



(12) 发明专利

(10) 授权公告号 CN 113297230 B

(45) 授权公告日 2024.03.08

(21) 申请号 202010732285.3

(22) 申请日 2020.07.27

(65) 同一申请的已公布的文献号  
申请公布号 CN 113297230 A

(43) 申请公布日 2021.08.24

(73) 专利权人 阿里巴巴集团控股有限公司  
地址 英属开曼群岛大开曼资本大厦一座四  
层847号邮箱

(72) 发明人 姚灵灵 杨泽 郭琦 庞枢

(74) 专利代理机构 北京智信禾专利代理有限公  
司 11637  
专利代理师 李晓庆

(51) Int. Cl.  
G06F 16/23 (2019.01)  
G06F 16/27 (2019.01)

(56) 对比文件

CN 106610876 A, 2017.05.03

CN 107391628 A, 2017.11.24

CN 106815097 A, 2017.06.09

CN 110471909 A, 2019.11.19

CN 105989139 A, 2016.10.05

US 2017371914 A1, 2017.12.28

CN 111125060 A, 2020.05.08

张燕超. 支持有效时间的时态RDF数据的不  
一致性分析. 计算技术与自动化. 2018, (03), 全  
文.

审查员 曹宁

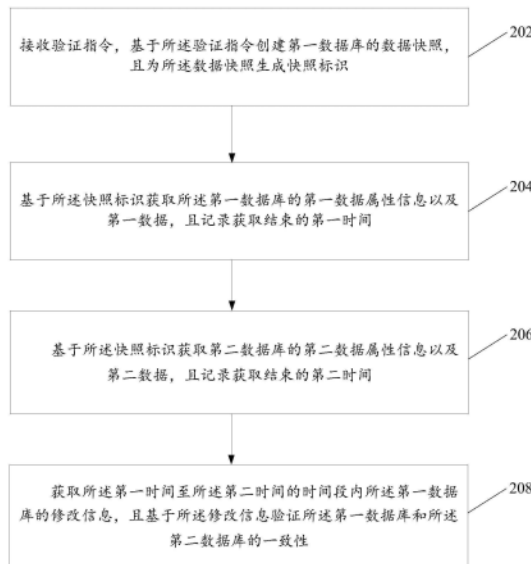
权利要求书3页 说明书13页 附图4页

(54) 发明名称

数据验证方法及装置

(57) 摘要

本说明书实施例提供一种数据验证方法及装置,其中,所述数据验证方法包括接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识;基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;获取所述第一时间至所述第二时间的时段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。



1. 一种数据验证方法,包括:

接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识,其中,所述快照标识用于获取对应的所述数据快照,所述数据快照用于表示数据库中的可查看数据,从而获取所述数据库中可查看数据的属性信息和数据;

基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;

基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;

获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。

2. 根据权利要求1所述的数据验证方法,所述获取所述第一时间至所述第二时间的时段内所述第一数据库的修改信息包括:

获取所述第一时间至所述第二时间的时段内所述第一数据库的日志记录,并根据所述日志记录确定所述第一数据库的修改信息。

3. 根据权利要求1或2所述的数据验证方法,所述基于所述修改信息验证所述第一数据库和所述第二数据库的一致性包括:

基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性,以及

所述第一数据库的第一数据与所述第二数据库的第二数据的一致性。

4. 根据权利要求3所述的数据验证方法,所述基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性包括:

在基于所述修改信息确定所述第一数据库的第一数据属性信息存在增加的情况下,删除增加的第一数据属性信息;

将删除后的所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息进行匹配;

若匹配成功,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息一致;

若匹配失败,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息不一致。

5. 根据权利要求3所述的数据验证方法,所述基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性包括:

在基于所述修改信息确定所述第一数据库的第一数据属性信息存在删除的情况下,添加删除的第一数据属性信息;

将添加后的所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息进行匹配;

若匹配成功,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息一致;

若匹配失败,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息不一致。

6. 根据权利要求3所述的数据验证方法,所述基于所述修改信息验证所述第一数据库的第一数据与所述第二数据库的第二数据的一致性包括:

在基于所述修改信息确定所述第一数据库的第一数据存在增加的情况下,删除增加的第一数据;

将删除后的所述第一数据库的第一数据与所述第二数据库的第二数据进行匹配;

若匹配成功,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据一致;

若匹配失败,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据不一致。

7. 根据权利要求3所述的数据验证方法,所述基于所述修改信息验证所述第一数据库的第一数据与所述第二数据库的第二数据的一致性包括:

在基于所述修改信息确定所述第一数据库的第一数据存在删除的情况下,添加删除的第一数据;

将添加后的所述第一数据库的第一数据与所述第二数据库的第二数据进行匹配;

若匹配成功,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据一致;

若匹配失败,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据不一致。

8. 根据权利要求1所述的数据验证方法,所述基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据包括:

基于所述快照标识确定所述数据快照,且根据所述数据快照获取并缓存所述第一数据库的第一数据属性信息以及第一数据。

9. 根据权利要求8所述的数据验证方法,所述基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据包括:

基于所述快照标识确定所述数据快照,且根据所述数据快照获取并缓存所述第二数据库的第二数据属性信息以及第二数据。

10. 根据权利要求1所述的数据验证方法,所述接收验证指令包括:

按照预设的时间间隔接收验证指令。

11. 一种数据验证装置,包括:

快照创建模块,被配置为接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识,其中,所述快照标识用于获取对应的所述数据快照,所述数据快照用于表示数据库中的可查看数据,从而获取所述数据库中可查看数据的属性信息和数据;

第一获取模块,被配置为基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;

第二获取模块,被配置为基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;

验证模块,被配置为获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。

12. 一种计算设备,包括:

存储器和处理器;

所述存储器用于存储计算机可执行指令,所述处理器用于执行所述计算机可执行指令,其中,所述处理器执行所述计算机可执行指令时实现权利要求1至10任意一项所述数据验证方法的步骤。

13. 一种计算机可读存储介质,其存储有计算机指令,该指令被处理器执行时实现权利要求1至10任意一项所述数据验证方法的步骤。

## 数据验证方法及装置

### 技术领域

[0001] 本说明书实施例涉及数据处理技术领域,特别涉及一种数据验证方法。本说明书一个或者多个实施例同时涉及一种数据验证装置,一种计算设备,以及一种计算机可读存储介质。

### 背景技术

[0002] 目前有很多针对数据库数据正确性验证的方式,但是分布式数据库存在“主备延迟”,很难确定备库同步结束的时间,并同时实时数据库对象及表数据一致性验证,因此针对分布式数据库的实时一致性验证,一直没有切实可行的方案。

