



(12) 发明专利申请

(10) 申请公布号 CN 101727480 A

(43) 申请公布日 2010.06.09

(21) 申请号 200910206036.4

(22) 申请日 2009.10.09

(30) 优先权数据

12/249,378 2008.10.10 US

(71) 申请人 ABB 技术有限公司

地址 瑞士苏黎世

(72) 发明人 M·卢托雅尔维 R·伊蒂南

T·维塔里

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华 庞淑敏

(51) Int. Cl.

G06F 17/30 (2006.01)

G06F 11/14 (2006.01)

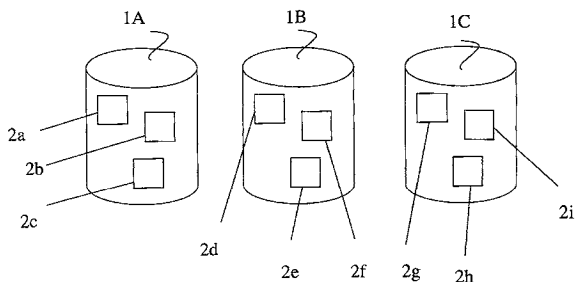
权利要求书 2 页 说明书 5 页 附图 3 页

(54) 发明名称

用于复制数据库的方法和系统

(57) 摘要

一种用于复制数据库的方法,在该方法中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和。在该方法中,数据库中相应表之间的数据的逻辑一致性通过比较数据库表的循环冗余校验值和来维持;且如果发现循环冗余校验值和之间的差异,则该树索引被逻辑地分割为两个子树索引,且该子树索引的循环冗余校验值和被相互比较,且该比较和分割一直持续到发现导致差异的数据,且在数据库之间复制不一致的数据。



1. 一种用于复制数据库的方法,在该方法中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和,在该方法中,通过比较数据库表的循环冗余校验值和来维持数据库中相应表之间的数据的逻辑一致性;且如果发现循环冗余校验值和之间的差异,则将该树索引逻辑地分割为两个子树索引并相互比较该子树索引的循环冗余校验值和,且使该比较和分割一直持续到发现导致差异的数据,并且在数据库之间复制不一致的数据。

2. 根据权利要求1所述的方法,其中该数据库中的索引结构是具有节点以及循环冗余校验值的B-树或B+树,且数据库表的循环冗余校验值和被存储在数据库表的唯一树索引的节点中。

3. 根据权利要求1所述的方法,其中该数据库存储在不同的计算机中。

4. 一种用于在信息管理系统中确保数据库一致性的复制数据库的方法,在该方法中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有实时数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和,在该方法中,通过比较该数据库表的循环冗余校验值和来维持该数据库中相应表之间的数据的逻辑一致性;且如果发现循环冗余校验值和之间的差异,则将该树索引逻辑地分割为两个子树索引且相互比较该子树索引的循环冗余校验值和,并且使该比较和分割一直持续到导致差异的数据被发现为止,并且不在数据库之间复制一致的数据,且将复制的数据库服务于该信息管理系统。

5. 根据权利要求4所述的方法,其中该数据库中的索引结构是具有节点和循环冗余校验值的B-树或B+树,且数据库表的循环冗余校验值和被存储在该数据库表的唯一树索引的节点中。

6. 根据权利要求4所述的方法,其中该数据库存储在不同的计算机中。

7. 根据权利要求4所述的方法,其中该信息管理系统是过程信息管理系统,且包含处理装置的测量值和/或操作参数的实时数据被供给到数据库表至少其一。

8. 一种用于复制数据库的设备,其中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和,所述设备包括:

用于通过比较该数据库表的循环冗余校验值和来维持该数据库中相应表之间的数据的逻辑一致性的装置;以及

用于如果发现循环冗余校验值和之间的差异,则将该树索引逻辑地分割为两个子树索引且相互比较该子树索引的循环冗余校验值和,且使该比较和分割一直持续到导致差异的数据被发现,并且在该数据库之间复制不一致的数据的装置。

9. 一种用于复制数据库的系统,包括:至少两个数据库,该数据库具有一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和;用于通过比较数据库表的循环冗余校验值和来维持数据库中相应表之间的数据的逻辑一致性的装置;用于发现循环冗余校验值和之间的差异的装置,用于将树索引逻辑分割为两个子树索引的装置,用于相互比较子树索引的循环冗余校验值和的装置,和用于使该比较和分割一直持续到导致差异的数据被发现的装置,以

