



(12) 发明专利申请

(10) 申请公布号 CN 113961641 A

(43) 申请公布日 2022. 01. 21

(21) 申请号 202111138224.5

(22) 申请日 2021.09.27

(71) 申请人 北京百度网讯科技有限公司

地址 100085 北京市海淀区上地十街10号
百度大厦2层

(72) 发明人 缪恒锋 陈梦林 卓泽城 张晓聪

(74) 专利代理机构 北京鸿德海业知识产权代理
有限公司 11412

代理人 田宏宾

(51) Int. Cl.

G06F 16/27 (2019.01)

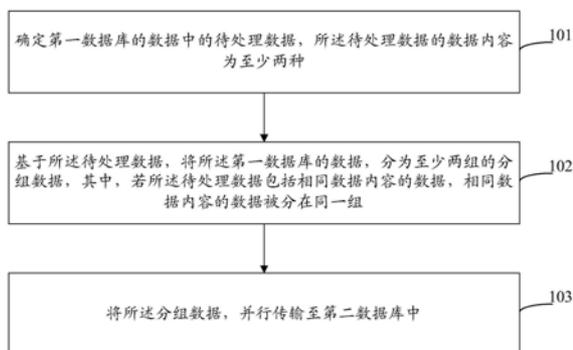
权利要求书3页 说明书9页 附图6页

(54) 发明名称

数据库同步方法、装置、设备和存储介质

(57) 摘要

本公开提供了一种数据库同步方法、装置、设备和存储介质，涉及计算机技术领域，具体涉及大数据、云计算等人工智能领域。数据库同步方法包括：确定第一数据库的数据中的待处理数据，所述待处理数据的数据内容为至少两种；基于所述待处理数据，将所述第一数据库的数据，分为至少两组的分组数据，其中，若所述待处理数据包括相同数据内容的数据，相同数据内容的数据被分在同一组；将所述分组数据，并行传输至第二数据库中。本公开可以提高数据库的同步效率。



1. 一种数据库同步方法,包括:
 - 确定第一数据库的数据中的待处理数据,所述待处理数据的数据内容为至少两种;
 - 基于所述待处理数据,将所述第一数据库的数据,分为至少两组的分组数据,其中,若所述待处理数据包括相同数据内容的数据,相同数据内容的数据被分在同一组;
 - 将所述分组数据,并行传输至第二数据库中。
2. 根据权利要求1所述的方法,其中,所述第一数据库中的数据以数据表形式存储,所述数据表包括数据列,所述确定第一数据库的数据中的待处理数据,包括:
 - 获取所述数据表中的所述数据列包括的不同数据内容的数量;
 - 将所述不同数据内容的数量最大的数据列中的数据,作为所述待处理数据。
3. 根据权利要求2所述的方法,其中,所述获取所述数据表中的所述数据列包括的不同数据内容的数量,包括:
 - 采用所述第一数据库的领域专用语言DSL,获取所述数据表中的所述数据列包括的不同数据内容的数量。
4. 根据权利要求2所述的方法,其中,所述不同数据内容的数量最大的数据列为当前列,所述数据表还包括数据行,所述基于所述待处理数据,将所述第一数据库中的数据,分为至少两个分组的分组数据,包括:
 - 确定所述当前列中的各个数据的分组信息;
 - 将所述当前列中的各个数据所在的数据行中的数据,分到所述各个数据的分组信息对应的分组内。
5. 根据权利要求4所述的方法,其中,所述确定所述当前列中的各个数据的分组信息,包括:
 - 将所述当前列包括的不同数据内容的数量作为最大分组数;
 - 选择不大于所述最大分组数的数值,作为当前分组数;
 - 获取所述当前列中的各个数据的哈希值;
 - 基于所述哈希值和所述当前分组数,确定所述当前列中的各个数据的分组信息。
6. 根据权利要求5所述的方法,其中,所述基于所述哈希值和所述当前分组数,确定所述当前列中的各个数据的分组信息,包括:
 - 获得所述哈希值对应的十进制数值;
 - 将所述当前分组数作为模,对所述十进制数值进行求余运算,以获得余数;
 - 将所述余数作为所述各个数据的分组信息。
7. 根据权利要求1-6任一项所述的方法,其中,所述将所述分组数据,并行传输至第二数据库中,包括:
 - 将所述分组数据,并行传输至消息队列中;
 - 从所述消息队列中,并行读取所述分组数据,以及将读取的所述分组数据,并行传输至第二数据库中。
8. 根据权利要求1-6任一项所述的方法,还包括:
 - 若所述分组数据传输至所述第二数据库时发生异常,响应于对所述异常的分组数据的重传指令,在所述第二数据库中删除所述异常的分组数据,并在删除所述异常的分组数据之后,重传所述异常的分组数据。