[0003] 因此需要提供一种可以对分布式数据库的主库和备库进行实时数据库对象以及表数据进行一致性验证的数据验证方法。

### 发明内容

[0004] 有鉴于此,本说明书实施例提供了一种数据验证方法。本说明书一个或者多个实施例同时涉及一种数据验证装置,一种计算设备,以及一种计算机可读存储介质,以解决现有技术中存在的技术缺陷。

[0005] 根据本说明书实施例的第一方面,提供了一种数据验证方法,包括:

[0006] 接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识;

[0007] 基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;

[0008] 基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;

[0009] 获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。

[0010] 根据本说明书实施例的第二方面,提供了一种数据验证装置,包括:

[0011] 快照创建模块,被配置为接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识;

[0012] 第一获取模块,被配置为基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;

[0013] 第二获取模块,被配置为基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;

[0014] 验证模块,被配置为获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。

[0015] 根据本说明书实施例的第三方面,提供了一种计算设备,包括:

[0016] 存储器和处理器;

[0017] 所述存储器用于存储计算机可执行指令,所述处理器用于执行所述计算机可执行指令,其中,所述处理器执行所述计算机可执行指令时实现所述数据验证方法的步骤。

[0018] 根据本说明书实施例的第四方面,提供了一种计算机可读存储介质,其存储有计算机可执行指令,该指令被处理器执行时实现所述数据验证方法的步骤。

[0019] 本说明书一个实施例实现了一种数据验证方法及装置,其中,所述数据验证方法包括接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识;基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性;

[0020] 所述数据验证方法可以通过第一数据库的审计日志对基于同一数据快照获取的第一数据库和第二数据库的数据属性信息以及具体数据进行修正,实现实时对第一数据库和第二数据库的一致性验证。

#### 附图说明

[0021] 图1是本说明书一个实施例提供的一种数据验证装置的架构图;

[0022] 图2是本说明书一个实施例提供的一种数据验证方法的流程图;

[0023] 图3是本说明书一个实施例提供的一种数据验证装置的结构示意图;

[0024] 图4是本说明书一个实施例提供的一种计算设备的结构框图。

#### 具体实施方式

[0025] 在下面的描述中阐述了很多具体细节以便于充分理解本说明书。但是本说明书能够以很多不同于在此描述的其它方式来实施,本领域技术人员可以在不违背本说明书内涵的情况下做类似推广,因此本说明书不受下面公开的具体实施的限制。

[0026] 在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。还应当理解,本说明书一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0027] 应当理解,尽管在本说明书一个或多个实施例中可能采用术语第一、第二等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本说明书一个或多个实施例范围的情况下,第一也可以被称为第二,类似地,第二也可以被称为第一。取决于语境,如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0028] 首先,对本说明书一个或多个实施例涉及的名词术语进行解释。

[0029] 云数据库:是指被优化或部署到一个虚拟计算环境中的数据库,可以实现按需付费、按需扩展、高可用性以及存储整合等优势。

[0030] GRV:英文全称:Global Read View,中文全称:全局读取视图,本说明书实施例中

可以理解为数据快照。

[0031] DDL:英文全称:Data Definition Language,中文全称:数据库模式定义语言,是用于描述数据库中要存储的现实世界实体的语言。

[0032] Redo:英文全称:Redo log,中文全称:重做日志。

[0033] 在本说明书中,提供了一种数据验证方法。本说明书一个或者多个实施例同时涉及一种数据验证装置,一种计算设备,以及一种计算机可读存储介质,在下面的实施例中逐一进行详细说明。

[0034] 参见图1,图1示出了根据本说明书一个实施例提供的一种数据验证装置的架构图,包括:

[0035] 分布式数据系统102、主库104,备库106,其中,主库104为主数据库,备库106可以理解为分布式数据库系统102的主库104的多个备份数据库。

[0036] 具体实施时,分布式数据系统102可以实时接收针对主库104和备库106的一致性验证指令,而在接收到验证指令后,分布式数据系统102对主库104创建数据快照,并为该数据快照生成唯一的快照标识;

[0037] 然后根据与所述快照标识对应的数据快照中的内容以及主库104的元数据去主库104中查询主库104的数据库对象所有信息以及主库104的具体数据,其中,数据库对象所有信息包括但不限于表、列、主键、外键、视图以及索引等信息,主库104的具体数据为主库的数据库表中存放的具体数据值,在查询到主库104的数据库对象所有信息以及主库104的具体数据后,将其保存并记录执行查询结束时间点为T1;

[0038] 再根据与查询主库104时相同的数据快照中的内容以及备库106的元数据,去备库106中查询备库106的数据库对象所有信息以及备库106的具体数据,将其保存并记录执行查询结束时间点为T2;

[0039] 最后通过主库104的Redo或审计日志获取主库104和备库106在T1至T2的时间段之间的数据库对象修改信息,将保存的主库104的数据库对象所有信息、具体数据结合T1至T2的时间段之间的数据库对象修改信息,与保存的备库106的数据库对象所有信息、具体数据进行分别比对,判断是否一致,若一致,则说明同一时间主备库的数据库对象所有信息以及具体数据一致;若不一致,则说明同一时间主备库的数据库对象所有信息以及具体数据不一致,可能存在主库104向备库106写入数据错误导致主库104和备库106同一时间点存储的数据不一致,又或者主库104的数据库对象信息修改(例如增加或删除列等)与备库106同步错误,导致主库104和备库106同一时间点的数据库对象信息不一致,此时可以发出错误提示,提醒维护人员对分布式系统102进行调整,以保证主备库实时数据的一致性。

[0040] 本说明书实施例提供的所述数据验证装置实现了任意时间查询主库和多个备库数据完全一致功能,利用该功能能够做到主库执行高并发任意读写操作下,可以实时验证主库和多个备库数据的一致性。

[0041] 下面对数据验证方法的具体处理过程进行详细说明,具体实施时,所述数据验证方法应用于分布式数据库中,例如云数据库。

[0042] 参见图2,图2示出了根据本说明书一个实施例提供的一种数据验证方法的流程图,包括以下步骤:

[0043] 步骤202:接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所

述数据快照生成快照标识。

[0044] 其中,第一数据库对应上述实施例的主库104,第二数据库对应上述实施例的备库106;验证指令可以理解为验证同一时间第一数据库和第二数据库一致性的指令,也可以理解为验证同一时间第一数据库中的某个数据表与对应的第二数据库中的数据表一致性的指令。

