的持久存储装置上的持久存储器。

[0075] 图 5 示出在其中实施例可以操作、被安装、被集成或被配置的系统 500 的图解表示。

[0076] 在一个实施例中,系统 500 包括存储器 595 和处理器或多个处理器 590。例如,存储器 595 可以存储需运行的指令,处理器 590 可以运行这样的指令。系统 500 包括总线 515,以在可通信地与总线 515 接合的多个外围装置之间传递事务和系统 500 中的数据。系统 500 进一步包括网页服务器 525,以例如接收请求、返回响应以及以别的方式与远程客户机(例如位于客户组织 105A-105C 内的客户机装置)接合。

[0077] 系统 500 进一步被示出为具有优化器代理 535,优化器代理 535 被设计为优化数据库查询和数据库子查询并可选地协调预查询,以确定查询底层数据存储的最优方法或优选方法。系统 500 进一步包括全局缓存层 550,以向可通信地接合的装置和系统提供缓存服务以及特别地提供客户概要数据(例如元数据等)的缓存。客户概要数据由与全局缓存层 550 协作操作的客户概要 530 提供,全局缓存层 550 指定例如所需的数据元素是由多租户数据库系统中的关系数据库还是由多租户数据库系统中的非关系数据库实现存储还是由关系数据库和非关系数据库实现存储,并且全局缓存层 550 为构成相应请求的数据集的一个或多个数据元素指定底层数据存储中的位置。客户概要 530 可以存储在硬盘驱动器、持久数据存储或系统 500 中的其它存储位置上。

[0078] 在系统 500 中不同的是基于硬件的查询层代理 501, 基于硬件的查询层代理 501 包括请求处理器 570、客户概要处理器 575、子查询生成器 580 和查询运行器 585。根据一个实施例, 请求处理器 570 接收指定需检索的数据的请求 (例如从网页服务器 525、从前面所述的主系统 110 或直接从连接的网络接口)。请求处理器 570 与客户概要处理器 575 协调, 以检索需从底层数据存储中检索的所请求数据的一个或多个位置。请求处理器 570 进一步与子查询处理器 580 协调来开发和生成必要的子查询, 以基于上述数据的所确定的一个或多个位置从适当的底层数据存储中检索所请求的一个或多个数据元素, 或者请求处理器 570 生成必要的子查询来发起导致数据子集从一个数据存储被同步至、被刷新至或被复制至另一数据存储的连接操作, 使得后续的子查询能够从单独的数据存储中检索完整的所请求的数据集。由子查询生成器 580 生成的这样的子查询可以依赖于可从优化器代理 535 获得的统计和预查询结果。查询运行器 585 对可通信地接合的数据库实现运行所生成的查询和子查询。

[0079] 图 6 是图示根据一个实施例用于执行多租户存储中的交叉存储连接的方法 600 的流程图, 该方法包括根据特定实施例指定和执行由数据库查询 (例如 217) 指定的连接操作。方法 600 可以由处理逻辑执行, 处理逻辑可以包括硬件 (例如电路、专用逻辑、可编程逻辑、微代码等)、软件 (例如在处理装置上运行的、执行诸如读取、写入、更新、优化、发起预查询、开发子查询等各种查询操作的指令) 或软件和硬件的结合。在一个实施例中, 由像在图 5 的要素 501 处示出的基于硬件的查询层代理这样的硬件逻辑执行方法 600。以下列出的框和 / 或操作中的一些框和 / 或操作根据特定实施例是可选的。所呈现的框的编号是为了清晰目的, 而不旨在规定各个框必须发生的操作顺序。

[0080] 方法 600 从处理逻辑开始, 处理逻辑在用于多租户数据库系统的主系统处接收请求, 该请求指定需从多租户数据库系统中检索的数据 (框 605)。在框 610 处, 经由主系统基于该请求, 处理逻辑检索需检索的数据的一个或多个位置。

[0081] 在框 615 处, 经由主系统, 处理逻辑响应于接收请求检索客户概要。例如客户概要可以描述需检索的数据的一个或多个位置, 客户概要指定数据的多个数据元素中的每个数据元素为存在于非关系数据存储中或存在于关系数据存储中或者既可从非关系数据存储又可从关系数据存储中获得。

[0082] 在框 620 处, 在主系统处, 处理逻辑基于请求生成数据库查询。例如, 数据库查询可以指定需检索的多个数据元素, 多个数据元素包括存在于非关系数据存储中的一个或多个数据元素和存在于关系数据存储中的一个或多个其它数据元素。数据库查询可以进一步包括多个子查询。在一个实施例中, 数据库查询经由子查询之一指定连接操作。清除、刷新、同步或复制操作可以类似地经由子查询指定。