9. 根据权利要求8所述的方法,其中,所述在所述第二数据库中删除所述异常的分组数据,包括:

采用所述第二数据库的DSL,在所述第二数据库中删除所述异常的分组数据。

10. 一种数据库同步装置,包括:

确定模块,用于确定第一数据库的数据中的待处理数据,所述待处理数据的数据内容为至少两种;

分组模块,用于基于所述待处理数据,将所述第一数据库的数据,分为至少两组的分组数据,其中,若所述待处理数据包括相同数据内容的数据,相同数据内容的数据被分在同一组;

传输模块,用于将所述分组数据,并行传输至第二数据库中。

11. 根据权利要求10所述的装置,其中,所述第一数据库中的数据以数据表形式存储,所述数据表包括数据列,所述确定模块具体用于:

获取所述数据表中的所述数据列包括的不同数据内容的数量;

将所述不同数据内容的数量最大的数据列中的数据,作为所述待处理数据。

12. 根据权利要求11所述的装置,其中,所述确定模块进一步具体用于:

采用所述第一数据库的领域专用语言DSL,获取所述数据表中的所述数据列包括的不同数据内容的数量。

13. 根据权利要求11所述的装置,其中,所述不同数据内容的数量最大的数据列为当前列,所述数据表还包括数据行,所述分组模块包括:

确定单元,用于确定所述当前列中的各个数据的分组信息;

分组单元,用于将所述当前列中的各个数据所在的数据行中的数据,分到所述各个数据的分组信息对应的分组内。

14. 根据权利要求13所述的装置,其中,所述确定单元具体用于:

将所述当前列包括的不同数据内容的数量作为最大分组数;

选择不大于所述最大分组数的数值,作为当前分组数;

获取所述当前列中的各个数据的哈希值;

基于所述哈希值和所述当前分组数,确定所述当前列中的各个数据的分组信息。

15. 根据权利要求14所述的装置,其中,所述确定单元进一步具体用于:

获得所述哈希值对应的十进制数值;

将所述当前分组数作为模,对所述十进制数值进行求余运算,以获得余数;

将所述余数作为所述各个数据的分组信息。

16. 根据权利要求10-15任一项所述的装置,其中,所述传输模块具体用于:

将所述分组数据,并行传输至消息队列中;

从所述消息队列中,并行读取所述分组数据,以及将读取的所述分组数据,并行传输至第二数据库中。

17. 根据权利要求10-15任一项所述的装置,还包括:

重传模块,用于若所述分组数据传输至所述第二数据库时发生异常,响应于对所述异常的分组数据的重传指令,在所述第二数据库中删除所述异常的分组数据,并在删除所述异常的分组数据之后,重传所述异常的分组数据。

18. 根据权利要求17所述的装置,其中,所述重传模块具体用于:

采用所述第二数据库的DSL,在所述第二数据库中删除所述异常的分组数据。

19. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-9中任一项所述的方法。

20. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-9中任一项所述的方法。

21. 一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-9中任一项所述的方法。

数据库同步方法、装置、设备和存储介质

技术领域

[0001] 本公开涉及计算机技术领域,具体涉及大数据、云计算等人工智能领域,尤其涉及一种数据库同步方法、装置、设备和存储介质。

背景技术

[0002] 数据库同步技术是让两个或多个数据库内容保持一致,或者按需要部分保持一致。

[0003] 相关技术中,可以采用日志分析的方式,进行数据库同步。

发明内容

[0004] 本公开提供了一种数据库同步方法、装置、设备和存储介质。

[0005] 根据本公开的一方面,提供了一种数据库同步方法,包括:确定第一数据库的数据中的待处理数据,所述待处理数据的数据内容为至少两种;基于所述待处理数据,将所述第一数据库的数据,分为至少两组的分组数据,其中,若所述待处理数据包括相同数据内容的数据,相同数据内容的数据被分在同一组;将所述分组数据,并行传输至第二数据库中。

[0006] 根据本公开的另一方面,提供了一种数据库同步装置,包括:确定模块,用于确定第一数据库的数据中的待处理数据,所述待处理数据的数据内容为至少两种;分组模块,用于基于所述待处理数据,将所述第一数据库的数据,分为至少两组的分组数据,其中,若所述待处理数据包括相同数据内容的数据,相同数据内容的数据被分在同一组;传输模块,用于将所述分组数据,并行传输至第二数据库中。

[0007] 根据本公开的另一方面,提供了一种电子设备,包括:至少一个处理器;以及与所述至少一个处理器通信连接的存储器;其中,所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如上述任一方面的任一项所述的方法。

[0008] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据上述任一方面的任一项所述的方法。

[0009] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据上述任一方面的任一项所述的方法。

[0010] 根据本公开的技术方案,可以提高数据库的同步效率。

[0011] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0012] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0013] 图1是根据本公开第一实施例的示意图;

- [0014] 图2是根据本公开第二实施例的示意图；
[0015] 图3是根据本公开第三实施例的示意图；
[0016] 图4是根据本公开第四实施例的示意图；
[0017] 图5是根据本公开第五实施例的示意图；
[0018] 图6是根据本公开第六实施例的示意图；
[0019] 图7是根据本公开第七实施例的示意图；
[0020] 图8是根据本公开第八实施例的示意图；
[0021] 图9是根据本公开第九实施例的示意图；
[0022] 图10是用来实现本公开实施例的数据库同步方法中任一方法的电子设备的示意图。

具体实施方式

[0023] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0024] 相关技术中,数据库同步时可以基于日志分析实现。然而,基于日志分析的方式,需要数据库类型和数据类型一致,以及,还需要串行执行,从而同步效率较差。

[0025] 为了提高同步性能,本公开提供如下实施例。

[0026] 图1是根据本公开第一实施例的示意图,本实施例提供一种数据库同步方法,该方法包括:

[0027] 101、确定第一数据库的数据中的待处理数据,所述待处理数据的数据内容为至少两种。

[0028] 102、基于所述待处理数据,将所述第一数据库的数据,分为至少两组的分组数据,其中,若所述待处理数据包括相同数据内容的数据,相同数据内容的数据被分在同一组。

[0029] 103、将所述分组数据,并行传输至第二数据库中。

[0030] 其中,第一数据库和第二数据库是待同步的两个数据库,第一数据库中存储数据,第二数据库可以为空,以便将第一数据库中的数据同步至第二数据库。

[0031] 本实施例提供的数据库同步方法的应用环境可以如图2所示,其中,该应用环境可以包括:第一数据库201、第二数据库202和控制装置203。

[0032] 控制装置可以为软件、硬件或者软硬结合。

[0033] 第一数据库、第二数据库和控制装置可以位于相同或不同的电子设备上,电子设备比如为服务器。

[0034] 控制装置可以接收到用户的指令后发起同步操作,或者,还可以预设触发周期,按照触发周期,周期性地发起同步操作。

[0035] 控制装置发起同步操作时,可以包括:若第一数据库具有同步能力,比如,具有分组能力,则控制装置向第一数据库中发送同步指令,第一数据库接收到同步指令后,执行上述的数据库同步方法。或者,

