



(12) 发明专利申请

(10) 申请公布号 CN 113342897 A

(43) 申请公布日 2021.09.03

(21) 申请号 202110724200.1

(22) 申请日 2021.06.29

(71) 申请人 深圳前海微众银行股份有限公司  
地址 518027 广东省深圳市前海深港合作区前湾一路1号A栋201室

(72) 发明人 石卫军 徐为恺 杨杨 江旻

(74) 专利代理机构 北京同达信恒知识产权代理有限公司 11291

代理人 宋正伟

(51) Int. Cl.

G06F 16/27 (2019.01)

G06F 16/28 (2019.01)

G06F 16/23 (2019.01)

G06F 16/22 (2019.01)

G06F 9/48 (2006.01)

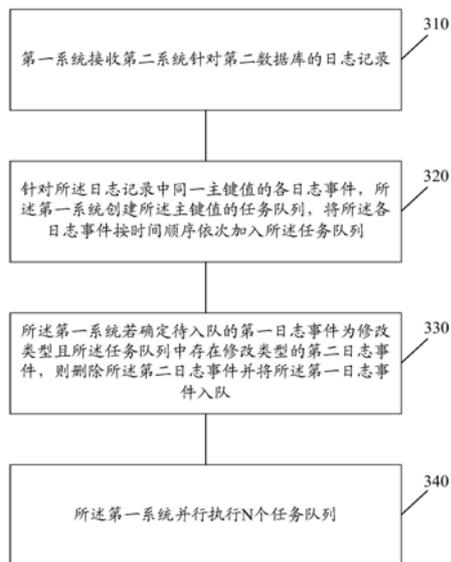
权利要求书2页 说明书10页 附图5页

(54) 发明名称

一种数据同步的方法及装置

(57) 摘要

本发明公开了一种数据同步的方法及装置,包括:第一系统接收第二系统针对第二数据库的日志记录;日志记录采用全量方式记录;针对日志记录中同一主键值的各日志事件,创建主键值的任务队列,将各日志事件按时间顺序依次加入所述任务队列;若确定待入队的第一日志事件为删除类型且任务队列中存在删除类型的第二日志事件,则删除第二日志事件并将第一日志事件入队;并行执行N个任务队列,以此减少任务队列中待写入的事件,降低数据写入的延迟,缩短数据写入的耗时,提升数据写入效率,增加实时搜索的效率。



1. 一种数据同步的方法,其特征在于,包括:

第一系统接收第二系统针对第二数据库的日志记录;所述第一系统中的第一数据库与所述第二数据库通过索引方式关联数据表;所述日志记录采用全量方式记录;

针对所述日志记录中同一主键值的各日志事件,所述第一系统创建所述主键值的任务队列,将所述各日志事件按时间顺序依次加入所述任务队列;

所述第一系统若确定待入队的第一日志事件为修改类型且所述任务队列中存在修改类型的第二日志事件,则删除所述第二日志事件并将所述第一日志事件入队;

所述第一系统并行执行N个任务队列。

2. 如权利要求1所述的方法,其特征在于,所述方法还包括:

所述第一系统若确定待入队的第一日志事件为新增类型,则将所述第一日志事件入队。

3. 如权利要求1所述的方法,其特征在于,所述方法还包括:

所述第一系统若确定待入队的第一日志事件为删除类型,则将所述任务队列中其他事件删除且将所述第一日志事件入队。

4. 如权利要求1所述的方法,其特征在于,针对所述日志记录中同一主键值的各日志事件,所述第一系统创建所述主键值的任务队列,包括:

所述第一系统确定所述日志记录中包含的各主键值及各主键值对应的日志事件;针对任一主键值,所述第一系统创建所述主键值的任务队列;或

针对任一日志事件,所述第一系统确定所述日志事件的主键值是否已创建任务队列,若没有,则创建所述主键的任务队列。

5. 如权利要求1所述的方法,其特征在于,所述第一系统并行执行N个任务队列,包括:

所述第一系统分别为所述N个任务队列分配各自的线程锁,并通过线程锁对应的线程将任务队列中的日志事件写入所述第一数据库。

6. 如权利要求5所述的方法,其特征在于,所述第一系统分别为所述N个任务队列分配各自的线程锁,包括:

针对任一任务队列,所述第一系统确定所述任务队列的主键值的哈希值;根据预设线程锁的数量N,对所述哈希值进行求余运算,得到余数;根据所述余数确定所述任务队列的线程锁。

7. 一种数据同步的装置,其特征在于,包括:

获取模块,用于接收第二系统针对第二数据库的日志记录;第一系统中的第一数据库与所述第二数据库通过索引方式关联数据表;所述日志记录采用全量方式记录;

处理模块,用于针对所述日志记录中同一主键值的各日志事件,创建所述主键值的任务队列,将所述各日志事件按时间顺序依次加入所述任务队列;

若确定待入队的第一日志事件为删除类型且所述任务队列中存在删除类型的第二日志事件,则删除所述第二日志事件并将所述第一日志事件入队;

并行执行N个任务队列。

8. 如权利要求7所述的装置,其特征在于,所述处理模块还用于:

若确定待入队的第一日志事件为新增类型,则将所述第一日志事件入队。

9. 一种计算机设备,其特征在于,包括:

存储器,用于存储程序指令;

处理器,用于调用所述存储器中存储的程序指令,按照获得的程序执行权利要求1至6任一项所述的方法。

10.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行权利要求1至6任一项所述的方法。

## 一种数据同步的方法及装置

### 技术领域

[0001] 本发明涉及金融科技(Fintech)领域,尤其涉及一种数据同步的方法及装置。

### 背景技术

[0002] 随着计算机技术的发展,越来越多的技术(例如:区块链、云计算或大数据)应用在金融领域,传统金融业正在逐步向金融科技转变,大数据技术也不例外,但由于金融、支付行业的安全性、实时性要求,也对大数据技术中数据同步提出了更高的要求。