及用于在数据库之间复制不一致的数据的装置。

10. 根据权利要求 9 所述的系统,其中该数据库存储在不同的计算机中。

用于复制数据库的方法和系统

技术领域

[0001] 本发明涉及用于复制数据库的方法、计算机可读介质和系统。

背景技术

[0002] 在计算机系统中数据常存储在数据库中。当数据库被复制时,相同的数据存储在多个存储装置中。复制被频繁地执行且它为相同数据库提供多个一致的副本。

[0003] 信息管理系统从若干个源搜集信息并管理该信息,且将信息分布到所限定的系统。信息例如可以从所有技术、商业或卫生保健领域收集。

[0004] 在诸如石油炼制的高级处理控制或者监控核电站的操作这样关键任务系统中,需要过程信息管理系统(PIMS)的高可靠性和不间断的操作。计算机系统中的任意单点故障必须不中断过程信息管理系统或在其上运行的应用程序的实时操作。

[0005] 典型地,已经使用PC集群技术来建立冗余过程信息管理系统。于是,系统和数据库此时运行在一个节点上,且在硬件或软件故障情况下,系统在独立的节点启动。故障总是导致系统操作中的不希望中断。

[0006] 关系数据库系统提供建立高可用性解决方案和数据库复制功能性的若干概念,但是它们的性能并不满足通常针对过程信息管理系统所设置的要求。

[0007] 一些过程历史产品能够提供过程历史的冗余,但是它们不能存储关系数据。

发明内容

[0008] 本发明的目的是创造一种用于复制数据库的方法、设备和系统。为了实现这一目的,本发明的特征在于权利要求1、4、8和9的特征部分中规定的特征。本发明的一些其他优选实施方式具有从属权利要求中规定的特性。

[0009] 一种用于复制数据库的方法,在该方法中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和。在该方法中,通过比较数据库表的循环冗余校验值和来维持数据库中相应表之间的数据的逻辑一致性;且如果发现循环冗余校验值和之间的差异,则将树索引逻辑地分割成两个子树索引且相互比较子树索引的循环冗余校验值和。使比较和分割一直持续到导致差异的数据被发现,并且在数据库之间复制不一致的数据。

[0010] 一种用于在信息管理系统中确保数据库一致性的复制数据库的方法。在该方法中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有实时数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和。通过比较数据库表的循环冗余校验值和来维持数据库中相应表之间的数据的逻辑一致性;且如果发现循环冗余校验值和之间的差异,则将树索引被逻辑地分割成两个子树索引且相互比较子树索引的循环冗余校验值和,且该比较和分割一直持续到导致差异的数据被发现。在数据库之间复制不一致的数据,并且将复制的数据库服务于信息管理系统。

统。

[0011] 一种用于复制数据库的设备,其中,至少两个数据库被复制,每个数据库包含一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和,所述设备包括:用于通过比较该数据库表的循环冗余校验值和来维持该数据库中相应表之间的数据的逻辑一致性的装置;以及用于如果发现循环冗余校验值和之间的差异,则将该树索引逻辑地分割为两个子树索引且相互比较该子树索引的循环冗余校验值和,且使该比较和分割一直持续到导致差异的数据被发现,并且在该数据库之间复制不一致的数据的装置。

[0012] 一种用于复制数据库的系统,包括:至少两个数据库,该数据库具有一个或多个表,且每个表具有数据和唯一树索引,该唯一树索引包含索引关键字和在每个树索引层上从数据计算的分层次的循环冗余校验值和;用于通过比较数据库表的循环冗余校验值和来维持数据库中相应表之间的数据的逻辑一致性的装置;用于发现循环冗余校验值和之间的差异的装置,用于将树索引逻辑分割为两个子树索引的装置,用于相互比较子树索引的循环冗余校验值和的装置和用于使得该比较和分割一直持续到导致差异的数据被发现的装置,以及用于在数据库之间复制不一致的数据的装置。

[0013] 在用于复制数据库的方法中,索引是引用(例如指向数据库文件或表中的记录或数据行的指针)的有序集。索引基于表的一列或多列。数据库中的索引结构例如是 B-树(B-tree)或 B+ 树。B-树是维持有序数据集且允许找到、删除、插入和浏览数据的有效操作的数据结构。