[0045] 具体的,在接收验证指令后,基于该验证指令为第一数据库创建在接收到该验证指令时刻的第一数据库的数据快照,例如接收验证指令为12:00,则在12:00为第一数据库创建数据快照,并为该12:00为第一数据库创建的数据快照生成唯一的快照标识,后续可以通过唯一的快照标识关联查找到该对应的数据快照。

[0046] 而在12:00时刻创建第一数据库的数据快照后,通过该数据快照可以看出在12:00这个时间点,第一数据库的数据表中的哪些行已经操作完成,哪些行可以看到,例如在12:00时刻创建的第一数据库的数据快照中可以看到前两行的数据,其他行的数据均无法看到。

[0047] 例如,验证指令也可以为基于SQL语句的验证指令,例如验证指令为查询数据库所有表的SQL语句:SELECT TABLE\_NAME FROM information\_schema.TABLES WHERE table\_schema='数据库名'。

[0048] 具体实施时,所述数据验证方法可以作为一个应用实现,那么在具体应用场景中,只要调用了该应用,就可以实现所述数据验证方法的处理流程,进行第一数据库和第二数据库的一致性验证,因此具体实施时,为了可以及时的验证分布式数据库系统中主库和备库的一致性,可以每隔一段时间就主动触发对分布式数据库系统中主库和备库的一致性验证流程,具体实现方式如下所述:

[0049] 所述接收验证指令包括:

[0050] 按照预设的时间间隔接收验证指令。

[0051] 其中,预设的时间间隔可以根据实际应用进行设置,例如设置为2秒或3秒等;过短的时间间隔会造成验证较为频繁,会过多的占用网络资源,造成系统负担,而过长的时间间隔可能会造成主备库数据一致性验证的不及时性,可能会错过主备库存储数据不一致的问题的发现。

[0052] 步骤204:基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间。

[0053] 其中,第一数据属性信息包括但不限于第一数据库的表、列、主键、外键、视图以及索引等信息,例如第一数据库里面有多少表、每个表有多少列、主键、外键以及索引是什么等信息;第一数据则可以理解为第一数据库中存放的具体数据值,例如第一数据库中存放的是学生档案信息,则第一数据可以理解为学生姓名、证件号码、联系方式以及家庭住址等具体数据。

[0054] 具体的,为了保证可以获取到基于同一版本的数据快照的第一数据库和第二数据库的数据属性信息和具体数据,则通过同一个快照标识进行第一数据库的第一数据属性信息以及第一数据,和第二数据库的第二数据属性信息以及第二数据的获取,具体实现方式如下所述:

[0055] 所述基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据



包括:

[0056] 基于所述快照标识确定所述数据快照,且根据所述数据快照获取并缓存所述第一数据库的第一数据属性信息以及第一数据。

[0057] 其中,每个数据快照对应一个唯一的快照标识;实际应用中,在进行第一数据库的第一数据属性信息以及第一数据获取时,首先基于快照标识确定对应的数据快照,然后基于该数据快照获取当前第一数据库的第一数据属性信息以及第一数据。

[0058] 例如数据快照中表示,可以查看第一数据库的前两行数据,那么基于该数据快照获取当前第一数据库的第一数据属性信息以及第一数据,则是基于该数据快照获取当前第一数据库的前两行数据的第一数据属性信息以及第一数据,例如基于该数据快照获取当前第一数据库的前两行数据的第一数据属性信息为数据表的表名a、列名c等表定义信息,该数据表a中包括3列:c1、c2、c3,第一数据为c1列对应的第一行的第一数据为1.1、第二行的第一数据为1.2,c2列对应的第一行的第一数据为2.1、第二行的第一数据为2.2,c3列对应的第一行的第一数据为3.1、第二行的第一数据为3.2,即参见表1:

[0059]

C1	C2	C3
1.1	2.1	3.1
1.2	2.2	3.2

[0060] 表1,即根据数据快照获取并缓存的第一数据库的第一数据属性信息以及第一数据参见表1,且在获取第一数据库的第一数据属性信息以及第一数据结束后记录获取结束的第一时间,例如基于数据快照获取第一数据库的数据的结束时间为12:05。

[0061] 步骤206:基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间。

[0062] 具体的,所述基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据包括:

[0063] 基于所述快照标识确定所述数据快照,且根据所述数据快照获取并缓存所述第二数据库的第二数据属性信息以及第二数据。

[0064] 其中,此处的快照标识即在获取第一数据库的第一数据属性信息以及第一数据时依据的快照标识。

[0065] 实际应用中,基于与获取第一数据库的第一数据属性信息以及第一数据时相同的快照标识,查找对应的、与获取第一数据库的第一数据属性信息以及第一数据时相同的数据快照,然后基于该数据快照获取当前第二数据库的第二数据属性信息以及第二数据。

[0066] 仍以上述数据快照为可以查看第一数据库的前两行数据为例,那么基于该数据快照获取当前第二数据库的第二数据属性信息以及第二数据,则是基于该数据快照获取当前第二数据库的前两行数据的第二数据属性信息以及第二数据;例如基于该数据快照获取当前第二数据库的前两行数据的第二数据属性信息为数据表的表名a、列名c等表定义信息,该数据表a中包括2列:c1、c2,第二数据为c1列对应的第一行的第一数据为1.1、第二行的第一数据为1.2,c2列对应的第一行的第一数据为2.1、第二行的第一数据为2.2,即参见表2:

[0067]

C1	C2
1.1	2.1
1.2	2.2

[0068] 表2,即根据数据快照获取并缓存的第二数据库的第二数据属性信息以及第二数据参见表2,且在获取第二数据库的第二数据属性信息以及第二数据结束后记录获取结束的第二时间,例如基于该数据快照获取第二数据库的数据的结束时间为12:08。

[0069] 本说明书实施例提供的所述数据验证方法根据同一个版本的数据快照获取第一数据库的第一数据属性信息以及第一数据,和获取第二数据库的第二数据属性信息以及第二数据,以保证获取的第一数据库和第二数据库的数据属性信息以及具体数据为同一时间点的内容,实现对同一时间点的第一数据库和第二数据库的一致性验证的精确性。

[0070] 步骤208:获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。

[0071] 具体的,所述获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息包括:

[0072] 获取所述第一时间至所述第二时间的时间段内所述第一数据库的日志记录,并根据所述日志记录确定所述第一数据库的修改信息。

[0073] 其中,第一数据库的日志记录为第一数据库的Redo或审计日志中的日志记录。

[0074] 以上述第一时间为12:05,第二时间为12:08为例,获取第一时间至第二时间的时间段内第一数据库的日志记录,则为获取12:05到12:08之间的时间段内第一数据库的日志记录,该日志记录中记载了12:05到12:08的时间段内第一数据库的增删改查等任何操作,例如在第一数据库中增加或删除列以及在第一数据库中对哪个数据在什么时间进行了修改等。

[0075] 具体的,获取第一时间至第二时间的时间段内第一数据库的日志记录,然后根据该日志记录中的内容查询出第一数据库在该时间段内的修改信息。

[0076] 本说明书实施例提供的所述数据验证方法,通过Redo或审计日志可以对第一时间至第二时间的时间段内第一数据库的修改信息进行准确以及快速的获取,保证了为后续通过该修改信息对第一数据库和第二数据库进行一致性验证的准确性。

[0077] 而具体实施时,所述基于所述修改信息验证所述第一数据库和所述第二数据库的一致性包括:

[0078] 基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性,以及

[0079] 所述第一数据库的第一数据与所述第二数据库的第二数据的一致性。

[0080] 具体的,基于修改信息验证第一数据库和第二数据库的一致性,则是基于修改信息验证同一时间点第一数据库的表结构以及具体存储的数据是否与第二数据库的表结构以及具体存储的数据一致,以确认主备库的表定义信息以及具体存储的数据的一致性,后续应用时可以直接从备库进行准确的数据读取,以减轻主库的负担,其中,表结构以及表定义信息即为第一数据属性信息,即数据库的表名、列名、主键、外键、视图以及索引等信息。

[0081] 仍以上述表1和表2为例,基于数据快照获取第一数据库的数据的结束时间为12:05,获取的第一数据库的第一数据属性信息为表名a、列名c,该数据表a中包括3列:c1、c2、c3;第一数据为c1列对应的第一行的第一数据为1.1、第二行的第一数据为1.2,c2列对应的第一行的第一数据为2.1、第二行的第一数据为2.2,c3列对应的第一行的第一数据为3.1、第二行的第一数据为3.2;

[0082] 在12:07的时候在第一数据库做了删除列的操作,即删除了c3列,这个时候基于与第一数据库同样的数据快照在第二数据库会获取到2列的数据,即基于数据快照获取第二数据库的数据的结束时间为12:08,获取当前第二数据库的前两行数据的第二数据属性信息为表名a、列名c,该数据表a中包括2列:c1、c2,第二数据为c1列对应的第一行的第一数据为1.1、第二行的第一数据为1.2,c2列对应的第一行的第一数据为2.1、第二行的第一数据为2.2;此时获取12:05至12:08的时间段内第一数据库的修改信息则为:删除了第一数据库的列c3,以及c3对应的具体数据3.1、3.2。

[0083] 此时,则可以基于该修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性,以及

[0084] 所述第一数据库的第一数据与所述第二数据库的第二数据的一致性。

[0085] 具体验证方式如下所述:

[0086] 所述基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性包括:

[0087] 在基于所述修改信息确定所述第一数据库的第一数据属性信息存在增加的情况下,删除增加的第一数据属性信息;

[0088] 将删除后的所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息进行匹配;

[0089] 若匹配成功,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息一致;

[0090] 若匹配失败,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息不一致。

[0091] 例如,基于修改信息确定第一数据库的列存在增加情况下,则删除该第一数据库中增加的该列,然后将删除该列后的第一数据库的列与第二数据库的列进行匹配,若列的数量相等,则确定同一时间点第一数据库的列的数量与第二数据库的列的数量相等;若列的数量不相等,则确定同一时间点第一数据库的列的数量与第二数据库的列的数量不相等,第二数据库可能存在表结构更新延迟,此时则可以同时维护人员进行调整,以保证主备库数据表结构的一致性,避免后续数据备份错误。

[0092] 而本说明书另一实施例中,所述基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性包括:

[0093] 在基于所述修改信息确定所述第一数据库的第一数据属性信息存在删除的情况下,添加删除的第一数据属性信息;

[0094] 将添加后的所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息进行匹配;

[0095] 若匹配成功,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息一致;

[0096] 若匹配失败,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息不一致。

[0097] 例如,基于修改信息确定第一数据库的列存在删除的情况下,则为第一数据库增加删除的该列,然后将增加该列后的第一数据库的列与第二数据库的列进行匹配,若列的

数量相等,则确定同一时间点第一数据库的列的数量与第二数据库的列的数量相等;若列的数量不相等,则确定同一时间点第一数据库的列的数量与第二数据库的列的数量不相等,第二数据库可能存在表结构更新延迟,此时则可以同时维护人员进行调整,以保证主备库数据表结构的一致性,避免后续数据备份错误。

[0098] 本说明书实施例提供的所述数据验证方法,在对数据库的数据属性信息的一致性验证时,首先通过元数据查询数据库对象所有信息(如表、列、主键、外键、视图、索引等数据属性信息)将其保存,并记录执行查询时间点t1,然后在在备库中执行主库相同查询信息,记录查询时间点t2,再通过Redo或审计日志等获取渠道获取主库和备库t1至t2时间段之间的数据库对象修改信息,最后验证主库数据库对象所有信息结合t1至t2时间段的主库修改信息与备库数据库对象所有信息是否一致,如果不一致则为主备数据不一致的错误,此时就需要对分布式数据库系统进行及时调整,避免从备库调用的数据出现应用问题。

[0099] 本说明书另一实施例中,所述基于所述修改信息验证所述第一数据库的第一数据与所述第二数据库的第二数据的一致性包括:

[0100] 在基于所述修改信息确定所述第一数据库的第一数据存在增加的情况下,删除增加的第一数据;

[0101] 将删除后的所述第一数据库的第一数据与所述第二数据库的第二数据进行匹配;

[0102] 若匹配成功,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据一致;

[0103] 若匹配失败,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据不一致。

[0104] 例如,基于修改信息确定第一数据库的列以及该列对应的数据存在增加情况下,则删除该第一数据库中增加的该列以及该列对应的数据,然后将删除该列以及该列数据后的第一数据库的第一数据与第二数据库的第二数据进行匹配,若第一数据和第二数据相同,则确定同一时间点第一数据库的第一数据与第二数据库的第二数据相同;若第一数据和第二数据不相同,则确定同一时间点第一数据库的第一数据与第二数据库的第二数据不相同,可以第二数据库可能存在数据备份延迟等情况,此时则可以同时维护人员进行调整,以保证主备库数据的一致性。