[0003] 在现有数据中,业务数据通常以表格的形式保存在关系型数据库(如MySQL数据库)中,在MySQL数据库中的数据表发生数据变化时,会产生日志记录,为了保证搜索数据库的数据与MySQL数据库的数据一致,需要通过采集工具,采集日志记录,并将日志记录中的多个DML(Data manipulation language,数据操纵语言)事件,按照顺序写入搜索数据库中,以此实现数据同步。

[0004] 但是,在将日志记录中的多个DML写入搜索数据库的过程中,耗时较长,影响实时搜索的效率。

[0005] 因此,现需要一种数据同步的方法,用于降低数据写入的延迟,缩短数据写入的耗时,提升数据写入效率,增加实时搜索的效率。

### 发明内容

[0006] 本发明实施例提供一种数据同步的方法及装置,用于降低数据写入的延迟,缩短数据写入的耗时,提升数据写入效率,增加实时搜索的效率。

[0007] 第一方面,本发明实施例提供一种数据同步的方法,包括:

[0008] 第一系统接收第二系统针对第二数据库的日志记录;所述第一系统中的第一数据库与所述第二数据库通过索引方式关联数据表;所述日志记录采用全量方式记录;

[0009] 针对所述日志记录中同一主键值的各日志事件,所述第一系统创建所述主键值的任务队列,将所述各日志事件按时间顺序依次加入所述任务队列;

[0010] 所述第一系统若确定待入队的第一日志事件为删除类型且所述任务队列中存在删除类型的第二日志事件,则删除所述第二日志事件并将所述第一日志事件入队;

[0011] 所述第一系统并行执行N个任务队列。

[0012] 上述技术方案中,按照日志记录中各事件的事件顺序对各事件入队,针对任一待入队的第一日志事件,根据第一日志事件的类型来得到任务队列,具体的,若确定待入队的第一日志事件为修改类型且任务队列中存在修改类型的第二日志事件,则删除第二日志事件并将第一日志事件入队,得到任务队列,以此来减少任务队列中待写入的事件,降低数据写入的延迟,缩短数据写入的耗时。

[0013] 通过并行执行N个任务队列,实现并发处理各任务队列,以此提升数据写入效率,增加实时搜索的效率。

[0014] 可选的,所述方法还包括:

[0015] 所述第一系统若确定待入队的第一日志事件为新增类型,则将所述第一日志事件入队。

[0016] 上述技术方案中,新增类型的日志事件一般发生在其他类型事件之前,因为日志事件指的是DML事件,而DML事件包括三种事件类型,分别为新增类型、修改类型和删除类型,其中,新增类型的事件指的是新增主键值对应的数据,也就是说,任一任务队列,最多拥有一个新增类型的事件,因此,在对于第一日志事件为新增类型,则将第一日志事件入队,防止日志记录丢失。

[0017] 可选的,所述方法还包括:

[0018] 所述第一系统若确定待入队的第一日志事件为删除类型,则将所述任务队列中其他事件删除且将所述第一日志事件入队。

[0019] 上述技术方案中,针对第一日志事件为删除类型时,第一日志事件,即删除类型的事件,其中一般包括了日志删除前的数据和删除后的数据,因此可以将其他类型的事件删除,且不影响数据同步的结果,以此来减少任务队列中待写入的事件,降低数据写入的延迟,缩短数据写入的耗时。

[0020] 可选的,针对所述日志记录中同一主键值的各日志事件,创建所述主键值的任务队列,包括:

[0021] 所述第一系统确定所述日志记录中包含的各主键值及各主键值对应的日志事件;针对任一主键值,所述第一系统创建所述主键值的任务队列;或针对任一日志事件,所述第一系统确定所述日志事件的主键值是否已创建任务队列,若没有,则创建所述主键的任务队列。

[0022] 上述技术方案中,通过主键值区分任务队列,以避免在后续并行处理时,数据同步异常,提升数据的写入效率和准确率。

[0023] 可选的,所述第一系统并行执行N个任务队列,包括:

[0024] 所述第一系统分别为所述N个任务队列分配各自的线程锁,并通过线程锁对应的线程将任务队列中的日志事件写入所述第一数据库。

[0025] 上述技术方案中,通过线程锁来实现并行处理,针对任一线程,通过线程锁来防止任务拥挤,以此实现多个线程并发处理各任务队列,以此提升数据写入效率,增加实时搜索的效率。

[0026] 可选的,所述第一系统分别为所述N个任务队列分配各自的线程锁,包括:

[0027] 针对任一任务队列,所述第一系统确定所述任务队列的主键值的哈希值;根据预设线程锁的数量N,对所述哈希值进行求余运算,得到余数;根据所述余数确定所述任务队列的线程锁。

[0028] 上述技术方案中,根据线程锁的数量和哈希值确定第一线程锁,也就是说,不同主键值的任务队列,所对应的线程锁是不同的,由此实现多个线程并发处理各任务队列,以此提升数据写入效率,增加实时搜索的效率。

[0029] 第二方面,本发明实施例提供一种数据同步的装置,包括:

[0030] 获取模块,用于接收第二系统针对第二数据库的日志记录;第一系统中的第一数据库与所述第二数据库通过索引方式关联数据表;所述日志记录采用全量方式记录;

[0031] 处理模块,用于针对所述日志记录中同一主键值的各日志事件,创建所述主键值

的任务队列,将所述各日志事件按时间顺序依次加入所述任务队列;