[0014] 用于复制数据库的方法和系统例如用于确保过程信息管理系统中的数据库的一致性。当在过程信息管理系统中提供数据库服务的服务器之一由于硬件故障或磁盘系统升级而停机时,其他服务器可以运行。提供给过程信息管理系统的数据被存储在运行的一个或多个服务器的数据库中。数据来自工厂的控制和监控系统,且包含来自各装置的测量得到的操作值或参数以及用于处理装置的计算得到的控制值。在服务器恢复之后,它丢失了停机期间提供给其他服务器的数据。然后,通过用于复制数据库的方法和系统来控制服务器中的数据库的一致性,即,将丢失或变化的数据传送到停机的服务器,且检查已经运行的服务器的一致性。当数据库的复制完成时,数据库具有相同的存储信息内容,且任意数据库可以在过程信息管理系统中工作。

[0015] 用于复制数据库的方法实现了在计算机系统中任意单点故障情况下的无中断操作以及在故障之后将系统复原为原始系统时的无中断操作。

[0016] 其他优点在于在负载共享情况下的高可用性和更好的数据存取性能。高可用性数据库提供了使用 PC 集群技术不能实现的冗余计算机之间的负载共享可能性。

[0017] 用于复制数据库的系统不需要事务存档或类似存档,而有一些关系数据库在复制中要使用该事务存档。事务存档消耗存储空间且降低性能。信息的复制和检查是独立的基于需要的过程。

[0018] 用于复制数据库的方法提供了数据库高可用性,且提供并发运行数据库系统的两个或多个实例而在单个计算机故障的情况下在系统中没有任何中断的可能性。用于复制数据库的系统提供了选项来建立在一个系统中含有若干复制数据库的系统以提供更高级别的可用性。负载共享功能性提供了将数据库系统扩展为包含大量计算和/或极大量并发用

户的解决方案的可能性。

[0019] 用于复制数据库的系统可以使用低成本的硬件实现。其实现了使用不昂贵的 PC 计算机来建立高度可靠的实时过程信息管理系统。

[0020] 在用于复制数据库的系统中,硬件可以是分散的,例如,数据库可以存储在不同的计算机中,无需任意专用硬件。从物理的视角,分散例如防止火险。

附图说明

[0021] 在下文中,将通过参考附图在特定实施方式的帮助下更详细地描述本发明,附图中:

[0022] 图 1 是用于数据库复制的系统的一般示图;

[0023] 图 2 是存储在 B+ 树中的分层次的循环冗余校验值的一般示图;

[0024] 图 3 是用于过程信息管理系统的数据库复制的系统的一般示图。

具体实施方式

[0025] 图 1 图示了数据库复制系统。

[0026] 在用于复制数据库的方法中,两个或更多数据库被彼此复制。在图 1 中,存在 3 个数据库 1A-1C。在本实施方式中,数据库 1A-1C 是符合关系模型的关系数据库。

[0027] 每个数据库 1A-1C 包括含有数据的表 2a-2i。在复制方法中,B- 树 (B-tree) 或 B+ 树 (B+tree) 应用为数据库表中的索引结构。B+ 树在图 2 中示出。它保持存储的数据且允许在逻辑分摊时间 (amortized time) 内进行搜索、插入和删除。存储在 B- 树中的数据片通常被称为关键字。在唯一索引中,每个关键字在逻辑上是唯一的,且可以只出现在 B- 树中的一个位置中。

[0028] 每个表 2a-2i 均具有唯一树索引,该唯一树索引包含索引关键字和从每个树索引层上的数据计算所得的分层次的循环冗余校验值和。树索引是条目的有序集。每个条目包含搜索关键字值和指向表中包含该值的特定行的指针。索引关键字是来自数据行的一个或多个字段组成的数据量。

[0029] 图 2 代表实施用于 48 个数据行的索引的简单三层 B+ 树的一部分。在叶节点 24 中,存储每个数据行的循环冗余校验值 ($C_{1...n}$) 及相应索引关键字 ($k_{1...n}$)。内部节点 23 包含指向叶节点的大量指针。相应叶节点中的相应索引关键字和所有循环冗余校验值和与每个指针存储在一起,最后的指针除外。最后的指针所指的叶节点的循环冗余校验值和被包括在上层内部节点的循环冗余校验和中,在该示例中,该上层内部节点是根节点 22。整个索引 21 中的所有循环冗余校验值和与指向根节点的指针存储在一起。

