



(12) 发明专利

(10) 授权公告号 CN 102521073 B

(45) 授权公告日 2014. 09. 17

(21) 申请号 201110383006. 8

CN 1581091 A, 2005. 02. 16, 全文.

(22) 申请日 2011. 11. 16

US 2007/0185852 A1, 2007. 08. 09, 全文.

US 6795830 B1, 2004. 09. 21, 全文.

(30) 优先权数据

12/948, 541 2010. 11. 17 US

Bettina Kemme 等. Online Reconfiguration in Replicated Databases Based on Group Communication. 《IEEE》. 2001, 第 5.2 节、图 2.

(73) 专利权人 微软公司

地址 美国华盛顿州

审查员 彭杰

(72) 发明人 V·沙阿 S·O·沃蒂莱宁

T·塔留斯

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 顾嘉运

(51) Int. Cl.

G06F 11/14 (2006. 01)

G06F 17/30 (2006. 01)

(56) 对比文件

US 6539381 B1, 2003. 03. 25, 说明书第 5 栏第 1-21 行, 第 9 栏第 33-60 行, 第 16 栏第 31-43 行、图 3-4 和 9.

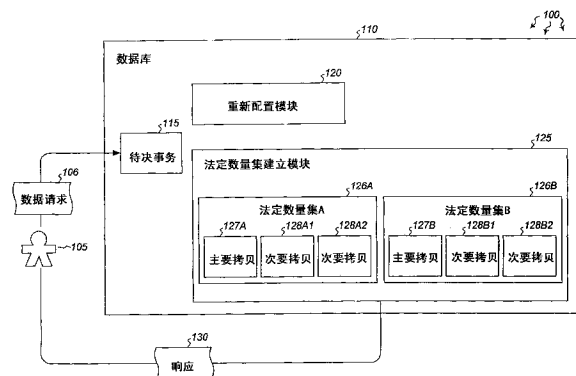
权利要求书 3 页 说明书 7 页 附图 4 页

(54) 发明名称

在故障恢复期间增加数据库的可用性

(57) 摘要

本申请涉及在故障恢复期间增加数据库的可用性。在此描述的实施例旨在在数据库重新配置期间提供数据库访问并在数据库重新配置期间维护复制连接。在一个实施例中, 计算机系统建立多个法定数量拷贝集以复制数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务。计算机系统确定数据分区重新配置已经开始, 并且使用法定数量拷贝集中的每一个中的至少法定数量个拷贝来在数据分区的重新配置期间提供对数据分区的数据的访问。



1. 一种在包括多个计算系统的计算机联网环境中,在包括处理器和存储器的计算机系统用于在数据库重新配置期间提供数据库访问的计算机实现的方法,所述方法包括:

对于数据库的给定数据分区,建立包括第一法定数量拷贝集和第二法定数量拷贝集(126A/126B)的两个法定数量拷贝集来复制所述给定数据分区的数据的动作,这样,对于同一数据分区而言存在两个法定数量拷贝集,其中所述法定数量拷贝集确保了至少最小数目的拷贝用于在分区重新配置(120)期间提交待决事务,并且其中所述法定数量集包括在这两个所述法定数量集中都具有成员关系的至少一个主要拷贝;

确定数据分区的重新配置已经启动的动作,执行所述数据分区的重新配置以将法定数量集的配置改变为新配置;以及

在所述数据分区的重新配置(120)期间通过下述动作使用所述两个法定数量拷贝集中的每一个中的至少法定数量个拷贝(127A-128A2)来提供对所述数据分区的数据的访问的动作:

当重新配置开始时,改变所述第一法定数量集中的所有拷贝的法定数量成员关系,以使得所述第一法定数量集中的所有拷贝是所述第一和第二法定数量集这两者中的成员;

将所述第二法定数量集改变为所述新配置,但所述第二法定数量集包括在这两个所述法定数量集中都具有成员关系的所述主要拷贝并具有足够数目的拷贝和拷贝分布来在重新配置期间提交待决事务,这样,所述第二法定数量集可以被用于在重新配置期间提交待决事务;

将被配置为所述新配置的所述第二法定数量集中的所有拷贝的成员关系改变为是所述第一法定数量集的成员,以便所述第一法定数量集被配置为所述新配置;以及

改变不属于所述新配置的任何拷贝的所述成员关系,以便它们不再是任意法定数量集的一部分。

2. 如权利要求1所述的方法,其特征在于,每个拷贝集包括多个次要拷贝。

3. 如权利要求2所述的方法,其特征在于,所述法定数量拷贝集中的至少一个是被实例化为在重新配置期间对请求进行应答的临时法定数量拷贝集。

4. 如权利要求3所述的方法,其特征在于,在重新配置结束之后,移除所述临时法定数量拷贝集。

5. 如权利要求1所述的方法,其特征在于,在重新配置的多个阶段期间提供对所述数据库分区的访问。

6. 如权利要求1所述的方法,其特征在于,由所述多个法定数量拷贝集中的拷贝离开或加入拷贝集来启动重新配置。

7. 如权利要求6所述的方法,其特征在于,进一步包括防止现有的数据库复制连接由于拷贝的离开而被断开。

8. 一种用于在数据库重新配置期间维护复制连接的方法,包括:

对于数据库的给定数据分区,建立包括第一法定数量拷贝集和第二法定数量拷贝集(126A/126B)的两个法定数量拷贝集来复制所述给定数据分区的数据的动作,这样,对于同一数据分区而言存在两个法定数量拷贝集,其中所述法定数量拷贝集确保了至少最小数目的拷贝(127A-128A2)用于在分区重新配置(120)期间提交待决事务,并且其中所述法定数量集包括在这两个所述法定数量集中都具有成员关系的至少一个主要拷贝;