[0032] 若确定待入队的第一日志事件为删除类型且所述任务队列中存在删除类型的第二日志事件,则删除所述第二日志事件并将所述第一日志事件入队;

[0033] 并行执行N个任务队列。

[0034] 可选的,所述处理模块还用于:

[0035] 若确定待入队的第一日志事件为新增类型,则将所述第一日志事件入队。

[0036] 可选的,所述处理模块还用于:

[0037] 若确定待入队的第一日志事件为删除类型,则将所述任务队列中其他事件删除且将所述第一日志事件入队。

[0038] 可选的,所述处理模块具体用于:

[0039] 确定所述日志记录中包含的各主键值及各主键值对应的日志事件;针对任一主键值,所述第一系统创建所述主键值的任务队列;或针对任一日志事件,确定所述日志事件的主键值是否已创建任务队列,若没有,则创建所述主键的任务队列。

[0040] 可选的,所述处理模块具体用于:

[0041] 分别为所述N个任务队列分配各自的线程锁,并通过线程锁对应的线程将任务队列中的日志事件写入所述第一数据库。

[0042] 可选的,所述处理模块具体用于:

[0043] 针对任一任务队列,确定所述任务队列的主键值的哈希值;根据预设线程锁的数量N,对所述哈希值进行求余运算,得到余数;根据所述余数确定所述任务队列的线程锁。

[0044] 第三方面,本发明实施例还提供一种计算机设备,包括:

[0045] 存储器,用于存储程序指令;

[0046] 处理器,用于调用所述存储器中存储的程序指令,按照获得的程序执行上述数据同步的方法。

[0047] 第四方面,本发明实施例还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行上述数据同步的方法。

## 附图说明

[0048] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0049] 图1为本发明实施例提供的一种数据同步的结构示意图;

[0050] 图2为本发明实施例提供的一种系统架构示意图;

[0051] 图3为本发明实施例提供的一种数据同步的方法的流程示意图;

[0052] 图4为本发明实施例提供的一种日志事件入队的流程示意图;

[0053] 图5为本发明实施例提供的一种数据同步的流程示意图;

[0054] 图6为本发明实施例提供的一种数据同步的装置的结构示意图。

## 具体实施方式

[0055] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0056] 现有技术中,在第二数据库(如MySQL数据库)中数据发生变化时,会产生日志记录,如binlog日志,日志记录中包括了预设时间窗口内的多个DML事件,其中,DML事件包括三中类型的事件,分别为新增类型、修改类型和删除类型,在进行数据同步时,需要将日志记录中的多个DML事件按照变化时间将DML事件写入第一数据库(如ES索引数据库)。

[0057] 图1示例性的示出了一种数据同步的结构示意图,如图1所示,MySQL系统生成binlog日志,然后ES系统通过采集工具获取binlog日志,然后将binlog日志写入ES数据库,以此实现数据同步。

[0058] 然而,对于DML事件,其包括了数据变化前的数据和变化后的数据,因此,针对某些类型的事件是无效的、多余的,因此导致在写入事件时,影响写入效率。

[0059] 如下述表1和表2示出的两个修改类型的DML事件,表1事件将姓名进行了修改,表2事件将性别进行修改。其中,表2所示的事件包括了表1事件的修改,因此,表1事件在写入时是无效的,多余的,可以进行删除。

[0060] 表1

[0061]		身份证号码	姓名	性别
	变化前	210***	崔**	男
	变化后	210***	王**	男

[0062] 表2

[0063]		身份证号码	姓名	性别
	变化前	210***	崔**	男
	变化后	210***	王**	女

[0064] 且在进行事件写入时,目前,为了确保数据不丢失,数据不乱序,按照时间顺序,通过一个线程依次将事件写入第一数据库,这导致写入时延迟较高,实时搜索的效率低,写入效率低。

[0065] 因此,现需要一种数据同步的方法,来降低数据写入的延迟,缩短数据写入的耗时,提升数据写入效率,增加实时搜索的效率。

[0066] 图2示例性的示出了本发明实施例所适用的一种系统架构,该系统架构包括数据采集模块210、数据聚合模块220、线程分配模块230和执行写入模块240。

[0067] 其中,数据采集模块210,用于获取第二系统针对第二数据库的日志记录,日志记录中包括多个DML事件。

[0068] 数据聚合模块220,用于根据主键值确定各任务队列,并将各日志事件按时间顺序和事件类型以此加入任务队列,确定出各任务队列中待写入的事件。

[0069] 线程分配模块230,用于根据主键值确定执行任务队列的线程锁。

[0070] 执行写入模块240,用于根据线程锁对应的线程将任务队列中的待写入事件写入至第一数据库中。

[0071] 需要说明的是,上述图2所示的结构仅是一种示例,本发明实施例对此不做限定。

[0072] 基于上述描述,图3示例性的示出了本发明实施例提供的一种数据同步的方法的流程示意图,该流程可由数据同步的装置执行。

[0073] 如图3所示,该流程具体包括:

[0074] 步骤310,第一系统接收第二系统针对第二数据库的日志记录;

[0075] 本发明实施例中,第一系统中的第一数据库与第二数据库通过索引方式关联数据表;其中,日志记录采用全量方式记录。

[0076] 步骤320,针对所述日志记录中同一主键值的各日志事件,所述第一系统创建所述主键值的任务队列,将所述各日志事件按时间顺序依次加入所述任务队列。

[0077] 本发明实施例中,根据各日志事件的主键值,将各日志事件划分为不同任务队列中的日志事件,以得到主键值对应的任务队列,以保证数据不乱序,保证数据的准确性。

[0078] 步骤330,所述第一系统若确定待入队的第一日志事件为修改类型且所述任务队列中存在修改类型的第二日志事件,则删除所述第二日志事件并将所述第一日志事件入队。