[0036] 若第一数据库不具有同步能力,控制装置可以向第一数据库发送读取指令,将第

一数据库中的数据读取到控制装置中,由控制装置执行上述的数据库同步方法。

[0037] 因此,本实施例的执行主体可以称为数据库同步装置,该装置可以位于第一数据库内,或者位于控制装置内。

[0038] 数据库中包括多个数据,不同数据的数据内容可以相同或不同,比如,数据库中的一个数据为“A市”,另一个数据为“B市”,则这两个数据的数据内容不同。又比如,数据库中的一个数据为“A市”,另一个数据也为“A市”,则这两个数据的数据内容相同。

[0039] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0040] 待处理数据是指具有至少两种数据内容的数据,比如,待处理数据包括:“A市”和“B市”。

[0041] 确定待处理数据后,可以基于待处理数据,对第一数据库中的数据进行分组。一般来讲,待处理数据为第一数据库中的部分数据,因此,也会对待处理数据进行分组。

[0042] 如待处理数据中包括相同数据内容的数据,则相同数据内容的数据被分在同一个分组。比如,待处理数据包括:“A市”、“B市”和“A市”,则两个“A市”被分到同一个分组。

[0043] 获得至少两组的分组数据后,比如,一个分组数据包括:“A市”和“B市”,另一组分组数据包括“C市”,则多个分组数据采用并行传输方式,传输至第二数据库中,即“A市”和“B市”,与“C市”是并行传输的。

[0044] 本实施例中,通过将分组数据并行传输至第二数据库中,由于采用了并行传输方式,可以提高数据库同步效率。进一步地,将相同数据内容的数据被分在同一组,可以保证相同数据内容的数据的处理一致性,保证数据库同步的准确性。

[0045] 一些实施例中,所述第一数据库中的数据以数据表形式存储,所述数据表包括数据列,所述确定第一数据库的数据中的待处理数据,包括:获取所述数据表中的所述数据列包括的不同数据内容的数量;将所述不同数据内容的数量最大的数据列中的数据,作为所述待处理数据。

[0046] 一般来讲,数据库中的数据是指数据表形式存储的。数据表包括数据行和数据列。

[0047] 如图3所示,假设第一数据库中的数据表包括4个数据列,分别用col_a、col_b、col_c、col_d表示。基于图3的示例,则可以获得这4个数据列中每个数据列包括的不同数据内容的数量。

[0048] 具体地,基于图3的示例,4个数据列包括的不同数据内容的数量分别为1、2、3、4。

[0049] 假设不同数据内容的数据最大的数据列称为当前列,则如图3所示,当前列为第4列。

[0050] 本实施例中,将不同数据内容的数量最大的数据列中的数据作为待处理数据,因此,本实施例中,待处理数据是指第4列中的数据,即,待处理数据包括:“A市”、“B市”、“C市”、“D市”。

[0051] 通过基于数据列确定待处理数据,可以符合数据库一般以数据表形式存储数据的实际情况。通过将不同数据内容的数量最大的数据列中的数据作为待处理数据,可以在后续流程选择尽量大的分组数,提高并行度。

[0052] 一些实施例中,所述获取所述数据表中的所述数据列包括的不同数据内容的数量,包括:采用所述第一数据库的领域专用语言(Domain Specified Language,DSL),获取

所述数据表中的所述数据列包括的不同数据内容的数量。

[0053] 例如,第一数据库为Mysql数据库,则可以采用如下语言获得col_a列的不同数据内容的数量:

[0054] `SELECT COUNT (DISTINCT columns)AS col FROM table_a.`

[0055] 通过采用第一数据库的DSL,不限于数据库的类型,只需要数据库对应的DSL,因此,在数据库同步时可以不限定数据库的类型。

[0056] 一些实施例中,所述不同数据内容的数量最大的数据列为当前列,所述数据表还包括数据行,所述基于所述待处理数据,将所述第一数据库中的数据,分为至少两个分组的分组数据,包括:确定所述当前列中的各个数据的分组信息;将所述当前列中的各个数据所在的数据行中的数据,分到所述各个数据的分组信息对应的分组内。

[0057] 基于图3所示的示例,假设分组信息为分组编号,分组编号包括0和1,假设“A市”和“B市”的分组信息均为分组_0,“C市”和“D市”的分组信息均为分组_1。