[0083] 在框 625 处, 处理逻辑对多租户数据库系统运行数据库查询, 以检索数据。

[0084] 在框 630 处, 处理逻辑在多租户数据库系统处接收新事务, 每个新事务均指定需写至非关系数据存储的新数据, 在框 635 处, 处理逻辑将新数据写至关系数据存储的附加日志。例如, 在一个实施例中, 数据库查询的子查询指定将从关系数据存储的附加日志中检索需检索的数据。

[0085] 在框 640 处, 当附加日志达到刷新阈值时, 处理逻辑把被写至关系数据存储的附加日志的新数据刷新至非关系数据存储。

[0086] 图 7 图示根据一个实施例的呈计算机系统的示例性形式的机器 700 的图解表示, 在机器 700 中可以运行用于导致机器 700 执行本申请介绍的方法中的任意一个或多个方法的一组指令。在可替代实施例中, 机器可以连接 (例如联网) 至局域网 (LAN)、内联网、外联网或互联网中的其它机器。机器可以在客户机—服务器网络环境中的服务器或客户机机器的容量下操作, 或者在对等 (或分布式) 网络环境中作为对等机器操作, 或者在请求式服务环境 (包括提供多租户数据库存储服务的请求式环境) 中作为服务器或一系列服务器操作。机器的特定实施例可以是以下形式: 个人计算机 (PC)、平板 PC、机顶盒 (STB)、个人数字助理 (PDA)、移动电话、web 工具、服务器、网络路由器、交换机或网桥、计算系统或能够运行指定需由该机器采取的动作的 (顺序的或其它方式的) 一组指令的任何机器。进一步, 尽管只图示单个机器, 但是术语“机器”还应当被认为包括单独地或共同地运行一组 (或多组) 指令以执行本申请中介绍的方法中的任意一个或多个方法的任何机器集合 (例如计算机)。

[0087] 示例性计算机系统 700 包括处理器 702、主存储器 704 (例如只读存储器 (ROM)、闪存、像同步 DRAM (SDRAM) 或存储器总线式 (Rambus) DRAM (RDRAM) 等这样的动态随机存取存储器 (DRAM)、像闪存这样的静态存储器、静态随机存取存储器 (SRAM)、易失性但高数据速率 RAM 等) 以及次存储器 718 (例如包括硬盘驱动器和持久多租户数据库实现的持久存储装置), 处理器 702、主存储器 704 和次存储器 718 经由总线 730 彼此通信。主存储器 704 包括客户概要 724 (例如指定构成两个或两个以上多样数据存储中指定数据或数据集的数据或数据元素的一个或多个位置, 例如跨越关系数据存储和非关系数据存储散布的且可经由例如基于硬件的查询层代理 734 检索的数据元素的位置)。主存储器 704 进一步包括全局缓存层 732, 例如系统广度可访问的全局缓存层, 以提供较大数据集的多个数据元素之间的元数据和与其它关联或对应信息, 例如经由客户概要 724 提供的那类信息。主存储器 704 和其子要素 (如 723 和 724) 可与处理逻辑 726 和处理器 702 协作操作, 以执行本申请中介绍的方法。

[0088] 处理器 702 表示一个或多个通用处理装置, 例如微处理器、中央处理单元等。更特别地, 处理器 702 可以是复杂指令集计算 (CISC) 微处理器、简化指令集计算 (RISC) 微处理器、超长指令字 (VLIW) 微处理器、实现其它指令集的处理或实现指令集组合的处理器。处理器 702 还可以是一个或多个专用处理装置, 例如专用集成电路 (ASIC)、现场可编程门阵列 (FPGA)、数字信号处理器 (DSP)、网络处理器等。处理器 702 被配置为运行处理逻辑 726, 以执行本申请中介绍的操作和功能。

[0089] 计算机系统 700 可以进一步包括网络接口卡 708。计算机系统 700 还可以包括用户界面 710 (例如视频显示单元、液晶显示器 (LCD) 或阴极射线管 (CRT))、字母数字输入装置 712 (例如键盘)、光标控制装置 714 (例如鼠标) 以及信号生成装置 716 (例如集成的扬声器)。计算机系统 700 可以进一步包括外围设备 736 (例如无线或有线通信装置、存储器装置、存储装置、音频处理装置、视频处理装置等)。计算机系统 700 可以进一步包括基于硬件的查询层代理 734, 基于硬件的查询层代理 734 管理数据库查询和子查询并与底层数据存储 (例如多租户数据库系统) 协调事务。

[0090] 次存储器 718 可以包括非瞬态机器可读存储介质 (或更特别地, 非瞬态机器可存取存储介质) 731, 在非瞬态机器可读存储介质上存储有体现本申请中描述的方法或功能中