[0079] 本发明实施例中,根据日志事件的类型,确定任务队列中待写入的事件,以减少待写入事件的数量,提升写入效率。

[0080] 步骤340,所述第一系统并行执行N个任务队列。

[0081] 本发明实施例中,根据任务队列的主键值,确定执行任务队列的线程,以此实现并行执行任务队列,提升写入效率。

[0082] 在步骤310中,全量方式记录指的是针对任一日志事件,日志事件中包括数据变化前的数据和变化后的一整行数据,如下述表3所示,第二数据库中的某数据表中包括用户信息,共四个字段,其中,字段“身份证号码”为主键,该事件为修改类型的事件,具体事件内容为将年龄21,修改为年龄18,因此在日志事件中,包括数据变化前的四个字段值和数据变化后的四个字段值。

[0083] 表3

	身份证号码	姓名	性别	年龄
变化前	210***	王**	男	21
变化后	210***	王**	男	18

[0085] 步骤320中,因为各日志事件中包括主键值,因此可以根据主键值来区分各日志事件。具体的,第一系统确定日志记录中包含的各主键值及各主键值对应的日志事件;针对任一主键值,创建所述主键值的任务队列;或针对任一日志事件,确定所述日志事件的主键值是否已创建任务队列,若没有,则创建所述主键的任务队列。

[0086] 举例来说,在某一日志记录中包括5个日志事件中,按照时间顺序分别为a1、b2、a3、b4和b5,其中,a1和a3是主键值A的事件,a2、a4和a5是主键值B的事件,因此可以创建两个任务队列,分别为任务队列A和任务队列B,然后按照时间顺序将对应的日志事件加入任务队列。

[0087] 以任务队列B进行举例,首先将日志事件b2加入至任务队列B,再将日志事件b4加入至任务队列B,最后将日志事件b5加入至任务队列B。

[0088] 在步骤330中,针对任一待入队的第一日志事件,根据第一日志事件的类型,来对

任务队列中的其他事件进行处理,具体的,第一日志事件为修改类型时,若任务队列中存在修改类型的第二日志事件,则将修改类型的第二日志事件删除,并将第一日志事件入队。

[0089] 结合上述示例进行举例,例如任务队列B中现包括日志事件b2、日志事件b4和日志事件b5,其中,日志事件b2为新增类型的日志事件,日志事件b4和日志事件b5为修改类型的日志事件,待入队的第一日志事件b6的事件类型为修改类型,则保留任务队列B中的日志事件b2,将日志事件b4和日志事件b5删除,将第一日志事件b6加入至任务队列B,此时,任务队列B中的日志事件包括日志事件b2和日志事件b6。

[0090] 在一种可实施的方式中,第一系统若确定待入队的第一日志事件为新增类型,则将所述第一日志事件入队。

[0091] 结合上述示例进行举例,日志事件b2为新增类型的日志事件且为待入队的第一日志事件,然后将日志事件b2加入至任务队列B。

[0092] 在另一种可实施的方式中,所述第一系统若确定待入队的第一日志事件为删除类型,则将所述任务队列中其他事件删除且将所述第一日志事件入队。

[0093] 结合上述示例进行举例,任务队列B中的日志事件包括日志事件b2和日志事件b6,其中,日志事件b2为新增类型的日志事件,日志事件b6为修改类型的日志事件,待入队的第一日志事件b7的事件类型为删除类型,则将任务队列B中的日志事件b2和日志事件b6删除,将第一日志事件b7加入至任务队列B,此时,任务队列B中的日志事件仅包括日志事件b7。

[0094] 以此实现减少任务队列中待写入的事件,避免无效事件、多余事件重复的写入第一数据库,降低数据写入的延迟,缩短数据写入的耗时。

[0095] 为了更好的阐述上述技术方案,图4示例性的示出了一种日志事件入队的流程示意图,如图4所示,包括:

[0096] 步骤410,接收日志记录。

[0097] 日志记录中包括4个日志事件,按照时间顺序分别为“新增用户信息”的日志事件c1、“修改姓名”日志事件c2、“修改身份证”日志事件c3和“修改性别”日志事件c4,也就是说,日志事件c1的事件类型为新增类型,日志事件c2的事件类型为修改类型,日志事件c3的事件类型为修改类型,日志事件c4的事件类型为修改类型。

[0098] 用户信息为主键,因此,日志事件c1、日志事件c2、日志事件c3和日志事件c4为同一任务队列C的日志事件。

[0099] 步骤420,确定是否为新增类型,若是,则执行步骤430,否则执行步骤440。

[0100] 步骤430,日志事件入队。

[0101] 例如待入队的第一日志事件为“新增用户信息”,则将日志事件c1加入至任务队列C。

[0102] 步骤440,确定日志事件是否为修改类型,若是,则执行步骤450,否则执行步骤480。

[0103] 步骤450,确定队列中是否存在修改类型的日志事件,若是,则执行步骤460,否则执行步骤470。

[0104] 步骤460,确定日志事件是否为删除类型,若是,则执行步骤490。

[0105] 步骤470,清空修改类型的日志事件。

[0106] 例如,若待入队的第一日志事件为“修改姓名”,此时的任务队列C中包括日志事件

c1,不包括其他修改类型的日志事件,则将日志事件c2加入至任务队列C。

[0107] 若待入队的第一日志事件为“修改身份证”,此时的任务队列C中包括日志事件c1和日志事件c2,包括其他修改类型的日志事件(日志事件c2),则将任务队列C中的日志事件c2删除,将日志事件c3加入至任务队列C。

[0108] 若待入队的第一日志事件为“修改性别”,此时的任务队列C中包括日志事件c1和日志事件c3,包括其他修改类型的日志事件(日志事件c3),则将任务队列C中的日志事件c3删除,将日志事件c4加入至任务队列C。