[0030] B- 树包括含有索引关键字的节点结构以及将 B- 树的节点链接到一起的指针。

[0031] 用于复制数据库的方法基于针对数据表 2a-2i 中所选逻辑唯一索引的 B 或 B+ 树的所有层计算和存储关于数据的循环冗余校验值和。其是存储结构且提供用以检查数据库表 2a-2i 的冗余副本的一致性且在发现不一致性时校正差异的服务。对于每个关键字,即,数据库中的数据行,计算循环冗余校验值,且将其存储到叶。关键字的循环冗余校验值相加在一起成为叶所属的内部节点的循环冗余校验值和,且内部节点的循环冗余校验值的和被进一步相加到它们所属的节点。最后,存在一个从数据库表中的所有数据计算的循环冗余

校验值和。

[0032] 通过比较数据库表的冗余副本的循环冗余校验值和,可以验证副本是否彼此逻辑一致。逻辑一致性意味着表包含有效数据。在冗余副本的循环冗余校验值和彼此不同的情况下,树索引被逻辑分割为两个子树索引,且将子树索引的循环冗余校验值和相互比较,一直到发现数据中的实际差异为止。不一致的数据在数据库之间复制。如果数据存在于表的若干副本中,数据的修改时间(行修改时间)用于决定哪个数据副本是最新的副本。使用保存在每个数据库中的删除日志检测删除的数据行。

[0033] 所述循环冗余校验值和的比较处理允许在 $O(\log n)$ 时间内发现两个任意大小的表的差异。例如,对于具有 1 000 000 000 行的两个表,将花费 30 次迭代发现副本之间的差异。然而,因为通常很多差异彼此靠近,所以使用例如 64 行的块反而使得迭代减小为 24。

[0034] 在 B- 树中,内部节点可以具有在某个预定范围内的数目可变的子节点。当数据被插入或者从节点删除时,其子节点的数目变化。为了维持该预定范围,内部节点可以结合或分离。因为一定范围的子节点被允许,B- 树不需要像其他自平衡搜索树那样频繁的重新平衡,但是可能会浪费一些空间,因为节点并不完全充满。子节点的数目的下限和上限对于特定实施方式一般是固定的。例如,在 2-3 的 B- 树中,每个内部节点可以仅具有 2 或 3 个子节点。

[0035] 当节点存取次数远超过节点内部的存取次数时,B- 树具有优于备选实施方式的实质性优点。这通常在大多数节点处于诸如硬盘驱动这样的辅助存储器中时发生。通过最大化每个内部节点内的子节点的数目,树的高度减小,平衡的发生更不频繁,且效率增加。通常,该值被设置为使得每个节点占据辅助存储器的整个磁盘存储块(disk block)或相当大小。尽管 2-3B- 树在主存储器中可能有用,且相当容易解释,但是如果节点大小变为磁盘存储盘块的大小,则结果可能是 257-513B- 树。

[0036] 每个 B- 树具有某种“次序 n ”,意味着节点包含从 n 到 $2n$ 个关键字,且节点的关键字由此总是至少半满。关键字以排列次序保持在每个节点内。相应指针列表有效地散布在关键字之间以指示如果关键字不在当前节点中则到哪里搜索该关键字。包含 k 个关键字的节点总是包含 $k+1$ 个指针。

[0037] B- 树的每个节点具有可变数目的关键字和子节点。关键字以非减小顺序存储。每个关键字具有相关的子节点,该子节点是包含的所有节点的关键字均小于或等于该关键字但大于前面关键字的子树的根。节点也具有附加的最右边的子节点,该子节点是包含的所有关键字均大于节点中任意关键字的子树的根。

[0038] B- 树具有针对每个节点的最小允许子节点数目,其已知为最小因子。如果 t 是最小因子,则每个节点必须具有至少 $t-1$ 个关键字。在某些状况下,根节点被允许违反这种属性,可以具有小于 $t-1$ 个关键字。每个节点最多可以具有 $2t-1$ 个关键字,或等价地具有 $2t$ 个子节点。