对于所述数据库的所述给定数据分区,确定拷贝(128A1)的离开已经启动数据分区的重新配置的动作,执行所述数据分区的重新配置以将法定数量集的配置改变为新配置;以及

在所述数据分区的重新配置(120)期间使用所述两个法定数量拷贝集(126A/126B)中的每一个中的至少法定数量个拷贝(127A-128A2)来提供对所述数据分区的数据的访问的动作,这样,可以通过下述动作在即使拷贝离开的情况下也能维护数据库复制连接:

当重新配置开始时,改变所述第一法定数量集中的所有拷贝的法定数量成员关系,以使得所述第一法定数量集中的所有拷贝是所述第一和第二法定数量集这两者中的成员;

将所述第二法定数量集改变为所述新配置,但所述第二法定数量集包括在这两个所述法定数量集中都具有成员关系的所述主要拷贝并具有足够数目的拷贝和拷贝分布来在重新配置期间提交待决事务,这样,所述第二法定数量集可以被用于在重新配置期间提交待决事务;

将被配置为所述新配置的所述第二法定数量集中的所有拷贝的成员关系改变为是所述第一法定数量集的成员,以便所述第一法定数量集被配置为所述新配置;以及

改变不属于所述新配置的任何拷贝的所述成员关系,以便它们不再是任意法定数量集的一部分。

9. 如权利要求8所述的方法,其特征在于,以所述离开的拷贝的分区在重新配置期间保持在事务上一致的方式来移除所述拷贝。

10. 一种用于在数据库重新配置期间提供数据库访问的方法,所述方法包括:

对于数据库的给定数据分区,建立包括第一法定数量拷贝集和第二法定数量拷贝集(126A/126B)的两个法定数量拷贝集(126A/126B)来复制所述给定数据分区的数据的动作,这样,对于同一数据分区而言存在两个法定数量拷贝集,其中所述法定数量拷贝集确保了至少最小数目的拷贝(127A-128A2)用于在分区重新配置期间提交待决事务,每个拷贝集包括一个主要拷贝(127A)和至少一个次要拷贝(128A1),并且其中所述法定数量集包括在这两个所述法定数量集中都具有成员关系的至少一个主要拷贝;

确定数据分区的重新配置(120)已经启动的动作,执行所述数据分区的重新配置以将法定数量集的配置改变为新配置;

在所述数据分区的重新配置(120)期间通过下述动作使用所述两个法定数量拷贝集(126A/126B)中的至少两个拷贝(127A/128A1)来提供对所述数据分区的数据的访问的动作:

当重新配置开始时,改变所述第一法定数量集中的所有拷贝的法定数量成员关系,以使得所述第一法定数量集中的所有拷贝是所述第一和第二法定数量集这两者中的成员;

将所述第二法定数量集改变为所述新配置,但所述第二法定数量集包括在这两个所述法定数量集中都具有成员关系的所述主要拷贝并具有足够数目的拷贝和拷贝分布来在重新配置期间提交待决事务,这样,所述第二法定数量集可以被用于在重新配置期间提交待决事务;

将被配置为所述新配置的所述第二法定数量集中的所有拷贝的成员关系改变为是所述第一法定数量集的成员,以便所述第一法定数量集被配置为所述新配置;以及

改变不属于所述新配置的任何拷贝的所述成员关系,以便它们不再是任意法定数量集

的一部分。

11. 一种用于在数据库重新配置期间维护复制连接的系统,包括:

对于数据库的给定数据分区,用于建立包括第一法定数量拷贝集和第二法定数量拷贝集(126A/126B)的两个法定数量拷贝集(126A/126B)来复制所述给定数据分区的数据的装置,这样,对于同一数据分区而言存在两个法定数量拷贝集,其中所述法定数量拷贝集确保了至少最小数目的拷贝(127A-128A2)用于在分区重新配置(120)期间提交待决事务,并且其中所述法定数量集包括在这两个所述法定数量集中都具有成员关系的至少一个主要拷贝;

对于所述数据库的所述给定数据分区,用于确定拷贝(128A1)的离开已经启动数据分区的重新配置并执行所述数据分区的重新配置以将法定数量集的配置改变为新配置的装置;

以及

用于在所述数据分区的重新配置(120)期间使用所述两个法定数量拷贝集(126A/126B)中的每一个中的至少法定数量个拷贝(127A-128A2)来提供对所述数据分区的数据的访问的装置,这样,可以通过下述动作在即使拷贝离开的情况下也能维护数据库复制连接:

当重新配置开始时,改变所述第一法定数量集中的所有拷贝的法定数量成员关系,以使得所述第一法定数量集中的所有拷贝是所述第一和第二法定数量集这两者中的成员;

将所述第二法定数量集改变为所述新配置,但所述第二法定数量集包括在这两个所述法定数量集中都具有成员关系的所述主要拷贝并具有足够数目的拷贝和拷贝分布来在重新配置期间提交待决事务,这样,所述第二法定数量集可以被用于在重新配置期间提交待决事务;

将被配置为所述新配置的所述第二法定数量集中的所有拷贝的成员关系改变为是所述第一法定数量集的成员,以便所述第一法定数量集被配置为所述新配置;以及

改变不属于所述新配置的任何拷贝的所述成员关系,以便它们不再是任意法定数量集的一部分。

## 在故障恢复期间增加数据库的可用性

### 技术领域

[0001] 本申请涉及在故障恢复期间增加数据库的可用性。特别是,在数据分区的重新配置期间提供对数据分区的数据的访问。

### 背景技术

[0002] 计算机已变得高度集成于工作、家庭、移动设备以及许多其他地方中。计算机可快速且有效地处理大量信息。被设计成在计算机系统上运行的软件应用允许用户执行包括商业应用、学校作业、娱乐等等在内的各种各样的功能。软件应用通常被设计成执行特定任务,诸如用于草拟文档的文字处理器应用或者用于发送、接收和组织电子邮件的电子邮件程序。