[0109] 此时,任务队列C中包括日志事件c1和日志事件c4,减少了待写入的日志事件,提升数据同步效率。

[0110] 步骤480,修改类型的日志事件入队。

[0111] 步骤490,清空任务队列。

[0112] 例如,日志记录中还包括了与日志事件c1、日志事件c2、日志事件c3和日志事件c4相同主键值的日志事件c5,且日志事件c5按照时间顺序,在日志事件c4之后,若待入队的第一日志事件为日志事件c5,此时的任务队列C包括日志事件c1和日志事件c4,则将日志事件c1和日志事件c4删除,将日志事件c5入队,使任务队列C中仅包括日志事件c5。

[0113] 在步骤340中,针对不同主键值的任务队列,可以根据不同的线程并行处理,而执行任务队列的线程是根据主键值来确定的。

[0114] 具体的,第一系统分别为N个任务队列分配各自的线程锁,并通过线程锁对应的线程将任务队列中的日志事件写入所述第一数据库。

[0115] 进一步地,针对任一任务队列,所述第一系统确定所述任务队列的主键值的哈希值;根据预设线程锁的数量N,对所述哈希值进行求余运算,得到余数;根据所述余数确定所述任务队列的线程锁。

[0116] 举例来说,根据哈希算法得到主键值的哈希值,然后对哈希值取预设位整数的绝对值,例如,哈希公式为: $ABS(s[0]*31^{(n-1)}+s[1]*31^{(n-2)}+\dots+s[n-1])$ ,其中,ABS为对整数求绝对值,s[i]是字符串的第i个字符,n是字符串的长度,^表示求幂,以此得到哈希值为1208821703。

[0117] 预设线程锁的数量为6个,线程锁的标识分别为L1、L2、L3、L4、L5和L6,对上述哈希值取余数,余数为5,以此确定哈希值为1208821703的任务队列通过线程锁L5的线程写入至第一数据库。

[0118] 需要说明的是,预设位整数可以是根据经验人为预设的,如8、10位等,本发明实施例以10为例,在此不做具体限定。

[0119] 在另一种可实施的方式中,通过一致性哈希算法,来确定任务队列的线程锁,例如,将预设的线程锁分布在哈希环上,将任务队列的主键值的哈希值映射至哈希环上,以此得到任务队列的线程锁。

[0120] 结合上述示例举例来说,将线程锁L1、L2、L3、L4、L5和L6顺时针映射至哈希环上,针对某一任务队列的主键值的哈希值,该哈希值映射在哈希环上线程锁L1和L2之间,根据顺时针规则,通过线程锁L2对应的线程将该任务队列中的日志事件写入至第一数据库。

[0121] 为了更好的解释上述技术方案,图5示例性的示出了一种数据同步的流程示意图,如图5所示,流程包括:

- [0122] 步骤510,数据采集。
- [0123] 在数据库中采集日志记录,日志记录中包括3个用户的日志事件,分别为用户D、用户E和用户F,其中,主键为身份证,用户D的主键值,即身份证号码为120\*\*\*35、用户E的主键值为120\*\*\*36、用户F的主键值为120\*\*\*37。
- [0124] 数据库可以为多个,在此不做具体限定。
- [0125] 步骤520,数据聚合。
- [0126] 根据用户的主键值将各日志事件进行划分,得到对应的任务队列,分别为任务队列D、任务队列E和任务队列F,其中,n表示日志事件的数量。
- [0127] 以用户F进行举例,用户F包括7条日志事件,按照时间顺序分别为“删除用户名事件F1”、“删除身份证事件F2”、“删除用户头像事件F3”、“删除住址信息事件F4”、“修改企业信用代码事件F5”、“修改个信用代码事件F6”和“修改手机号事件F7”。
- [0128] 根据上述图4所示的将日志事件入队的技术方案可以确定出,首先,将日志事件F1入队,然后在日志事件F2入队时,将日志事件F1删除,日志事件F2入队,在日志事件F3入队时,将日志事件F2删除,日志事件F3入队,在日志事件F4入队时,将日志事件F3删除,日志事件F4入队,此时任务队列F包括日志事件F4。
- [0129] 将日志事件F5入队,然后在日志事件F6入队时,将日志事件F5删除,日志事件F6入队,在日志事件F7入队时,将日志事件F6删除,日志事件F7入队,此时任务队列F包括日志事件F4和日志事件F7。
- [0130] 同理,得到任务队列D和任务队列E。
- [0131] 步骤530,锁分配。
- [0132] 预设线程锁包括L1、L2、L3、L4、L5和L6,确定主键值的哈希值,得到用户D的哈希值为1208829034、用户E的哈希值为1208833841、用户F的哈希值为1208829040。
- [0133] 分别进行求余计算,得到用户D的余数为4、用户E的余数为5、用户F的余数为4,以此确定用户D的线程锁为L4、用户E的线程锁为L5、用户F的线程锁为L4。
- [0134] 步骤540,写入第一数据库。
- [0135] 通过线程锁L4的线程将任务队列D和任务队列F写入第一数据库。
- [0136] 通过线程锁L5的线程将任务队列E写入第一数据库。
- [0137] 针对任一任务队列,按照任务队列中日志事件的时间顺序将日志事件写入第一数据库。对于线程锁L4的线程,按照任务队列获取线程锁L4的顺序,以此将任务队列写入第一数据库,以保证数据不乱序,保证数据准确性。
- [0138] 基于相同的技术构思,图6示例性的示出了本发明实施例提供的一种数据同步的装置的结构示意图,该装置可以执行的流程。
- [0139] 如图6所示,该装置具体包括:
- [0140] 获取模块610,用于接收第二系统针对第二数据库的日志记录;第一系统中的第一数据库与所述第二数据库通过索引方式关联数据表;所述日志记录采用全量方式记录;
- [0141] 处理模块620,用于针对所述日志记录中同一主键值的各日志事件,创建所述主键值的任务队列,将所述各日志事件按时间顺序依次加入所述任务队列;
- [0142] 若确定待入队的第一日志事件为删除类型且所述任务队列中存在删除类型的第二日志事件,则删除所述第二日志事件并将所述第一日志事件入队;