[0105] 本说明书另一实施例中,所述基于所述修改信息验证所述第一数据库的第一数据与所述第二数据库的第二数据的一致性包括:

[0106] 在基于所述修改信息确定所述第一数据库的第一数据存在删除的情况下,添加删除的第一数据;

[0107] 将添加后的所述第一数据库的第一数据与所述第二数据库的第二数据进行匹配;

[0108] 若匹配成功,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据一致;

[0109] 若匹配失败,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据不一致。

[0110] 例如,基于修改信息确定第一数据库的列以及该列对应的数据存在删除情况下,则增加该第一数据库中删除的该列以及该列对应的数据,然后将增加该列以及该列数据后的第一数据库的第一数据与第二数据库的第二数据进行匹配,若第一数据和第二数据相

同,则确定同一时间点第一数据库的第一数据与第二数据库的第二数据相同;若第一数据和第二数据不相同,则确定同一时间点第一数据库的第一数据与第二数据库的第二数据不相同,可以第二数据库可能存在数据备份延迟等情况,此时则可以同时维护人员进行调整,以保证主备库数据的一致性。

[0111] 本说明书实施例提供的所述数据验证方法,在接收到验证指令后,通过分布式数据库的GRV功能,为第一数据库创建数据快照,通过GRV,可以看到GRV创建时点应该看到的数据,而不会读到在GRV创建后提交的数据,GRV可以在主库和多个备库中间共享,可以通过相同的唯一id来访问,基于上述功能主库和多个备库可以通过该唯一的GRV的id查询数据库中所有表和数据,同时基于GRV的查询可以并发执行,首先在主库使用GRV获取唯一gid(即GRV的id),在主库中使用该gid对应的GRV查询库中表dd1和具体数据,并将结果缓存,以及记录主库查询时间t1;在各备库中使用相同gid对应的GRV查询备库中表dd1和数据将结果缓存,并记录各备库查询时间t2,然后获取主库和备库查询时间段t1至t2之间数据库表修改信息,获取渠道可以为Redo或审计日志等,通过ddl修改记录修正主库和备库返回数据,验证修正后的主库和备考返回数据是否完全一致,如果不一致,则为存在主库和备库数据不一致的错误,此时就需要对分布式数据库系统进行及时调整,避免从备库调用的数据出现应用问题。

[0112] 举例说明:例如,主库数据库db1包含tab1表,tab1在t1时间为两列col1、col2,在t1到t2之间用户操作表结构删除col2列剩下col1列。

[0113] 首先,在主库使用GRV获取唯一gid,主库查询表结束时间为t1,保存tab1的ddl信息,使用gid查询结果为两列数据10条记录;备库查询表结束时间为t2,保存tab1的ddl信息,使用gid查询结果为一列数据10条记录,从主库的Redo或审计日志里面找t1和t2之间的ddl记录,发现删除col2列,此时,则可以删除主库col2查询结果,只保留col1列查询结果,然后将该删除了col2查询结果的主库与备库进行一致性比对,则可以确定主备库的ddl信息以及存储数据均一致,则结束验证流程。

[0114] 本说明书实施例提供的所述数据验证方法,实时验证了主库和备库的数据库对象定义的一致性及解决主和备库执行期间存在表ddl修改导致无法验证的情况,首先在主库和备库对应时间点通过系统元数据查询数据库对象所有信息(如表、列、主键、外键、视图、索引等),以及基于分布式数据系统提供的GRV功能查到的主库、备库对应时间点的具体数据,然后通过Redo或审计日志根据算法对查到的数据库对象所有信息以及具体数据进行校验以及修正;本说明书实施例中,主库可以执行任意的种类的压力,并且可以并发执行,且验证一致性用的query语句,在主库、备库上都支持并发执行,实现了可以在任意时间查询主库和多个备库数据完全一致的功能,利用该功能能够做到在主库执行高并发任意读写操作情况下,实时验证主库和多个备库数据一致性。

[0115] 与上述方法实施例相对应,本说明书还提供了数据验证装置实施例,图3示出了本说明书一个实施例提供的一种数据验证装置的结构示意图。如图3所示,该装置包括:

[0116] 快照创建模块302,被配置为接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识;

[0117] 第一获取模块304,被配置为基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;

- [0118] 第二获取模块306,被配置为基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;
- [0119] 验证模块308,被配置为获取所述第一时间至所述第二时间的时间段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性。
- [0120] 可选的,所述验证模块308,进一步被配置为:
- [0121] 获取所述第一时间至所述第二时间的时间段内所述第一数据库的日志记录,并根据所述日志记录确定所述第一数据库的修改信息。
- [0122] 可选的,所述验证模块308,进一步被配置为:
- [0123] 基于所述修改信息验证所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息的一致性,以及
- [0124] 所述第一数据库的第一数据与所述第二数据库的第二数据的一致性。
- [0125] 可选的,所述验证模块308,进一步被配置为:
- [0126] 在基于所述修改信息确定所述第一数据库的第一数据属性信息存在增加的情况下,删除增加的第一数据属性信息;
- [0127] 将删除后的所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息进行匹配;
- [0128] 若匹配成功,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息一致;
- [0129] 若匹配失败,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息不一致。
- [0130] 可选的,所述验证模块308,进一步被配置为:
- [0131] 在基于所述修改信息确定所述第一数据库的第一数据属性信息存在删除的情况下,添加删除的第一数据属性信息;
- [0132] 将添加后的所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息进行匹配;
- [0133] 若匹配成功,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息一致;
- [0134] 若匹配失败,则确定同一时间点所述第一数据库的第一数据属性信息与所述第二数据库的第二数据属性信息不一致。
- [0135] 可选的,所述验证模块308,进一步被配置为:
- [0136] 在基于所述修改信息确定所述第一数据库的第一数据存在增加的情况下,删除增加的第一数据;
- [0137] 将删除后的所述第一数据库的第一数据与所述第二数据库的第二数据进行匹配;
- [0138] 若匹配成功,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据一致;
- [0139] 若匹配失败,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据不一致。
- [0140] 可选的,所述验证模块308,进一步被配置为:

[0141] 在基于所述修改信息确定所述第一数据库的第一数据存在删除的情况下,添加删除的第一数据;

[0142] 将添加后的所述第一数据库的第一数据与所述第二数据库的第二数据进行匹配;

[0143] 若匹配成功,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据一致;

[0144] 若匹配失败,则确定同一时间点所述第一数据库的第一数据与所述第二数据库的第二数据不一致。

[0145] 可选的,所述第一获取模块304,进一步被配置为:

[0146] 基于所述快照标识确定所述数据快照,且根据所述数据快照获取并缓存所述第一数据库的第一数据属性信息以及第一数据。

[0147] 可选的,所述第一获取模块306,进一步被配置为:

[0148] 基于所述快照标识确定所述数据快照,且根据所述数据快照获取并缓存所述第二数据库的第二数据属性信息以及第二数据。

[0149] 可选的,所述快照创建模块302,被配置为:

[0150] 按照预设的时间间隔接收验证指令。

[0151] 本说明书一个实施例实现了一种数据验证装置,其中,所述数据验证装置包括接收验证指令,基于所述验证指令创建第一数据库的数据快照,且为所述数据快照生成快照标识;基于所述快照标识获取所述第一数据库的第一数据属性信息以及第一数据,且记录获取结束的第一时间;基于所述快照标识获取第二数据库的第二数据属性信息以及第二数据,且记录获取结束的第二时间;获取所述第一时间至所述第二时间的时段内所述第一数据库的修改信息,且基于所述修改信息验证所述第一数据库和所述第二数据库的一致性;所述数据验证装置可以通过第一数据库的审计日志对基于同一数据快照获取的第一数据库和第二数据库的数据属性信息以及具体数据进行修正,实现实时对第一数据库和第二数据库的一致性验证。

[0152] 上述为本实施例的一种数据验证装置的示意性方案。需要说明的是,该数据验证装置的技术方案与上述的数据验证方法的技术方案属于同一构思,数据验证装置的技术方案未详细描述的细节内容,均可以参见上述数据验证方法的技术方案的描述。

[0153] 图4示出了根据本说明书一个实施例提供的一种计算设备400的结构框图。该计算设备400的部件包括但不限于存储器410和处理器420。处理器420与存储器410通过总线430相连接,数据库450用于保存数据。

[0154] 计算设备400还包括接入设备440,接入设备440使得计算设备400能够经由一个或多个网络460通信。这些网络的示例包括公用交换电话网(PSTN)、局域网(LAN)、广域网(WAN)、个域网(PAN)或诸如因特网的通信网络的组合。接入设备440可以包括有线或无线的任何类型的网络接口(例如,网络接口卡(NIC))中的一个或多个,诸如IEEE802.11无线局域网(WLAN)无线接口、全球微波互联接入(Wi-MAX)接口、以太网接口、通用串行总线(USB)接口、蜂窝网络接口、蓝牙接口、近场通信(NFC)接口,等等。

[0155] 在本说明书的一个实施例中,计算设备400的上述部件以及图4中未示出的其他部件也可以彼此相连接,例如通过总线。应当理解,图4所示的计算设备结构框图仅仅是出于示例的目的,而不是对本说明书范围的限制。本领域技术人员可以根据需要,增添或替换其

他部件。

[0156] 计算设备400可以是任何类型的静止或移动计算设备,包括移动计算机或移动计算设备(例如,平板计算机、个人数字助理、膝上型计算机、笔记本计算机、上网本等)、移动电话(例如,智能手机)、可佩戴的计算设备(例如,智能手表、智能眼镜等)或其他类型的移动设备,或者诸如台式计算机或PC的静止计算设备。计算设备400还可以是移动式或静止式的服务器。

[0157] 其中,处理器420用于执行如下计算机可执行指令,其中,所述处理器执行所述计算机可执行指令时实现所述数据验证方法的步骤。

[0158] 上述为本实施例的一种计算设备的示意性方案。需要说明的是,该计算设备的技术方案与上述的数据验证方法的技术方案属于同一构思,计算设备的技术方案未详细描述的细节内容,均可以参见上述数据验证方法的技术方案的描述。

[0159] 本说明书一实施例还提供一种计算机可读存储介质,其存储有计算机指令,该指令被处理器执行时实现所述数据验证方法的步骤。

[0160] 上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是,该存储介质的技术方案与上述的数据验证方法的技术方案属于同一构思,存储介质的技术方案未详细描述的细节内容,均可以参见上述数据验证方法的技术方案的描述。

[0161] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[0162] 所述计算机指令包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括电载波信号和电信信号。

[0163] 需要说明的是,对于前述的各方法实施例,为了简便描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本说明书实施例并不受所描述的动作顺序的限制,因为依据本说明书实施例,某些步骤可以采用其它顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定都是本说明书实施例所必须的。

[0164] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其它实施例的相关描述。

[0165] 以上公开的本说明书优选实施例只是用于帮助阐述本说明书。可选实施例并没有详尽叙述所有的细节,也不限制该发明仅为所述的具体实施方式。显然,根据本说明书实施例的内容,可作很多的修改和变化。本说明书选取并具体描述这些实施例,是为了更好地解释本说明书实施例的原理和实际应用,从而使所属技术领域技术人员能很好地理解和利用