[0003] 在许多情况下,软件应用程序被设计为与其他软件应用程序或其他计算机系统交互。例如,因特网浏览器将用户请求发送给web服务器,并且,这些web服务器通过对用户的请求进行响应来应答。web服务器和其它计算机系统可以被配置为访问数据存储以作为对用户请求的响应的一部分。这些数据存储可以存储有大量的信息,并且可包括复制数据以用于附加的冗余性的拷贝。在一些情况下,这些拷贝被分组在一起作为拷贝集或群集。当拷贝集中的一个拷贝变得不可用且随后恢复在线时,拷贝集必须被更新并重新配置。在这种重新配置期间,拷贝集不可用于对数据读或写请求进行响应。

### 发明内容

[0004] 在此描述的实施例旨在在数据库重新配置期间提供数据库访问并在数据库重新配置期间维护复制连接。在一个实施例中,计算机系统建立多个法定数量拷贝集以复制数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务。计算机系统确定数据分区重新配置已经启动,并且使用在法定数量拷贝集中的每一个中的至少法定数量个拷贝来在数据分区的重新配置期间提供对数据分区的数据的访问。

[0005] 在另一个实施例中,计算机系统建立多个法定数量拷贝集以复制数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务。计算机系统确定拷贝的离开已经启动数据分区的重新配置。计算机系统防止现有的数据库复制连接在拷贝离开时就断开,并且使用数据分区重新配置期间所维护的法定数量拷贝集中的每一个中的至少法定数量个拷贝来在重新配置期间提供对数据分区的数据的访问。

[0006] 提供本概述以便以简化形式介绍在以下详细描述中进一步描述的一些概念。本发明内容并非旨在标识所要求保护的的主题的关键特征或必要特征,也不旨在用于帮助确定所要求保护的的主题的范围。

[0007] 另外的特征和优点将在以下的描述中阐述,并且部分可从该描述中显而易见,或者可以从此处的教示实践中习得。本发明的各特征和优点可以通过在所附的权利要求书中

特别指出的手段和组合来实现和获得。本发明的特征将从以下描述和所附权利要求书中变得完全显而易见,或者可通过如下所述对本发明的实践而获知。

### 附图说明

[0008] 为了进一步阐明本发明的各实施例的以上和其他优点和特征,将参考附图来呈现本发明的各实施例的更具体的描述。可以理解,这些附图只描绘本发明的典型实施例,因此将不被认为是对其范围的限制。本发明将通过使用附图用附加特征和细节来描述和解释,附图中:

[0009] 图 1 示出本发明的实施例可以在其中操作的计算机体系结构,所述操作包括在数据库重新配置期间提供数据库访问以及在数据库重新配置期间维护复制连接。

[0010] 图 2 示出用于在数据库重新配置期间提供数据库访问的示例方法的流程图。

[0011] 图 3 示出用于在数据库重新配置期间维护复制连接的示例方法的流程图。

[0012] 图 4 示出重新配置过程的流程图。

### 具体实施方式

[0013] 在此描述的实施例旨在在数据库重新配置期间提供数据库访问并在数据库重新配置期间维护复制连接。在一个实施例中,计算机系统建立多个法定数量(quorum)拷贝集以复制数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务。计算机系统确定数据分区重新配置已经开始,并且使用在法定数量拷贝集中的每一个中的至少法定数量个拷贝来在数据分区的重新配置期间提供对数据分区的数据的访问。

[0014] 在另一个实施例中,计算机系统建立多个法定数量拷贝集以复制数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务。计算机系统确定拷贝的离开已经启动数据分区的重新配置。计算机系统防止现有的数据库复制连接在拷贝离开时就断开,并且使用法定数量拷贝集中的每一个中的至少法定数量个拷贝来在数据分区的重新配置期间提供对数据分区的数据的访问。

[0015] 以下讨论现涉及可以执行的多种方法以及方法动作。应当注意,虽然这些方法动作可能是按一定次序讨论的,或者是在流程图中被描绘为是按照特定顺序进行的,然而并非必然需要特定的次序,除非特别声明,或者是因为一个动作依赖于另一动作在该动作被执行之前完成而需要的。

[0016] 本发明的各实施例可包括或利用专用或通用计算机,该专用或通用计算机包括诸如例如一个或多个处理器和系统存储器等计算机硬件,如以下更详细讨论的。本发明范围内的各实施例还包括用于携带或存储计算机可执行指令和/或数据结构的物理介质和其他计算机可读介质。这些计算机可读介质可以是通用或专用计算机系统能够访问的任何可用介质。存储计算机可执行指令的计算机可读介质是计算机存储介质。携带计算机可执行指令的计算机可读介质是传输介质。由此,作为示例而非限制,本发明的各实施例可包括至少两种完全不同类型的计算机可读介质:计算机存储介质和传输介质。

[0017] 计算机存储介质包括 RAM、ROM、EEPROM、CD-ROM 或其他光盘存储、磁盘存储或其他磁存储设备、或可用于存储计算机可执行指令或数据结构形式的所需程序代码装置且可由