[0058] 则如图4所示,“A市”和“B市”所在的数据行中的数据均被分到分组_0,“C市”和“D市”所在的数据行中的数据均被分到分组_1。

[0059] 由于数据库在同步时,一般是以数据行为单位进行,比如,分别写入第一行的数据,第二行的数据等,因此,本实施例中,通过将各个数据所在的数据行中的数据分到对应数据的分组内,可以实现以数据行为单位的处理。

[0060] 一些实施例中,所述确定所述当前列中的各个数据的分组信息,包括:将所述当前列包括的不同数据内容的数量作为最大分组数;选择不大于所述最大分组数的数值,作为当前分组数;获取所述当前列中的各个数据的哈希值;基于所述哈希值和所述当前分组数,确定所述当前列中的各个数据的分组信息。

[0061] 比如,基于图3的示例,最大分组数为4,当前分组数为不大于4的数值,具体数值可以根据经验选择,比如,选择为2。

[0062] 当前列的数据包括:“A市”、“B市”、“C市”、“D市”,可以对上述数据进行哈希运算,以获得各个数据的哈希值。哈希运算比如为MD5运算。

[0063] 获得各个数据的哈希值后,可以基于哈希值和当前分组数(如,2),确定对应数据的分组信息。

[0064] 通过基于哈希值和当前分组数确定分组信息,可以保证分组信息的准确度。

[0065] 具体地,可以将可以依次计算每个数据列中的不同数据内容的数量,aColumnCount;比较aColumnCount和maxCount,替换对应的列名以及maxCount值,其中,maxCount的列名和值的初始值为第一次计算的数据列,比如,为列名为col_a以及col_a中不同数据内容的数量1。输出maxCount值最大的列名及其数量,本实施例中即clo_d和4。

[0066] 一些实施例中,所述基于所述哈希值和所述当前分组数,确定所述当前列中的各个数据的分组信息,包括:获得所述哈希值对应的十进制数值;将所述当前分组数作为模,对所述十进制数值进行求余运算,以获得余数;将所述余数作为所述各个数据的分组信息。

[0067] 其中,获得哈希值后,为了避免数据溢出,可以先对哈希值进行截断,将截断后的哈希值转换为10进制,进而进行求余运算。

[0068] 如图5所示,以哈希运算为MD5运算为例,可以经过MD5运算后,获得当前列(第四列)中各个数据的MD5值,再对MD5进行截断,比如,可以选择MD5值的最后10位,再将选择的

最后10位转换为10进制。再对10进制以2(当前分组数)为模进行求余运算,获得余数,将余数作为分组信息。

[0069] 通过求余运算,可以简便地获得各个数据对应的分组信息。

[0070] 一些实施例中,所述将所述分组数据,并行传输至第二数据库中,包括:将所述分组数据,并行传输至消息队列中;从所述消息队列中,并行读取所述分组数据,以及将读取的所述分组数据,并行传输至第二数据库中。

[0071] 如图6所示,以消息队列为kafka为例,消息队列中可以包括多个主题(topic),每个topic对应一组分组数据。基于上述示例,消息队列可以包括topic_0和topic_1,topic_0用于存储一组分组数据,即第一行和第二行的数据,topic_1用于存储另一组分组数据,即第三行和第四行的数据。

[0072] topic_0中的数据和topic_1中的数据为并行传输的。

[0073] 进一步地,每个主题内的数据可以采用多种传输实例,因此,在一个传输实例出现异常时,可以由其他传输实例继续进行对应主题内的数据的传输,保证传输的正常进行。

[0074] 如图7所示,以消息队列到第二数据库的传输过程为例,不同的topic内的数据可以采用不同的传输实例执行,通过不同的传输实例可以将不同的topic内的数据,并行传输至第二数据库中。

[0075] 并行传输还可以称为分布式传输,通过分布式传输,可以将传输的压力均衡到不同的传输实例上;数据进行分组传输,比起单点传输的速度更快;分布式部署之后,整个传输系统的容错能力更强,当某个实例发生异常时,只要有其他备份的实例就可以正常的传输。