- [0143] 并行执行N个任务队列。
- [0144] 可选的,所述处理模块620还用于:
- [0145] 若确定待入队的第一日志事件为新增类型,则将所述第一日志事件入队。
- [0146] 可选的,所述处理模块620还用于:
- [0147] 若确定待入队的第一日志事件为删除类型,则将所述任务队列中其他事件删除且将所述第一日志事件入队。
- [0148] 可选的,所述处理模块620具体用于:
- [0149] 确定所述日志记录中包含的各主键值及各主键值对应的日志事件;针对任一主键值,所述第一系统创建所述主键值的任务队列;或针对任一日志事件,确定所述日志事件的主键值是否已创建任务队列,若没有,则创建所述主键的任务队列。
- [0150] 可选的,所述处理模块620具体用于:
- [0151] 分别为所述N个任务队列分配各自的线程锁,并通过线程锁对应的线程将任务队列中的日志事件写入所述第一数据库。
- [0152] 可选的,所述处理模块620具体用于:
- [0153] 针对任一任务队列,确定所述任务队列的主键值的哈希值;根据预设线程锁的数量N,对所述哈希值进行求余运算,得到余数;根据所述余数确定所述任务队列的线程锁。
- [0154] 基于相同的技术构思,本发明实施例还提供一种计算机设备,包括:
- [0155] 存储器,用于存储程序指令;
- [0156] 处理器,用于调用所述存储器中存储的程序指令,按照获得的程序执行上述数据同步的方法。
- [0157] 基于相同的技术构思,本发明实施例还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行上述数据同步的方法。
- [0158] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。
- [0159] 本申请是参照根据本申请的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。
- [0160] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。
- [0161] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计

计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0162] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