通用或专用计算机访问的任何其他介质。

[0018] “网络”被定义为允许在计算机系统和 / 或模块和 / 或其他电子设备之间传输电子数据的一个或多个数据链路。当信息通过网络或另一个通信连接（硬连线、无线、或者硬连线或无线的组合）传输或提供给计算机时，该计算机将该连接适当地视为传输介质。传输介质可包括可用于携带计算机可执行指令或数据结构形式的所需程序代码装置且可由通用或专用计算机访问的网络和 / 或数据链路。上述的组合也应被包括在计算机可读介质的范围内。

[0019] 此外，在到达各种计算机系统组件之后，计算机可执行指令或数据结构形式的程序代码装置可从传输介质自动传输到计算机存储介质（或反之亦然）。例如，通过网络或数据链路接收到的计算机可执行指令或数据结构可被缓存在网络接口模块（例如，“NIC”）内的 RAM 中，然后最终被传输到计算机系统 RAM 和 / 或计算机系统处的较不易失性的计算机存储介质。因而，应当理解，计算机存储介质可被包括在还利用（或甚至主要利用）传输介质的计算机系统组件中。

[0020] 计算机可执行指令包括，例如使通用计算机、专用计算机、或专用处理设备执行某一功能或某组功能的指令和数据。计算机可执行指令可以是例如二进制代码、诸如汇编语言之类的中间格式指令、或甚至源代码。尽管用结构特征和 / 或方法动作专用的语言描述了本主题，但可以理解，所附权利要求书中定义的主题不必限于上述特征或动作。相反，上述特征和动作是作为实现权利要求的示例形式而公开的。

[0021] 本领域的技术人员将理解，本发明可以在具有许多类型的计算机系统配置的网络计算环境中实践，这些计算机系统配置包括个人计算机、台式计算机、膝上型计算机、消息处理器、手持式设备、多处理器系统、基于微处理器的或可编程消费电子设备、网络 PC、小型计算机、大型计算机、移动电话、PDA、寻呼机、路由器、交换机等等。本发明也可在其中通过网络链接（或者通过硬连线数据链路、无线数据链路，或者通过硬连线和无线数据链路的组合）的本地和远程计算机系统两者都执行任务的分布式系统环境中实施。在分布式系统环境中，程序模块可以位于本地和远程存储器存储设备中。

[0022] 图 1 示出了可在其中采用本发明的原理的计算机体系结构 100。计算机体系结构 100 包括数据库 110。数据库可以是任意类型的数据库或数据存储系统，并且可以包括在一个或多个计算系统上的存储设备。例如，数据库可以位于一个组织或机构本地内，或者可以分布在跨宽广地理区域的许多不同的计算机系统上。数据库 110 可以包括存储区域网络（SAN）或其它存储方案。数据库是通过因特网可访问的，并且可被配置为从用户接收请求。例如，用户 105 可以发送请求数据库所提供的数据和 / 或服务的数据请求。这些请求可以由数据库作为待决事务 115 来存储。

[0023] 数据库事务确保了当执行用户请求时没有数据被丢失。例如，如果用户正在请求更新他们的银行数据，则数据库事务将保证该数据如用户所请求那样被更新。可以以拷贝的形式来备份数据库的数据。例如，每个数据存储分区可以具有一个或多个数据拷贝。如在图 1 中所示，这些拷贝可以是法定数量拷贝（例如 126A/126B）集的一部分。每个法定数量集可以包括多个不同的拷贝。虽然示出法定数量集 A 和 B 具有一个主要拷贝集（127A/127B）和两个次要拷贝集（128A1/128A2/128B1/128B2），应该可以理解可以使用不同数目的拷贝。法定数量集建立模块 125 可以基于各种准则建立各种不同数目的法定数量集。在一些情况

中,每个数据分区具有一个法定数量集,并且随后被分配第二、临时的法定数量以在重新配置期间使用。

[0024] 数据库重新配置可以在当拷贝停工(即由于计算机故障、网络故障或某个其它问题而停止工作)或恢复(即再次开始工作)时发生。这样,例如,如果次要拷贝 128A1 将要停工,就将需要重新配置法定数量集 126A。类似地,如果次要拷贝 128A1 将要在稍后某个时间点上再次恢复,则将需要再次重新配置法定数量集 126A。可以使用重新配置模块 120 来以这样的方式来重新配置法定数量拷贝集:在进行重新配置的同时向用户提供数据库服务。这些和其他概念将在下面参考图 2 和 3 更详细地解释。

[0025] 如上所示,即使在数据库分区正在经历重新配置过程期间,也可以在数据库分区上执行读和/或写操作。在一些实施例中,这可以通过维护多个动态法定数量集以便允许对分区的读/写访问同时保持其在重新配置处理期间在事务上一致来实现。在重新配置过程期间可以防止在现有拷贝之间的复制连接的断开。这可以允许用户(例如 105)在重新配置过程期间执行读/写操作。可以以如下方式将拷贝加入到各法定数量集中或从法定数量集中移除:分区在重新配置期间并且在存在用户事务的情况下保持在事务上一致。而且,可以防止依赖于从数据库的主要拷贝中进行读取的操作在数据库经历重新配置时被重置。这样的操作可以包括为分区创建新的拷贝,或创建分区的副本。

[0026] 在分布式数据存储系统(例如数据库 110)中,当要改变分区的配置时,执行重新配置过程。重新配置过程涉及改变分区的活动配置。作为该过程的一部分,维护在该分区的拷贝之间的通常要被断开的现有复制连接。因此,用户能够在该过程期间在该分区上执行读/写操作。

[0027] 在一些实施例中,可以实现重新配置或最小重新配置的专门情况。最小重新配置可确保用户可以在重新配置过程的持续时间期间在分区上执行读/写操作。在一些情况中,为了执行最小重新配置,要建立写法定数量拷贝。该写法定数量可以被定义为  $(n+1)/2$  的上限,其中  $n$  是在配置中的拷贝总数。当前的主要拷贝将启动并运行,并且在最小重新配置之后将还是作为主要拷贝。在建立这些项之后,启动最小重新配置。