的任意一个或多个方法或功能的一个或多个指令集（例如软件 722）。软件 722 在其由计算机系统 700 运行过程中还可以完全地或至少部分地存在于主存储器 704 和 / 或存在于处理器 702 中,主存储器 704 和处理器 702 还构成机器可读存储介质。软件 722 可以进一步在网络 720 上经由网络接口卡 708 发送或接收。

[0091] 尽管已经通过示例和关于具体实施例描述了本申请中公开的主题,但应当理解,所请求保护的实施例不限于所公开的明确列举的实施例。相反,本公开内容旨在覆盖对本领域技术人员来说显然的各种改变和类似布置。因此,所附权利要求的范围应与最宽的解释相一致,以便包含所有这样的修改和类似布置。要理解,上面的描述旨在是图解性的而不是限制性的。当本领域技术人员阅读和理解上面的描述时,许多其它实施例将对本领域技术人员来说是明显的。因此,所公开主题的范围应参照所附权利要求连同权利要求享有权利的等价物的全部范围来确定。

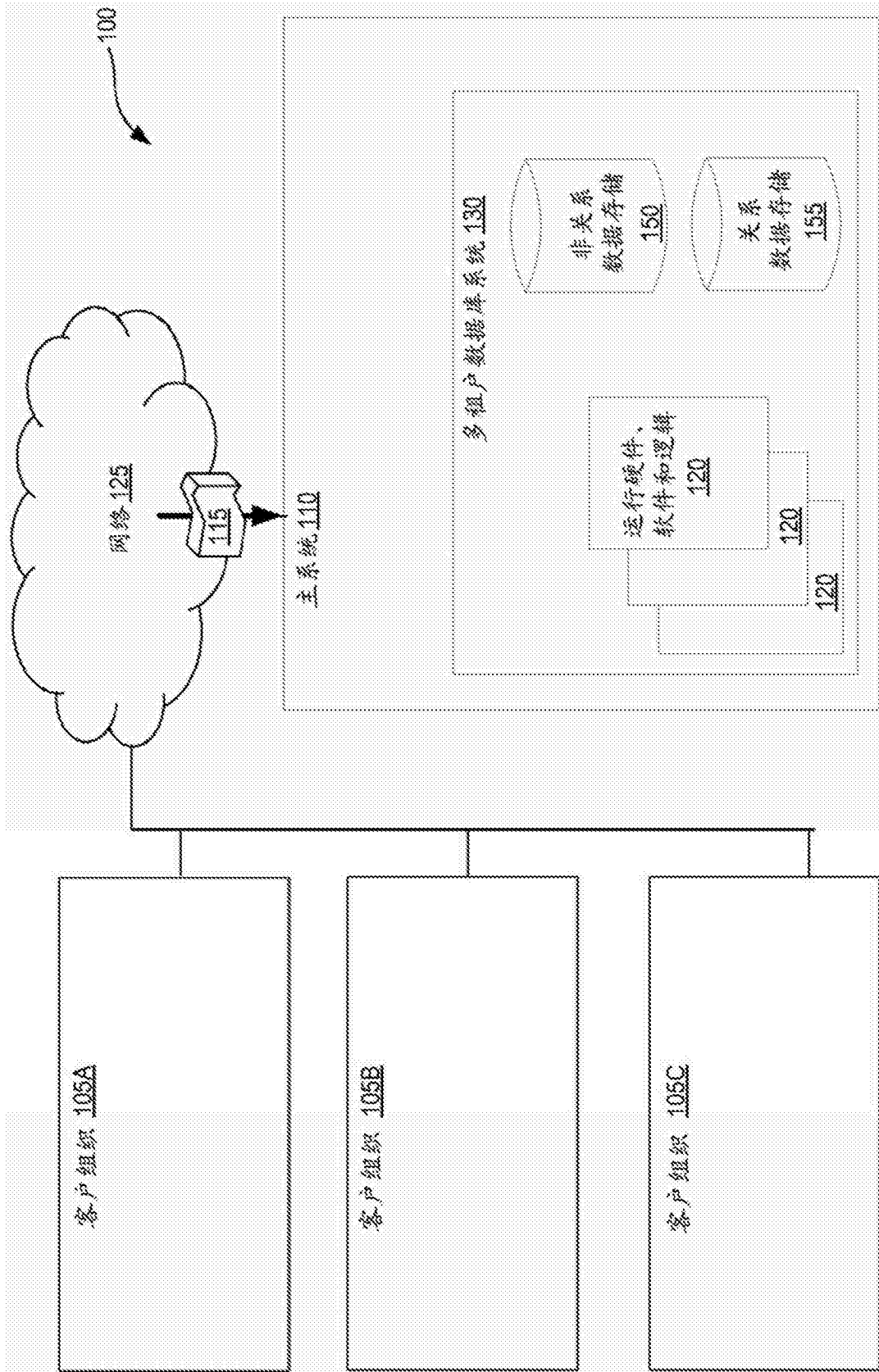


图 1

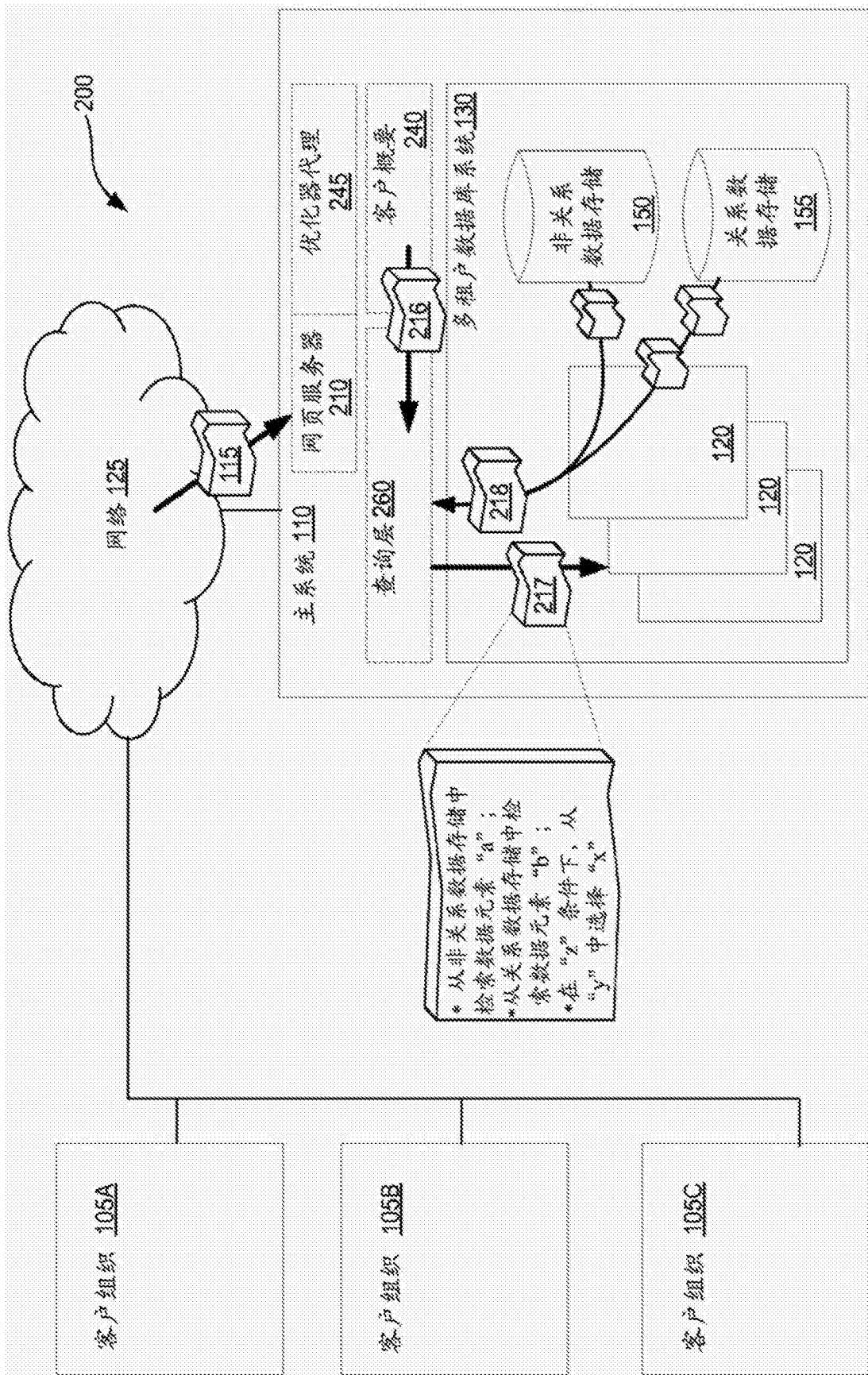