[0039] 因为每个节点倾向于具有大的分支因子(大量的子节点),典型地,必须在定位所需关键字之前遍历相对少的节点。如果对于每个节点的存取需要磁盘存取,则 B- 树将最小化所需的磁盘存取的数目。最小因子通常被选择为使得每个节点的总大小对应于底层存储装置的存储块大小的倍数。这种选择简化且优化了磁盘存取。因此,B- 树是用于当所有数据不能存储在主存储器中且存取辅助存储器相对昂贵或耗时的情况时的理想数据结构。

[0040] 图 3 图示了用于过程信息管理系统的数据复制系统。

[0041] 过程信息系统是信息管理系统的一个示例。过程信息管理系统 (PIMS) 30 与各种控制或其他系统通信接口,以使用一秒或更多或更少的采样时间收集过程数据。过程信息管理系统提供应用程序 31、分布控制系统 (DCS) 32、可编程逻辑控制器 (PLC) 33、实验室信息管理系统 (LIMS) 34、制造执行系统 (MES) 35、协调生产管理系统 (CPM) 36、计算机化维护管理系统 (CMMS) 37、先进过程控制应用程序 (APC) 38 以及各方面的其他信息技术系统之间的数据交换。过程信息管理系统提供数据处理服务、数据采集和存取接口、数据可视化和分析工具、计算和应用程序开发工具,且集成所有有效值和信息,并将它们存储在大型、有效的实时关系数据库中。

[0042] 过程信息管理系统通常具有两个或更多向其他计算机程序或计算机提供数据库服务的独立服务器。在该示例中,存在两个服务器 ID、IE,且它们均包含在彼此复制的数据库之间独立地工作的、数据产生以及历史记录和应用。在该实施方式中,数据库存储在不同的计算机中。

[0043] 实时过程信息管理具有以下特性,例如:以每秒 10 至 100000 个数据行的速度连续地将来自处理装置的数据流输入到数据库中,用于历史记录;将输入数据流的连续数据处理、聚集和存储为聚集历史值;在相同数据库中存储其他关联应用数据;1 秒响应时间,对于应用和用户而言具有高数据存取性能。

[0044] 当过程信息管理系统的服务器之一 ID 例如由于硬件故障或磁盘系统升级而停机时,另一服务器 IE 运行。于是例如包含过程装置的测量值和/或操作参数的实时数据仅被供给到被升级的运行服务器 IE 的数据库表 2j-2l。在服务器 ID 恢复之后,在运行时间连续地验证和控制服务器中的数据库的一致性。数据库复制系统具有自动一致性控制和恢复。数据库是很多表,且一致性控制和恢复基于关于每一行的时间戳和校验和以及逻辑唯一索引中的循环冗余校验值和。具有最新时间戳的行被设定为正确的行,且在检测行的列表上搜索丢失的行。

[0045] 在服务器之一中硬件故障的情况下,在过程信息管理系统的操作中不存在中断。磁盘故障也不导致信息流的故障。在故障之后可以不中断地重建系统,且系统能够承受计算机的若干次连续故障而没有过程信息管理系统的操作的中断或者数据的丢失。

[0046] 用于复制数据库的系统还允许磁盘系统在不中断过程信息管理系统的情况下升级。

[0047] 用于复制数据库的该方法有利地使用计算机实施。使用的程序被存储在计算机的内存或存储在可以被装载到计算装置上的计算机可读介质(例如 DVD)中。这些计算机可读介质具有使得计算机执行方法的指令。