本说明书。本说明书仅受权利要求书及其全部范围和等效物的限制。

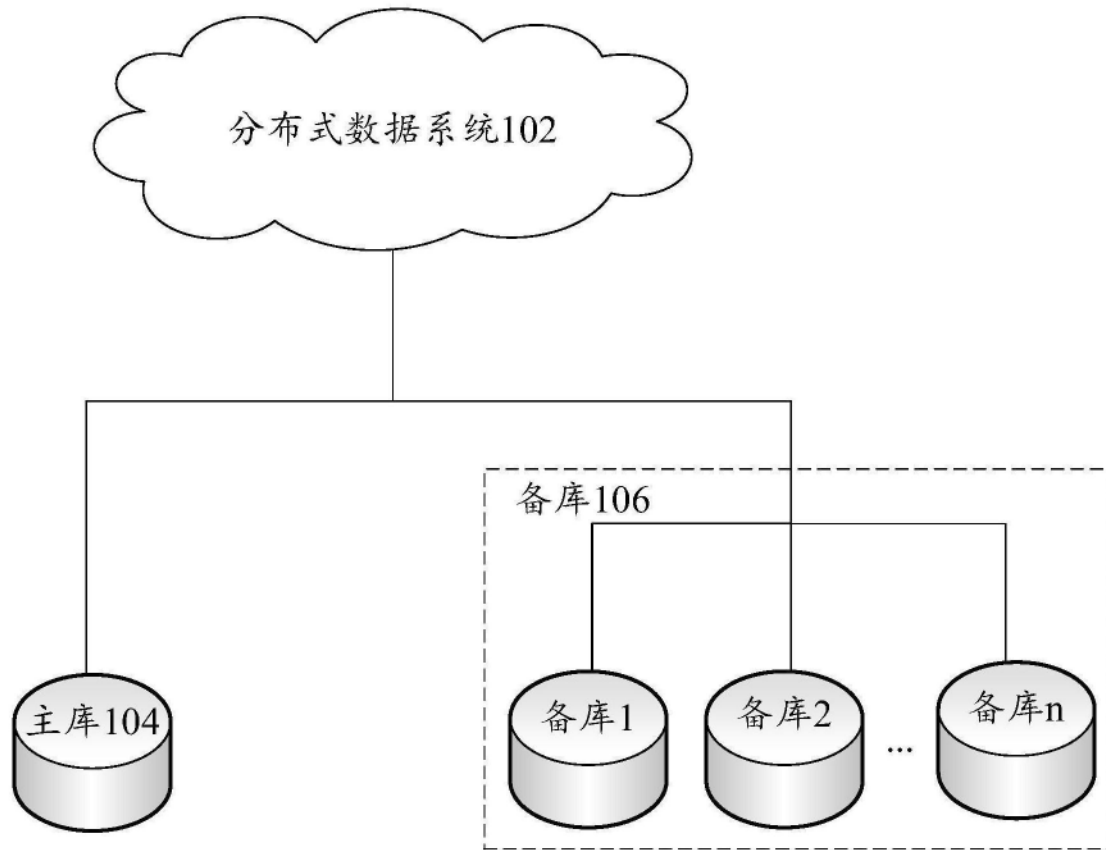


图1

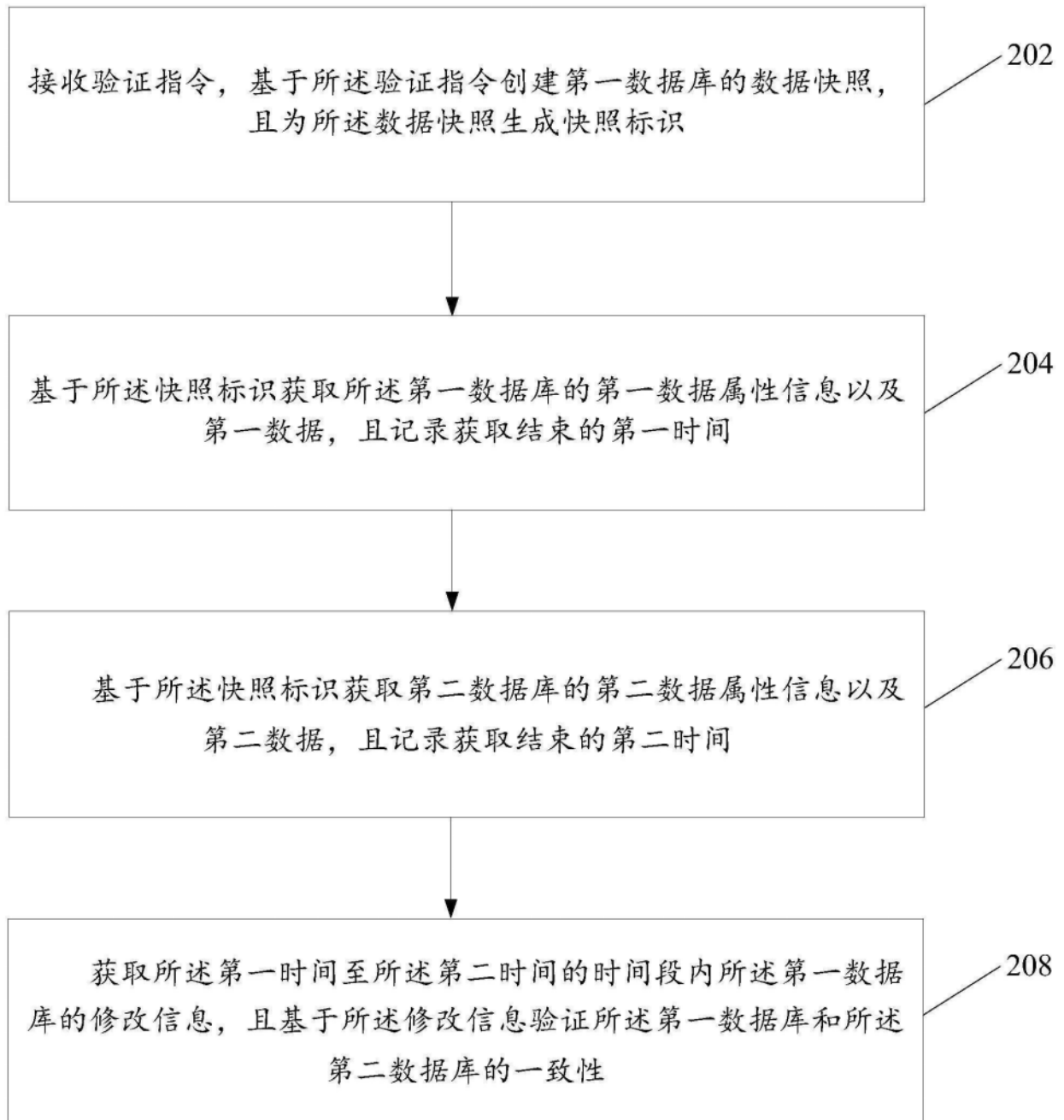