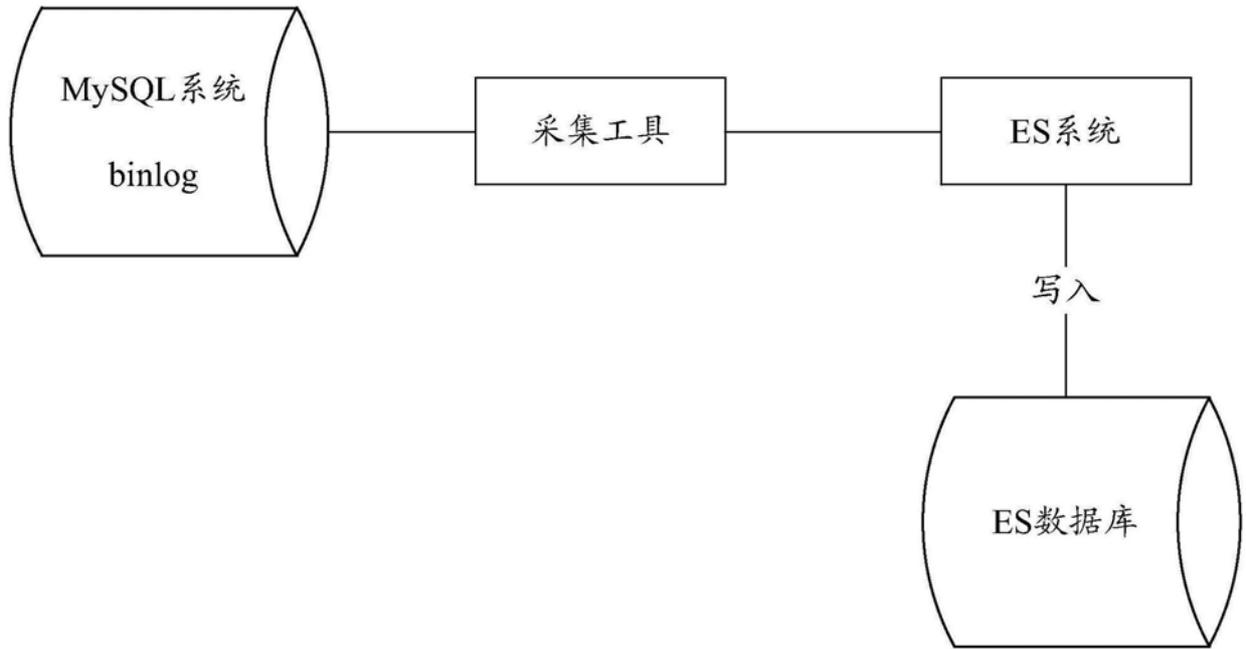


图1

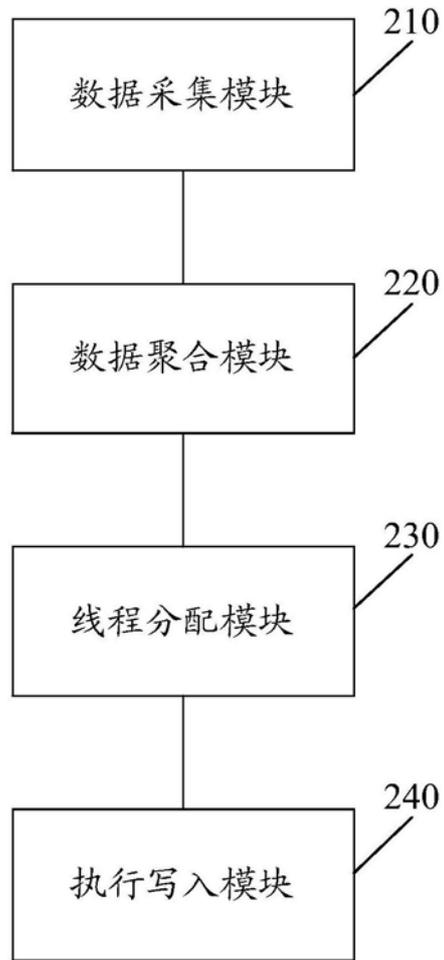


图2

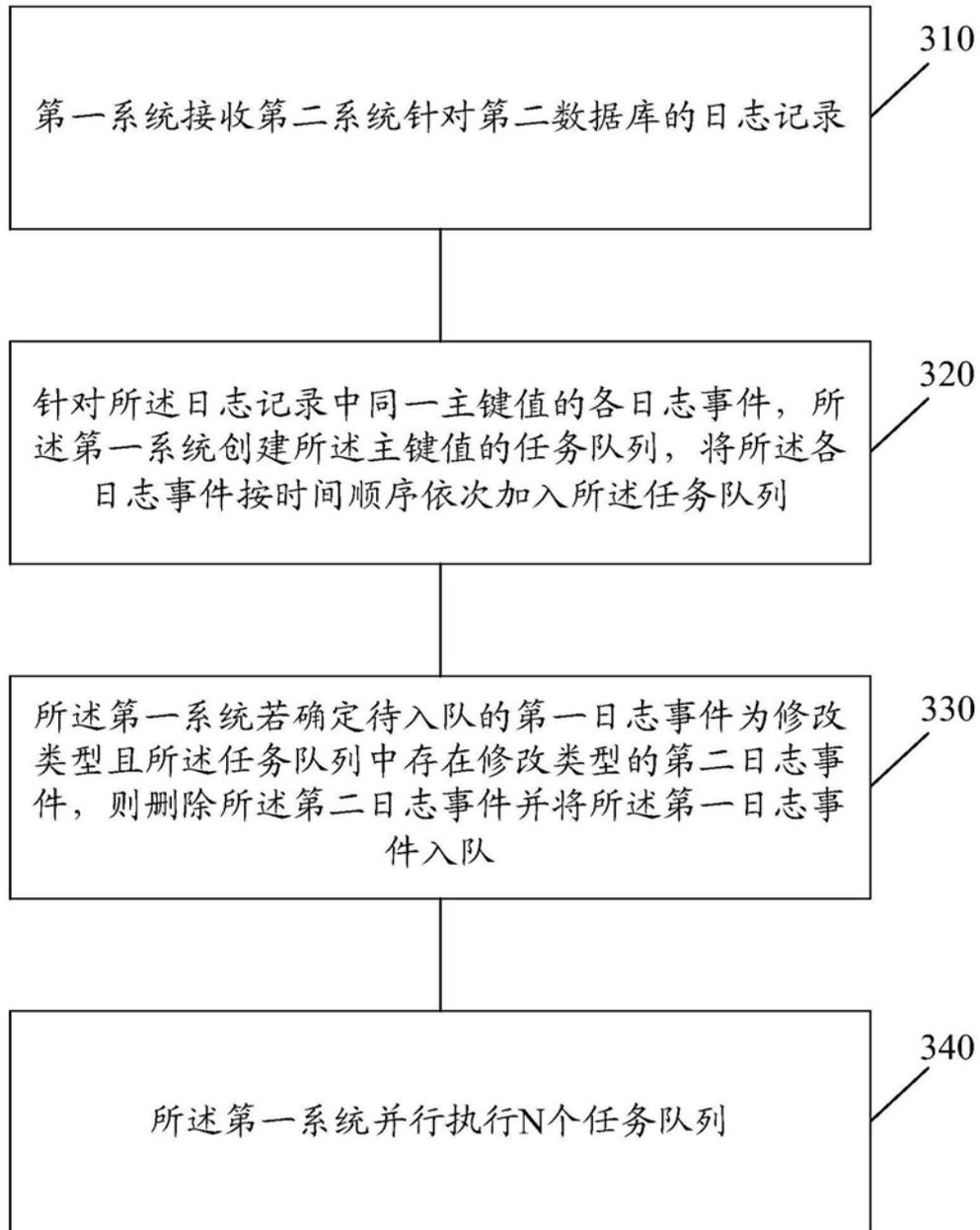


图3