[0028] 在一些实施例中,最小重新配置在下述方面不同于常规的重新配置:1) 维护多个法定数量集,这些法定数量集在重新配置过程期间被动态更新,2) 主要拷贝是多个法定数量集的一部分,3) 最初,所有的次要拷贝要么在第一法定数量集中,要么在法定数量之外,4) 在重新配置结束时,所有的次要拷贝要么在第一法定数量集中,要么在法定数量之外,5) 在重新配置过程持续时间中维护至多两个法定数量拷贝集,6) 在重新配置持续时间期间,根据配置成员来修改法定数量集,以及 7) 要在每个法定数量集上提交用户事务才能将其认为是已提交的。

[0029] 由于上述第 7 条,在重新配置期间的任意点处,在每个法定数量集中要存在可用的写法定数量拷贝。因此,用户可以成功地完成在分区上的写事务。因为至少写法定数量个拷贝是可用的,所以读事务也是可能的(其中,读法定数量是  $(n+1)/2$  的下限,并且写法定数量是  $(n+1)/2$  的上限)。在这个操作期间,如果在前一配置中的可用拷贝的数目下降到其写法定数量之下,则重新配置代理将检测到该情况,终止当前的最小重新配置,并将所述重新配置作为常规的重新配置来重启。

[0030] 考虑到以上描述的系统 and 体系结构,参考图 2 和图 3 的流程图将更好地理解根据



所公开的主题实现的方法。为了解释简明起见,这些方法被示出和描述为一系列框。然而,应该理解和了解,所要求保护的主题不受框的次序的限制,因为一些框可按不同的次序进行和/或与此处所描绘和描述的其他框同时进行。此外,并非全部所示的框都是实现下面所述的方法所必需的。

[0031] 图 2 示出用于在数据库重新配置期间提供数据库访问的方法 200 的流程图。现在将频繁参照环境 100 的组件和数据来描述方法 200。

[0032] 方法 200 包括建立多个法定数量拷贝集来复制给定数据分区的数据的动作,其中所述法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务(动作 210)。例如,法定数量集建立模块 125 将建立法定数量集 A 和 B(126A/126B) 来复制给定数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务 115。

[0033] 在一些情况下,跨最小拷贝集来复制事务(或来自事务的数据)。例如,最小的拷贝集可以包括主要拷贝(例如 127A)和至少一个(或至少两个,等等)次要拷贝(例如 128A1 和 128A2)。每个法定数量拷贝集包括至少一个主要拷贝,并且包括任意数目的次要拷贝。在一些情况下,单个主要拷贝可以是多个法定数量拷贝集的成员。这样,在图 1 中,法定数量集 A 中的主要拷贝 127A 可以是法定数量集 B 中相同的主要拷贝。

[0034] 在一些实施例中,法定数量拷贝集可以是实例化以应答重新配置期间的请求的临时法定数量拷贝集。这样,例如,法定数量集 B(126B) 可以是临时法定数量集,该临时法定数量集被建立以在数据库正被重新配置时对数据请求(例如 106)或其它待决事务进行应答。在一些情况下,在结束重新配置之后移除这样的临时法定数量拷贝集。

[0035] 在重新配置的多个不同的阶段,可以提供对给定数据分区的访问,如图 4 中所示。在图 4 示出的示例中,当次要拷贝 C 加入现有的法定数量集时启动最小重新配置。如在 410 中所示,次要拷贝 C 将加入到包括主要拷贝 A 和次要拷贝 B 的现有法定数量集 AB。如在 415 中所示,法定数量集 1 包括 A 和 B,而法定数量集 2 包括主要拷贝 A。

[0036] 在阶段 1(420),当重新配置开始时,属于前一配置并且属于新配置的所有拷贝的法定数量集成员关系,它们的成员关系被改变以便它们既成为法定数量集 1 又成为法定数量集 2 的一部分,并且仅属于新配置的所有的拷贝被加入成为法定数量集 2 的一部分。这样,在 425,法定数量集 1 具有拷贝 A 和 B,而法定数量集 2 具有拷贝 A、B 和 C。可以启动阶段 1 增额(catch-up)430,其中将加入节点 C 更新为与次要拷贝 B 相同。在 435 中更新法定数量集 2,并且如 440 所示,法定数量集 2 具有主要拷贝 A 和次要拷贝 B 和 C,并且具有足够数目的拷贝和拷贝分布来提交事务。

[0037] 在阶段 2(445)期间,不改变法定数量集,如在 450 中所示。在阶段 3(455)期间,改变属于新配置的所有拷贝的法定数量成员关系,以便它们现在是法定数量集 1 的一部分。而且,改变不属于新配置的所有拷贝的法定数量成员关系,以便它们不再是任意法定数量集的一部分(在法定数量之外)。这样,法定数量集 1 的新配置具有拷贝 A、B 和 C,而法定数量集 2 仅具有主要拷贝 A,如在 460 中所示。在阶段 4(465)期间,发送提交消息,并且,法定数量集 1 是在带有经更新的次要拷贝 C 的情况下完全可操作的。

[0038] 返回图 2,方法 200 包括确定数据分区重新配置已经启动的动作(动作 220)。例如,重新配置模块 120 可以确定已经为法定数量集 A(126A)启动数据分区重新配置。可以