图 2

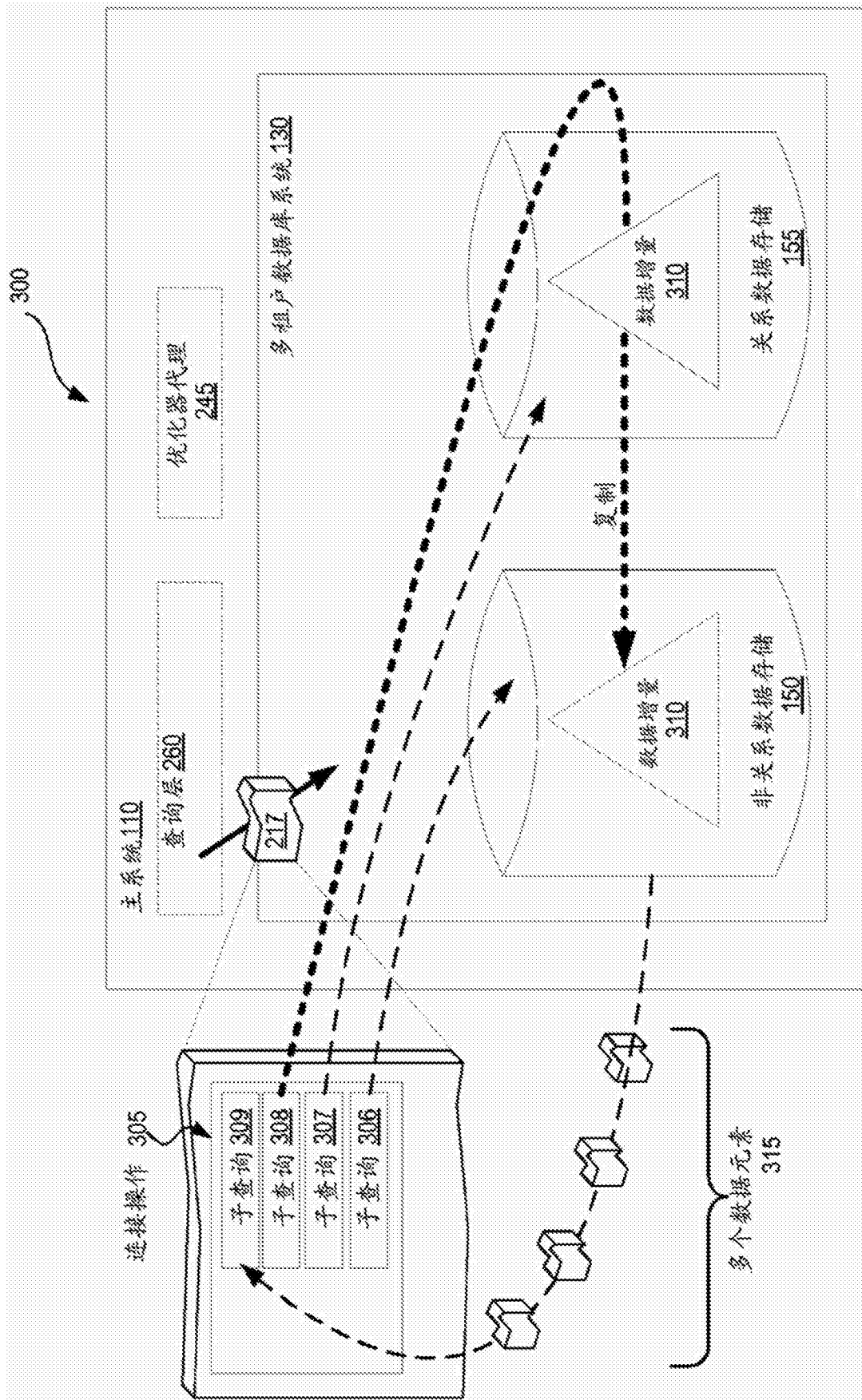


图 3

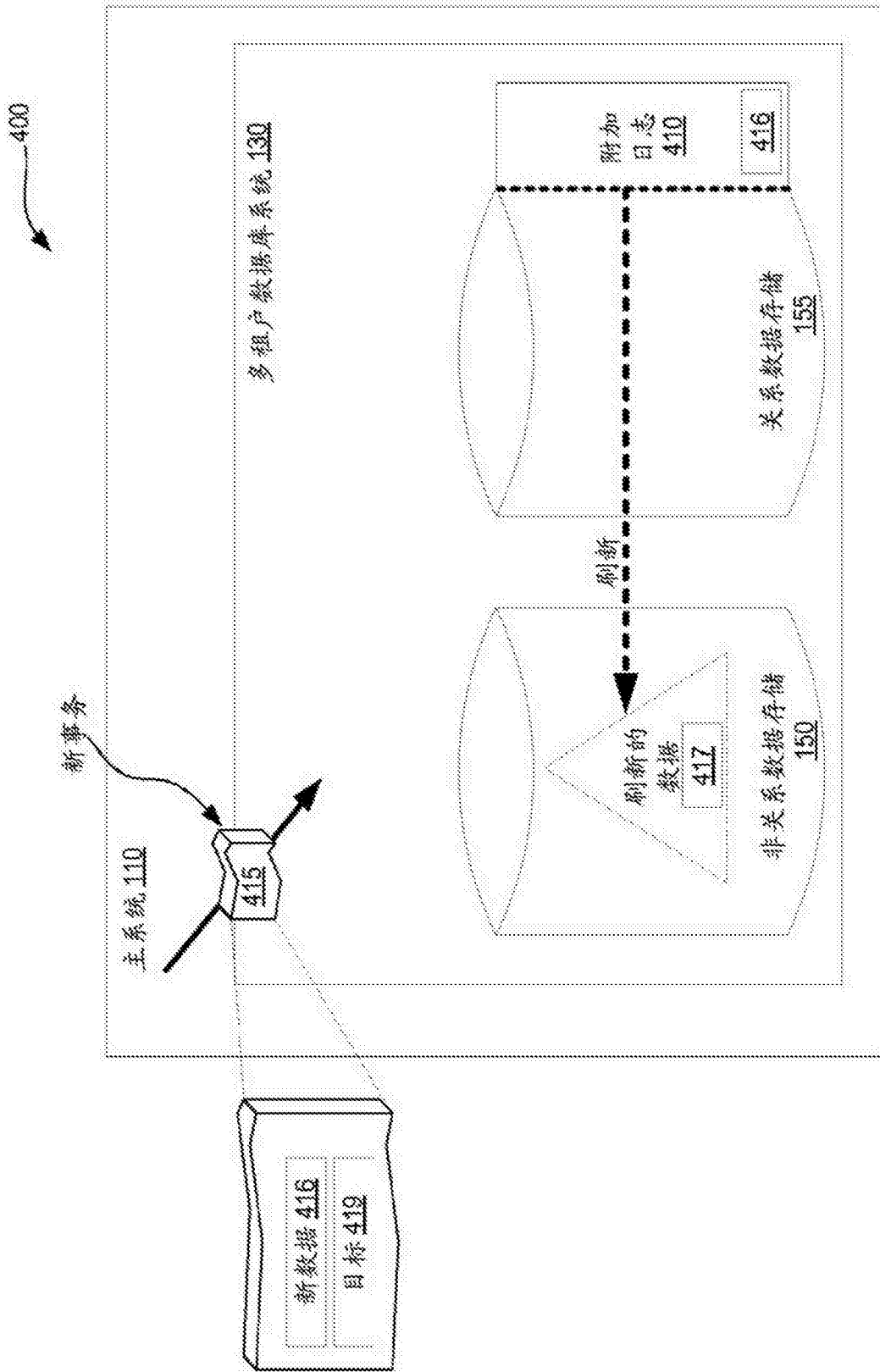


图 4