图2

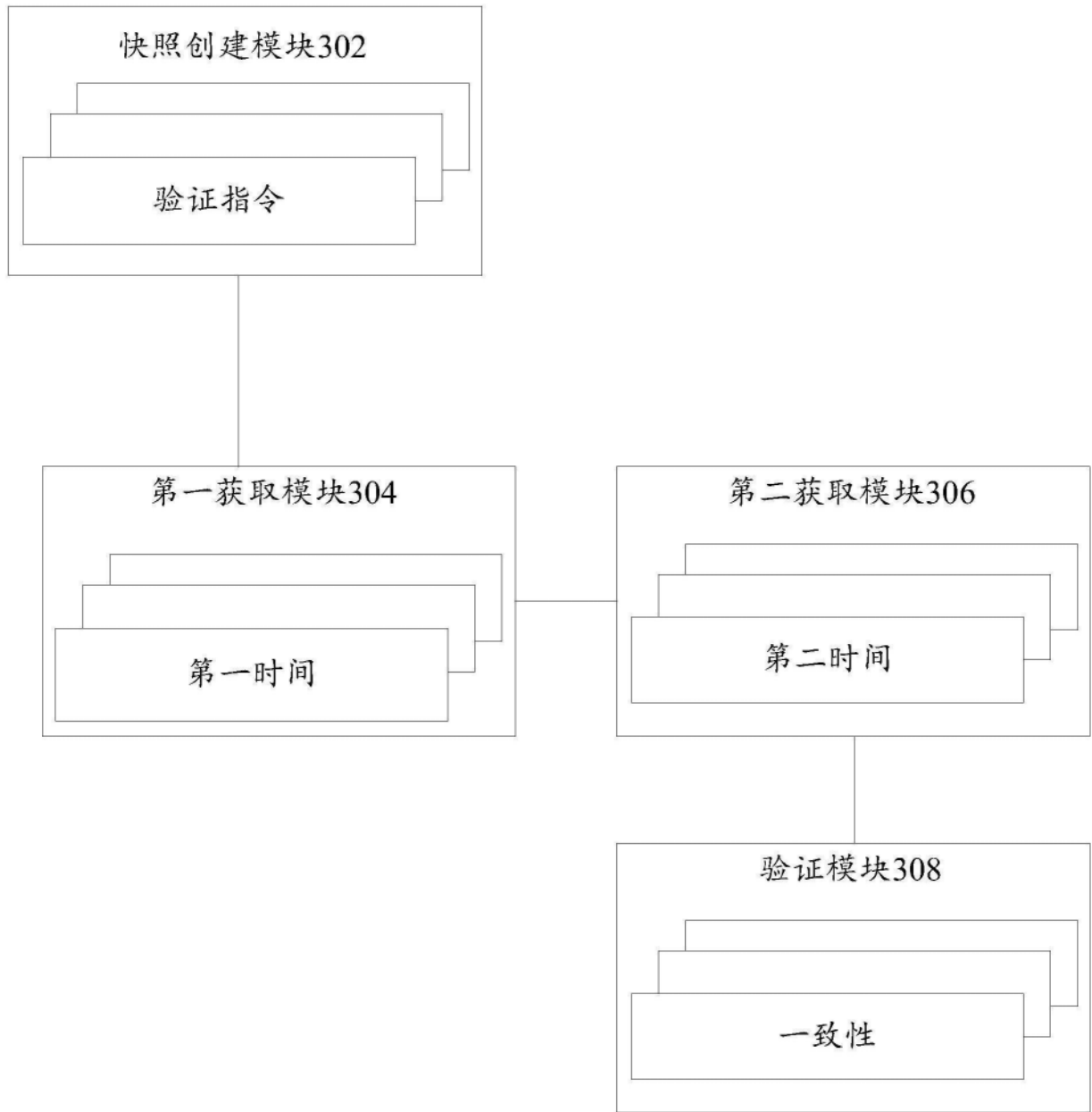


图3

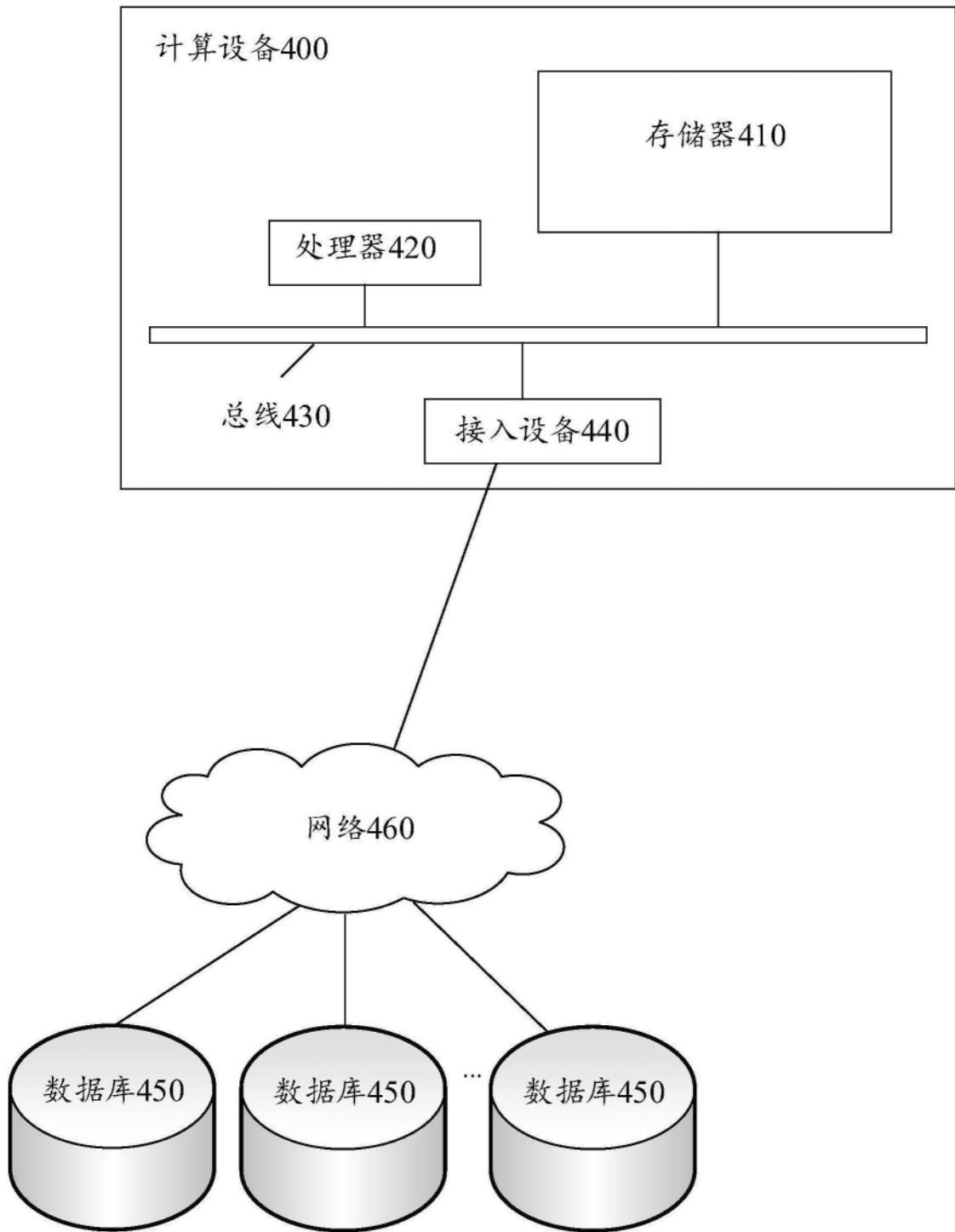


图4