由法定数量拷贝集的拷贝（例如 128A2）离开或加入拷贝集来启动重新配置。在重新配置期间，加入拷贝所加入的法定数量拷贝集被修改为包括该加入的拷贝。类似地，当拷贝离开法定数量集时，重新配置该法定数量集。当拷贝离开法定数量集时，重新配置模块 120 可以防止现有的数据库复制连接由于拷贝的离开而被断开。这样，如果次要拷贝 128A2 将要离开法定数量集 A (126A)，则在数据库和主要拷贝 127A 和次要拷贝 128A1 之间的现有数据库复制连接将不被断开。

[0039] 方法 200 包括在数据分区的重新配置期间使用法定数量拷贝集中的每一个中的至少法定数量个拷贝来提供对数据分区的数据的访问。例如，数据库 110 可以在数据分区的重新配置期间使用法定数量集 A 的主要拷贝 127A 和次要拷贝 128A1 提供对给定分区的数据的访问。在一些情况中，可由法定数量拷贝集中的大多数拷贝（图 1 的法定数量集的三个拷贝中的两个拷贝）来对数据库事务进行确认。当法定数量成员在重新配置的不同阶段期间被移动到不同的法定数量集时，可以在事务上一致的方式来维护分区上的数据。这样，无论有多少拷贝被改变或如何改变拷贝，都将以在事务上一致的方式来提供对底层数据的访问。这就确保了在任意事务中没有数据丢失。而且，可以防止依赖于从数据库的主要拷贝中进行读取的操作在重新配置过程期间被重置。

[0040] 现在转向图 3，图 3 示出用于在数据库重新配置期间维护复制连接的方法 300 的流程图。现在将频繁参照环境 100 的组件和数据来描述方法 300。

[0041] 方法 300 包括建立多个法定数量拷贝集来复制给定数据分区的数据的动作，其中所述法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务（动作 310）。例如，法定数量集建立模块 125 可建立法定数量集 A 和 B (126A/126B) 来复制给定数据分区的数据。法定数量拷贝集确保了至少最小数目的拷贝可用于在分区重新配置期间提交待决事务 115。

[0042] 方法 300 包括确定拷贝的离开启动了数据分区重新配置的动作（动作 320）。例如，重新配置模块 120 可以确定已经为法定数量集 A (126A) 启动了数据分区重新配置。可以由例如次要拷贝加入或离开法定数量集 A 来启动重新配置。方法 300 还包括防止现有的数据库复制连接在拷贝离开时被断开的动作（动作 330）。

[0043] 例如，重新配置模块 120 可以防止任意现有的到法定数量集 A 的其它拷贝的数据库复制连接（例如到主要拷贝 127A 或次要拷贝 128A1 的连接）被断开或移除。这样，使得没有改变的拷贝的数据库复制连接保持完好。随后，保持完好的拷贝在重新配置期间能够继续处理事务。对于移除（即离开法定数量集）的拷贝来说，可以以拷贝的分区在重新配置期间保持在事务上一致的方式来移除正离开的拷贝。这样，被处理的任意事务将是一致的，并将提供数据库用户所期望的事务性保证。

[0044] 方法 300 还包括在数据分区的重新配置期间使用法定数量拷贝集中的每一个中的至少法定数量个拷贝来提供对数据分区的数据的访问（动作 340）。例如，法定数量集 A (126A) 可以在数据库的数据分区重新配置期间提供对该分区的访问。法定数量集可以使用主要拷贝 127A 和次要拷贝 (128A1/128A2) 中的至少一个来提供这样的访问。在一些实施例中，在重新配置期间可以防止各种应用程序被重置。特别地，可以防止依赖于从数据库的主要拷贝中读取的应用程序被重置。这样，至少在一些情况中，在重新配置期间可以防止分区复制操作被重置。另外地，或替换地，在重新配置期间可以防止新的拷贝创建操作被重

置。

[0045] 这样,提供了在数据库重新配置期间提供数据库访问的系统、方法和计算机程序产品。在重新配置期间可以以在事务上一致的方式来继续处理事务。这样,提供了在数据库重新配置期间维护复制连接的系统、方法和计算机程序产品。这样,没有作为重新配置一部分来改变的拷贝可以维护它们的复制连接,并且能在重新配置期间继续提供数据库访问。

[0046] 本发明可具体化为其它具体形式而不背离其精神或本质特征。所描述的实施例在所有方面都应被认为仅是说明性而非限制性的。因此,本发明的范围由所附权利要求书而非前述描述指示。落入权利要求书的等效方案的含义和范围内的所有改变被权利要求书的范围所涵盖。