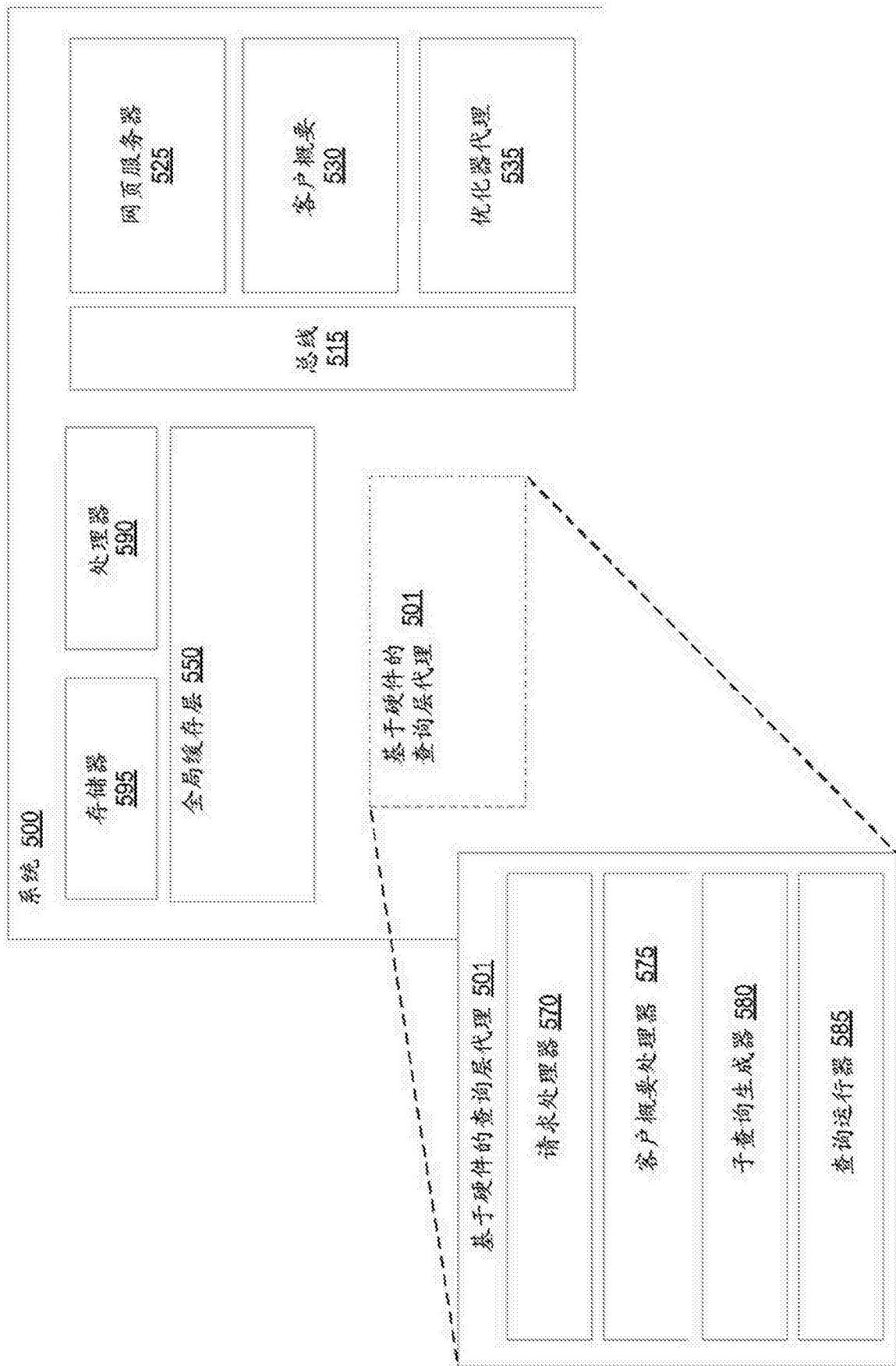


图 5

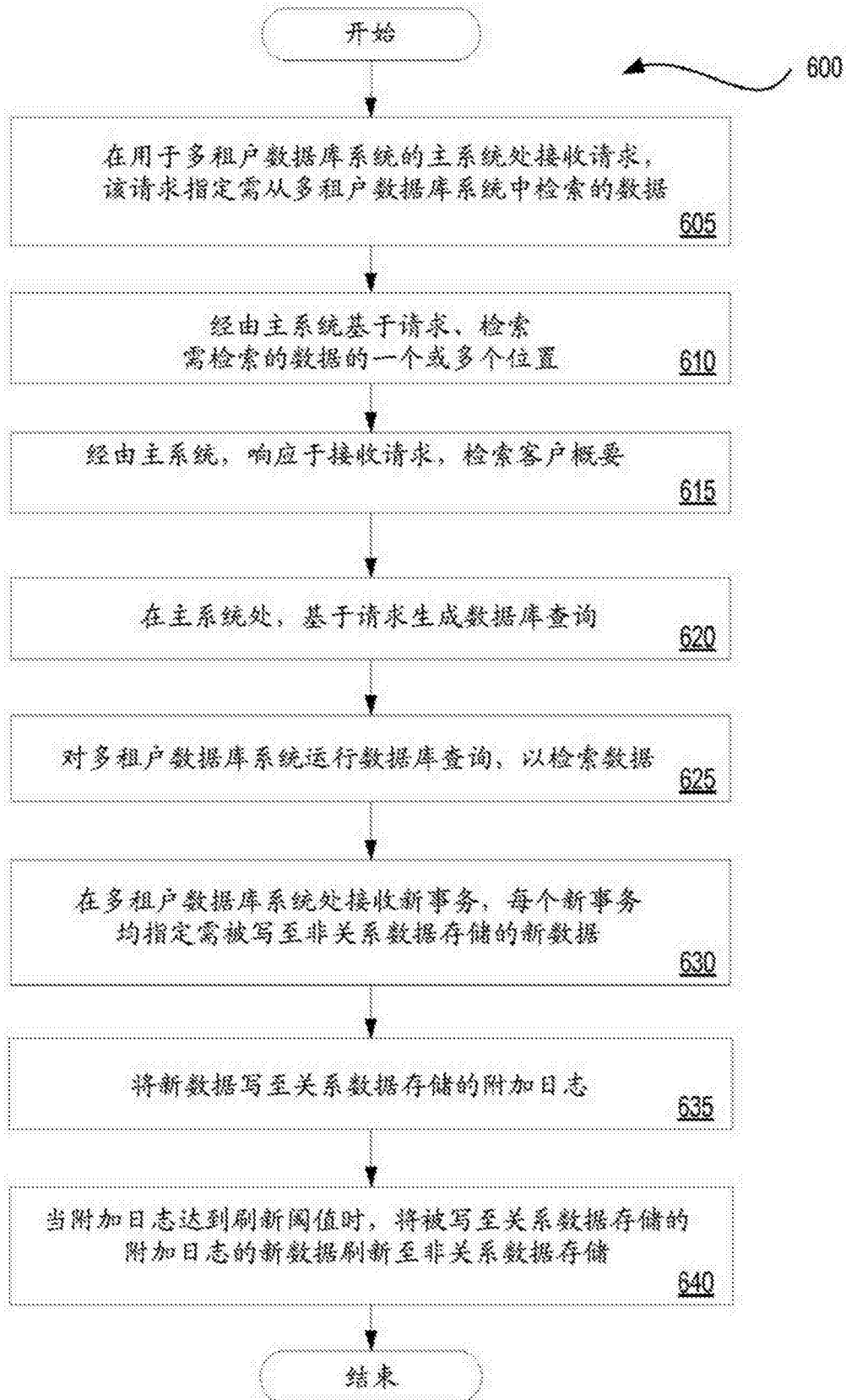


图 6

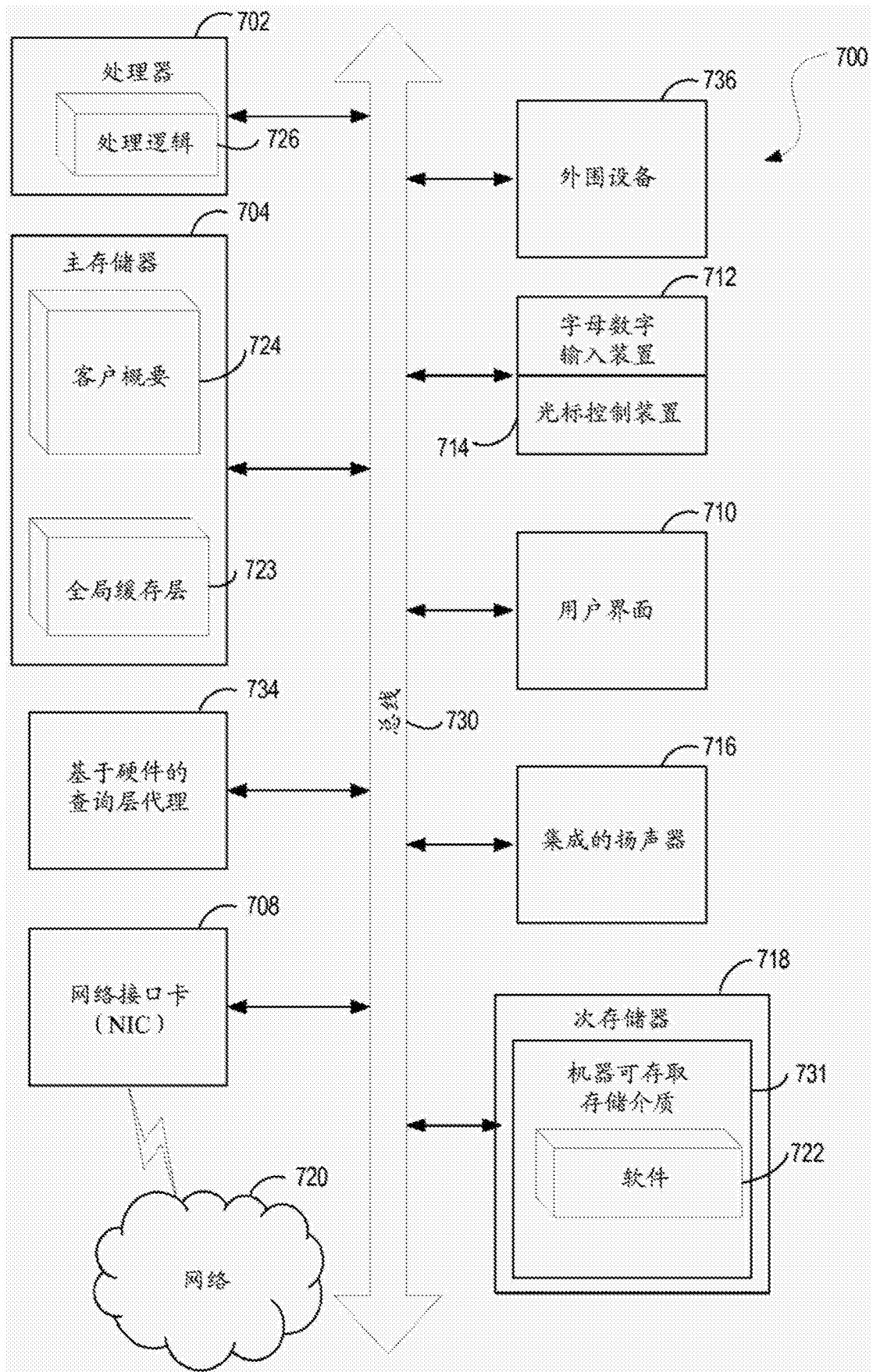


图 7