[0076] 通过消息队列,可以将第一数据库中的分组数据传输至第二数据库中。进一步地,若第三数据库也需要同步第一数据库中的数据,则第三数据库可以直接从消息队列中读取分组数据,相对于数据库直接通信的方式,在多个数据库间进行同步时,可以节省资源。

[0077] 一些实施例中,所述方法还包括:若所述分组数据传输至所述第二数据库时发生异常,响应于对所述异常的分组数据的重传指令,在所述第二数据库中删除所述异常的分组数据,并在删除所述异常的分组数据之后,重传所述异常的分组数据。

[0078] 数据库同步过程中,若发生数据异常,相关技术中,一般采用新增数据的方式。比如,同步第一组分组数据(第一行数据和第二行数据)时,假设第一行数据同步完成,第二行数据未同步完成时,发生异常,此时,第二数据库中已经存在第一行数据,相关技术中,在异常消除后,会重新同步该第一组分组数据,即,第二数据库中会再增加第一行数据和第二行数据,由于第一行数据已经存在,那么就存在重复数据,造成资源浪费。

[0079] 为此,本实施例中,在重新同步分组数据之前,先删除之前已经同步的数据,比如,先删除第一行数据,之后再重新同步第一行数据和第二行数据。

[0080] 通过在重新分组数据之前先删除已有的数据,可以实现数据回滚,避免第二数据库中出现重复数据,避免资源浪费。

[0081] 一些实施例中,所述在所述第二数据库中删除所述异常的分组数据,包括:采用所述第二数据库的DSL,在所述第二数据库中删除所述异常的分组数据。

[0082] 比如,第二数据库为Mysql数据库时,删除topic_0的分组数据的语句可以为:

[0083] DELETE FROM table_name WHERE MOD(CONV(SUBSTRING(MD5('maxCol'),-10,

10), 16, 10), 2) = 0;

[0084] 通过采用第二数据库的DSL, 不限于数据库的类型, 只需要数据库对应的DSL, 因此, 在数据库同步时可以不限定数据库的类型。

[0085] 上述分别描述了方法的相关步骤, 下面以分组数据通过消息队列传输为例进行说明。

[0086] 图8是根据本公开第八实施例的示意图, 本实施例提供一种数据库同步方法, 该方法包括:

[0087] 801、确定第一数据库的数据中的待处理数据, 所述待处理数据的数据内容为至少两种。

[0088] 802、基于所述待处理数据, 将所述第一数据库的数据, 分为至少两组的分组数据, 其中, 若所述待处理数据包括相同数据内容的数据, 相同数据内容的数据被分在同一组。

[0089] 803、将所述分组数据, 并行传输至消息队列中。

[0090] 804、从所述消息队列中, 并行读取所述分组数据, 以及将读取的所述分组数据, 并行传输至第二数据库中。

[0091] 具体内容可以参见上述实施例中的相关描述。

[0092] 本公开实施例中, 不限于数据库类型, 同步的数据库可以为Mysql、ES、MongoDB等, 在同步过程中, 可以采用数据库的DSL执行相关操作。如果第一数据库不支持分组, 还可以将第一数据库中的数据读取在控制装置内执行分组, 实现过程灵活。通过并行传输分组数据, 可以实现分布式部署, 提高整体同步架构的容错能力和性能。通过在重传分组数据之前执行删除操作, 可以实现数据回滚, 确保事务性传输, 确保传输数据的准确性。

[0093] 图9是根据本公开第九实施例的示意图, 本实施例提供一种数据库同步装置, 该装置900包括: 确定模块901、分组模块902和传输模块903。

[0094] 确定模块901用于确定第一数据库的数据中的待处理数据, 所述待处理数据的数据内容为至少两种; 分组模块902用于基于所述待处理数据, 将所述第一数据库的数据, 分为至少两组的分组数据, 其中, 若所述待处理数据包括相同数据内容的数据, 相同数据内容的数据被分在同一组; 传输模块903用于将所述分组数据, 并行传输至第二数据库中。