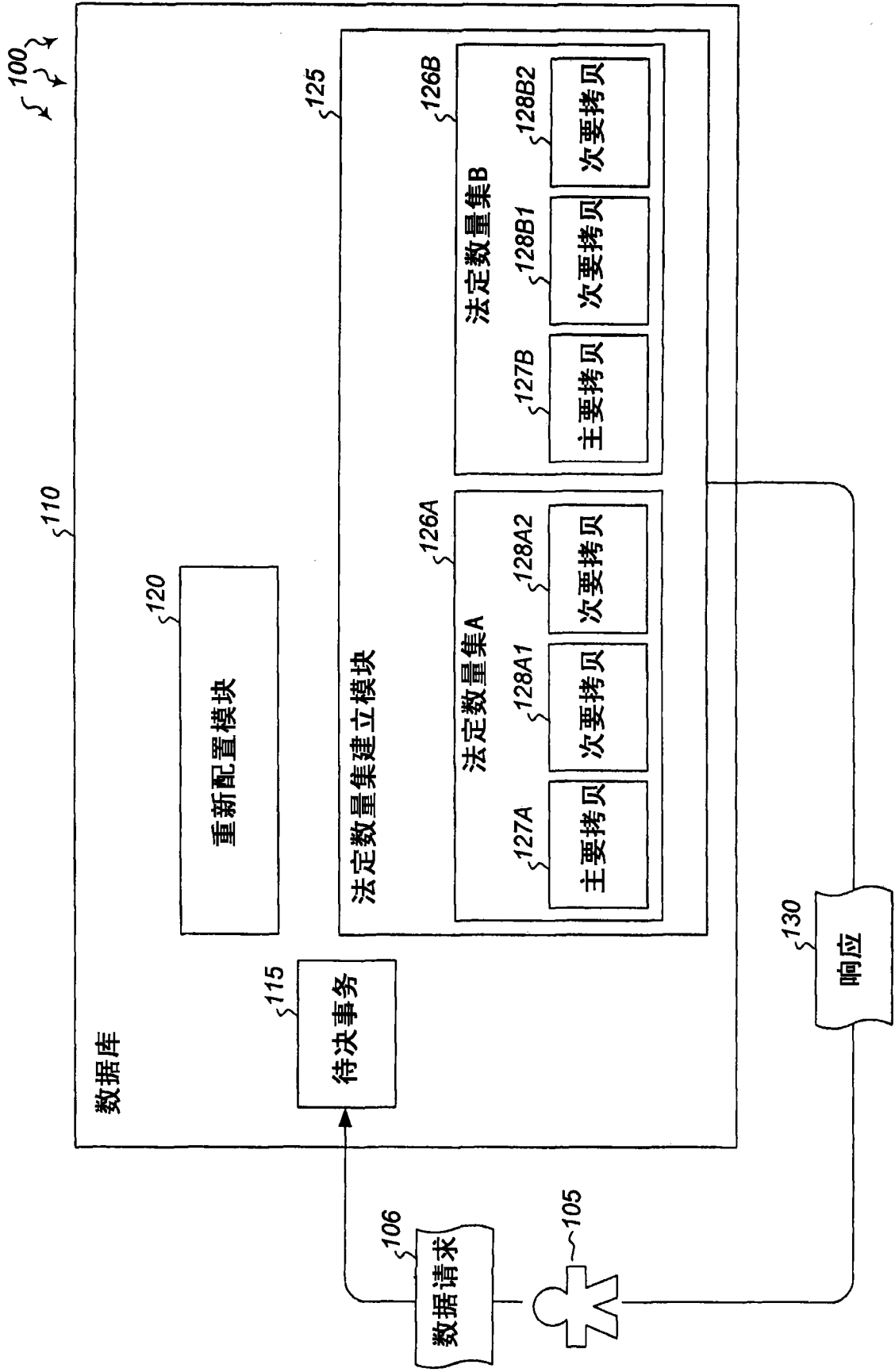


图 1

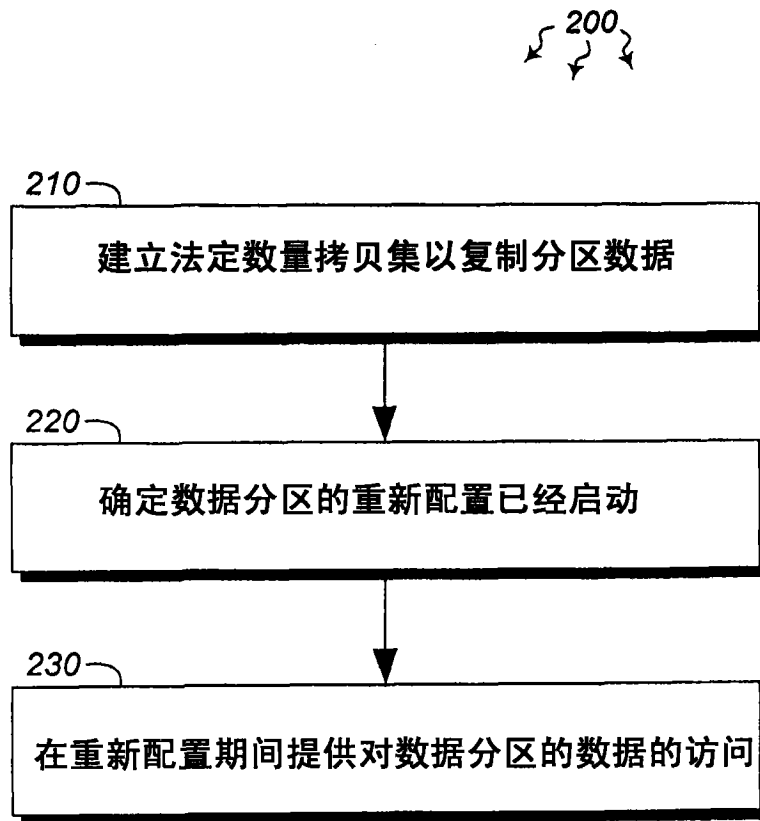


图 2