[0048] 已经在特定实施方式的帮助下描述了本发明。然而,描述不应理解为限制了专利保护的范围;本发明的实施方式可以在下述权利要求书的范围内变化。

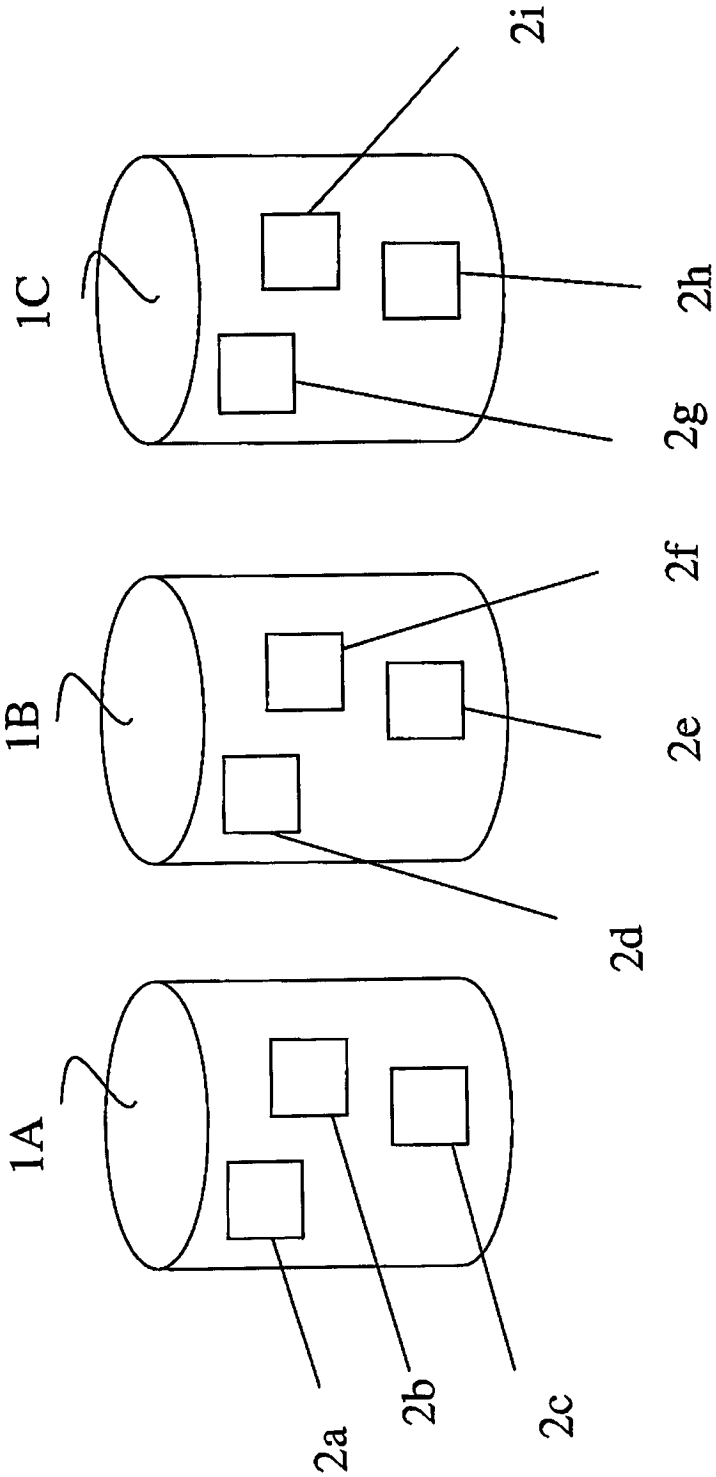