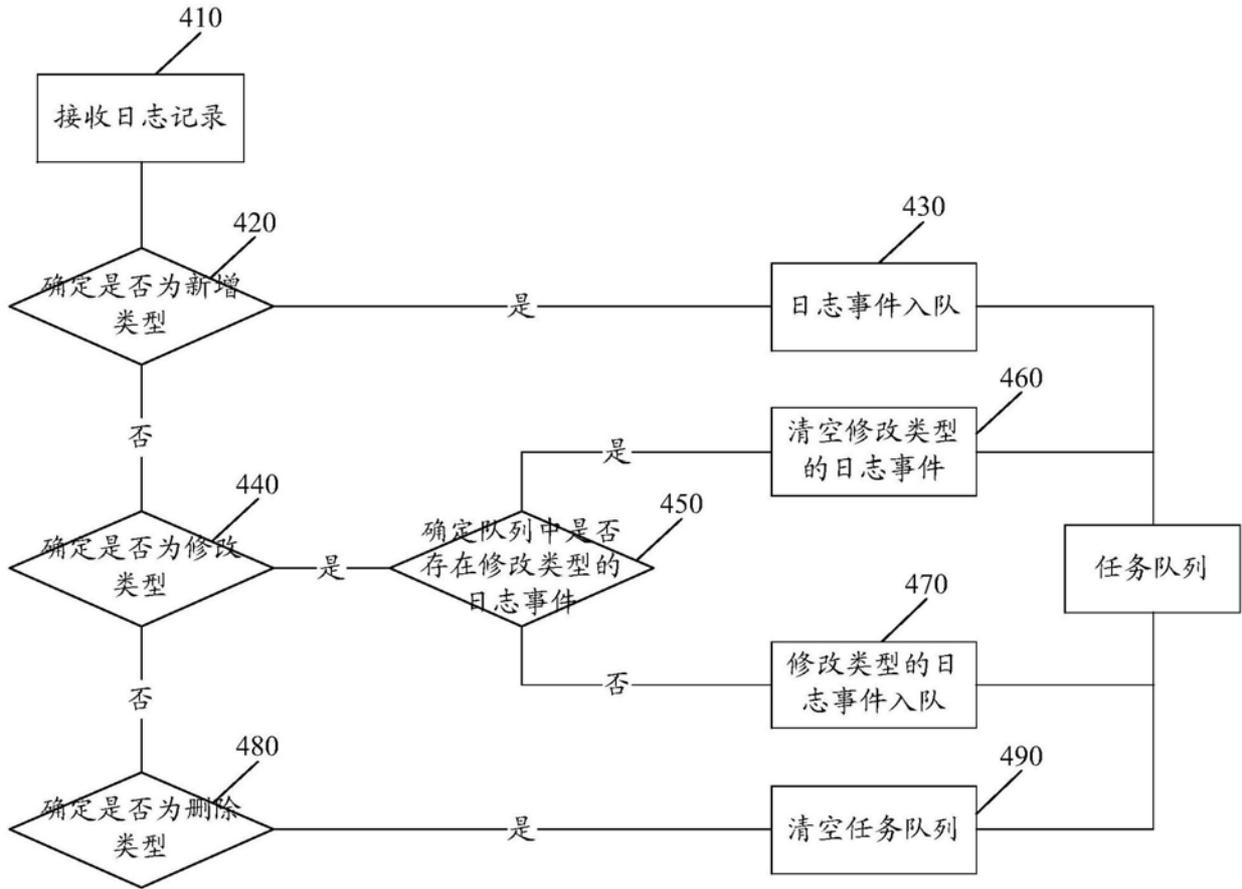


图4

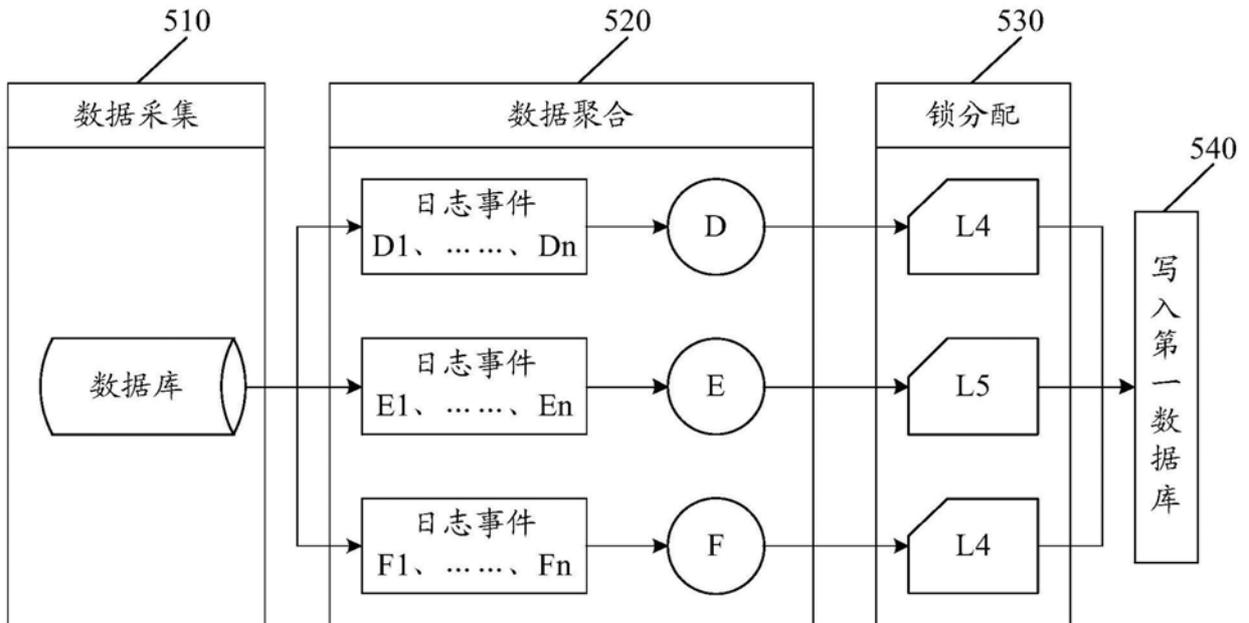


图5

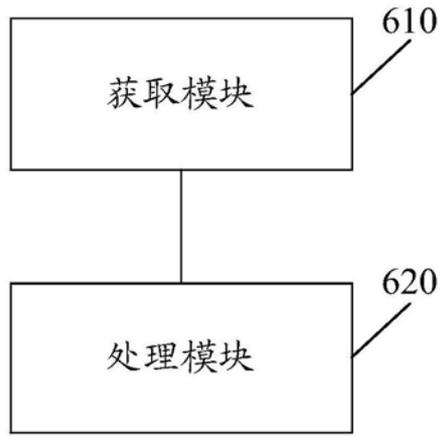


图6