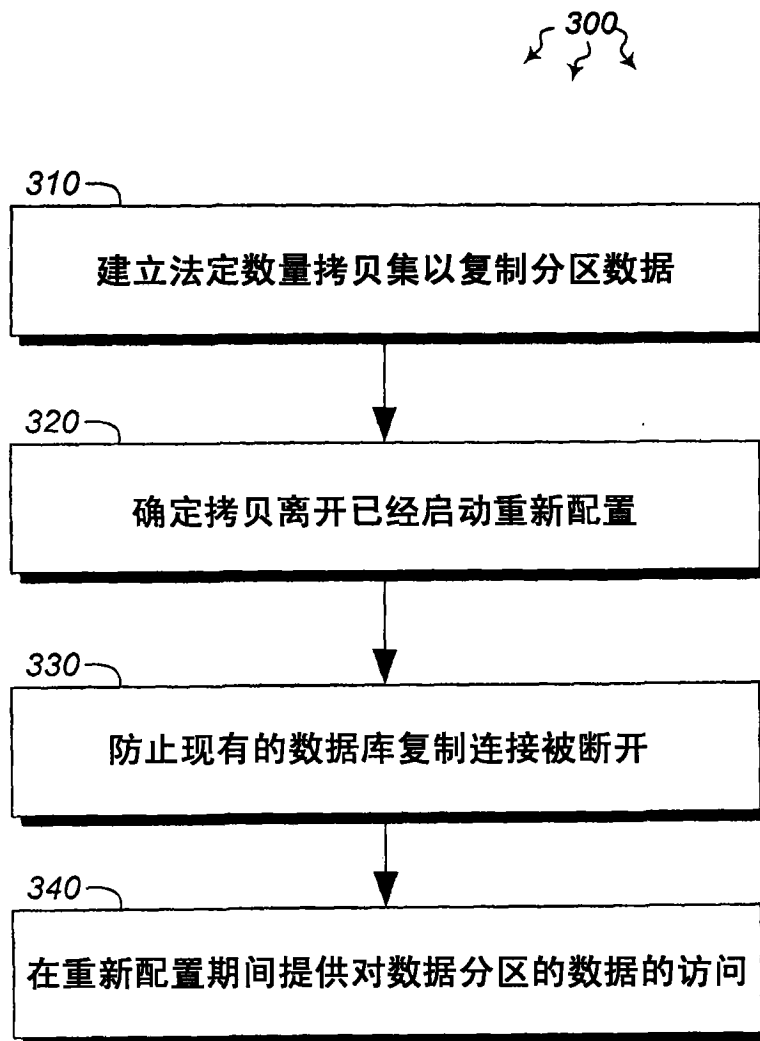


图 3

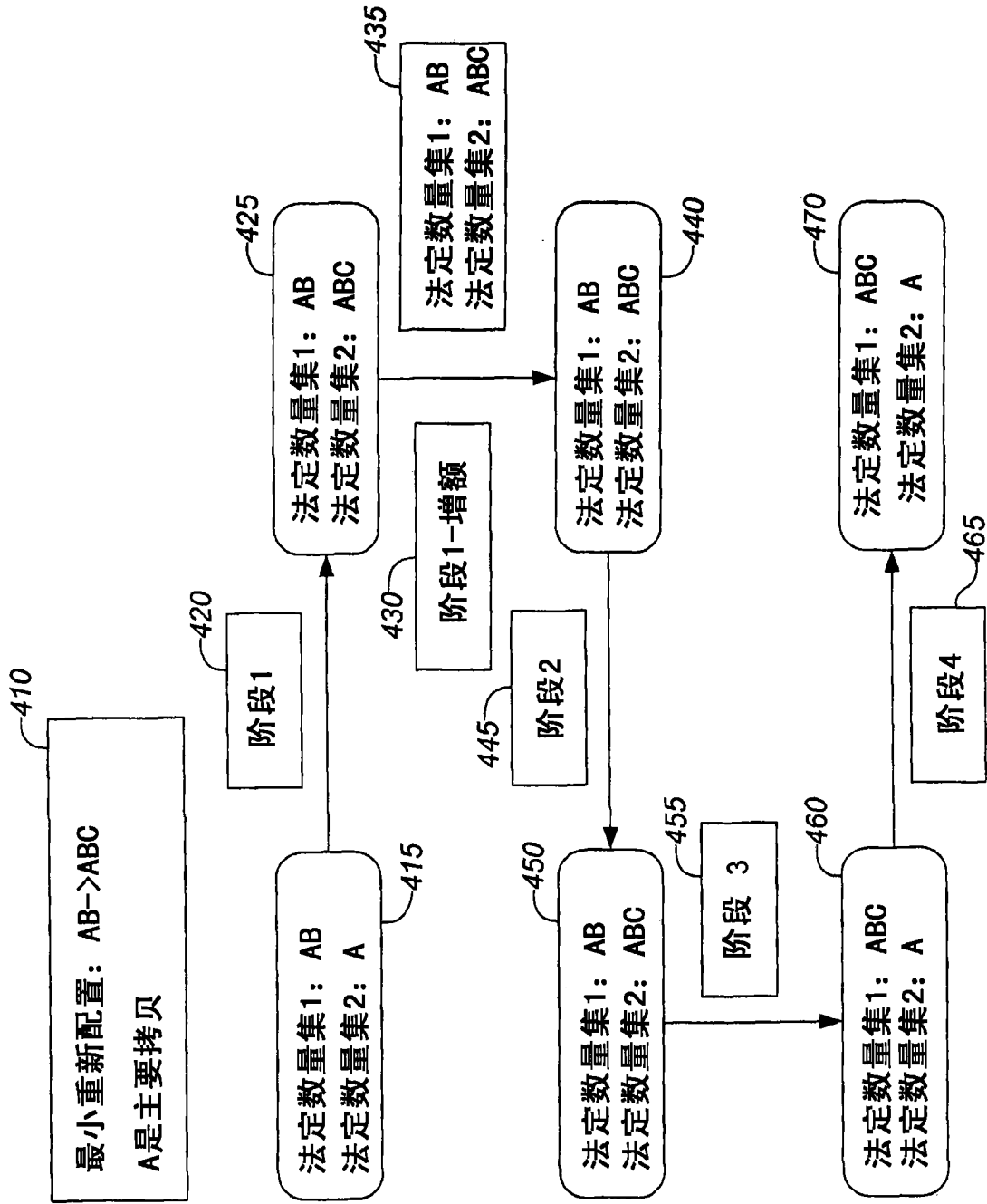


图 4