图 1

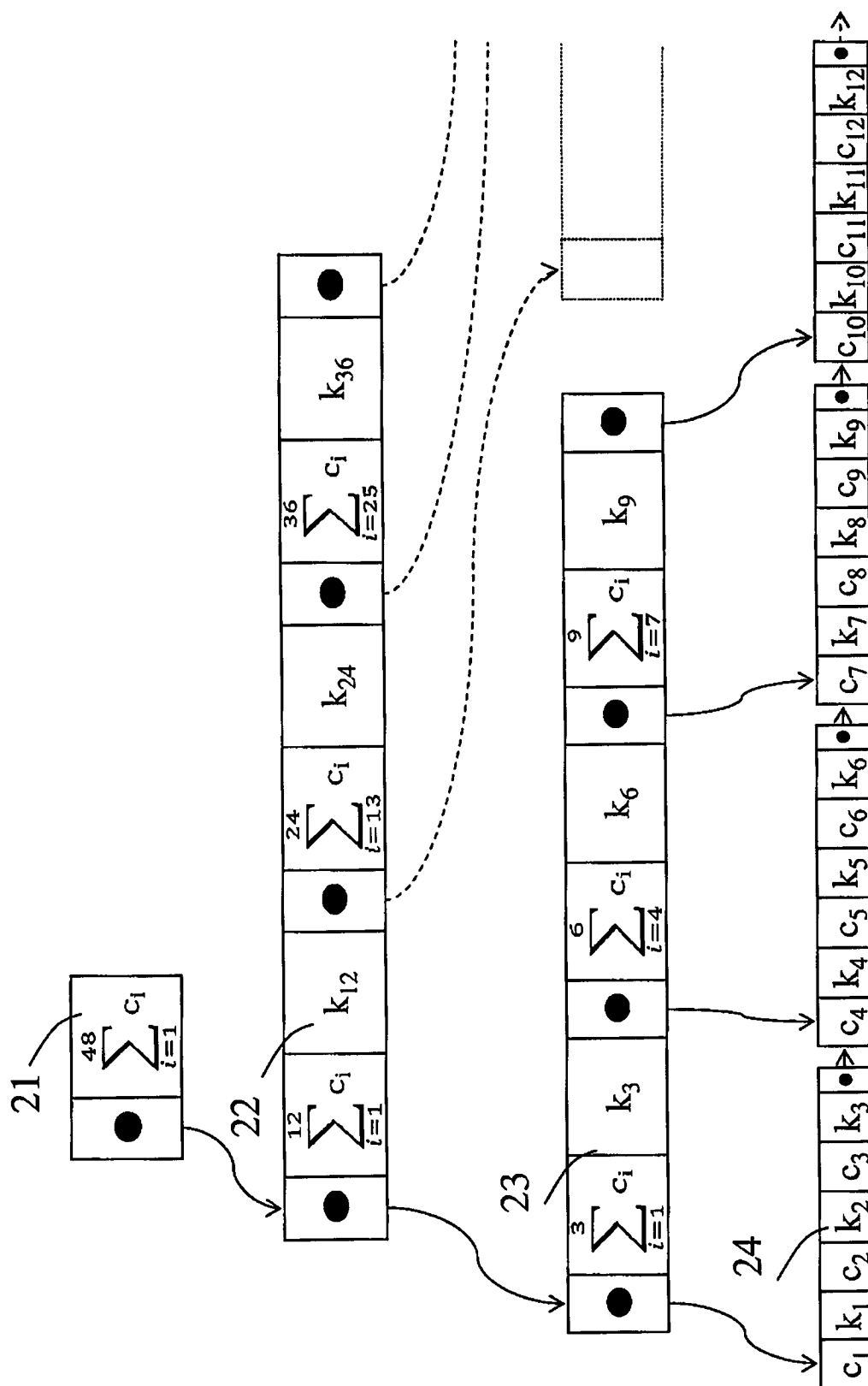


图 2

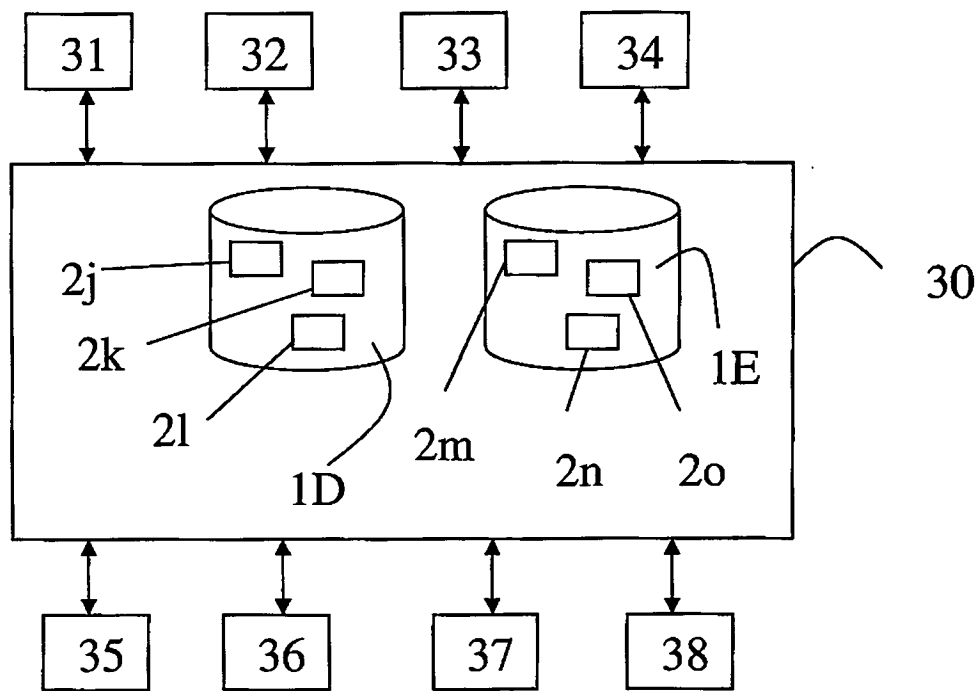


图 3