[0095] 一些实施例中, 所述第一数据库中的数据以数据表形式存储, 所述数据表包括数据列, 所述确定模块901具体用于: 获取所述数据表中的所述数据列包括的不同数据内容的数量; 将所述不同数据内容的数量最大的数据列中的数据, 作为所述待处理数据。

[0096] 一些实施例中, 所述确定模块901进一步具体用于: 采用所述第一数据库的DSL, 获取所述数据表中的所述数据列包括的不同数据内容的数量。

[0097] 一些实施例中, 所述不同数据内容的数量最大的数据列为当前列, 所述数据表还包括数据行, 所述分组模块902包括: 确定单元, 用于确定所述当前列中的各个数据的分组信息; 分组单元, 用于将所述当前列中的各个数据所在的数据行中的数据, 分到所述各个数据的分组信息对应的分组内。

[0098] 一些实施例中, 所述确定单元具体用于: 将所述当前列包括的不同数据内容的数量作为最大分组数; 选择不大于所述最大分组数的数值, 作为当前分组数; 获取所述当前列中的各个数据的哈希值; 基于所述哈希值和所述当前分组数, 确定所述当前列中的各个数据的分组信息。

[0099] 一些实施例中,所述确定单元进一步具体用于:获得所述哈希值对应的十进制数值;将所述当前分组数作为模,对所述十进制数值进行求余运算,以获得余数;将所述余数作为所述各个数据的分组信息。

[0100] 一些实施例中,所述传输模块903具体用于:将所述分组数据,并行传输至消息队列中;从所述消息队列中,并行读取所述分组数据,以及将读取的所述分组数据,并行传输至第二数据库中。

[0101] 一些实施例中,所述装置900还包括:重传模块,用于若所述分组数据传输至所述第二数据库时发生异常,响应于对所述异常的分组数据的重传指令,在所述第二数据库中删除所述异常的分组数据,并在删除所述异常的分组数据之后,重传所述异常的分组数据。

[0102] 一些实施例中,所述重传模块具体用于:采用所述第二数据库的DSL,在所述第二数据库中删除所述异常的分组数据。

[0103] 本实施例中,通过将分组数据并行传输至第二数据库中,由于采用了并行传输方式,可以提高数据库同步效率。进一步地,将相同数据内容的数据被分在同一组,可以保证相同数据内容的数据的处理一致性,保证数据库同步的准确性。

[0104] 可以理解的是,本公开实施例中,不同实施例中的相同或相似内容可以相互参考。

[0105] 可以理解的是,本公开实施例中的“第一”、“第二”等只是用于区分,不表示重要程度高低、时序先后等。

[0106] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0107] 图10示出了可以用来实施本公开的实施例的示例电子设备1000的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字助理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0108] 如图10所示,电子设备1000包括计算单元1001,其可以根据存储在只读存储器(ROM) 1002中的计算机程序或者从存储单元10010加载到随机访问存储器(RAM) 1003中的计算机程序,来执行各种适当的动作和处理。在RAM 1003中,还可存储电子设备1000操作所需的各种程序和数据。计算单元1001、ROM 1002以及RAM 1003通过总线1004彼此相连。输入/输出(I/O)接口1005也连接至总线1004。

[0109] 电子设备1000中的多个部件连接至I/O接口1005,包括:输入单元1006,例如键盘、鼠标等;输出单元1007,例如各种类型的显示器、扬声器等;存储单元1008,例如磁盘、光盘等;以及通信单元1009,例如网卡、调制解调器、无线通信收发机等。通信单元1009允许电子设备1000通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0110] 计算单元1001可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元1001的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元1001执行上文所描述的各个方法和处理,例如数据库同步方法。例如,在一些实施例中,数据库同步方法可被实现为计算机软

件程序,其被有形地包含于机器可读介质,例如存储单元1008。在一些实施例中,计算机程序的部分或者全部可以经由ROM 1002和/或通信单元1009而被载入和/或安装到电子设备1000上。当计算机程序加载到RAM 1003并由计算单元1001执行时,可以执行上文描述的数据库同步方法的一个或多个步骤。备选地,在其他实施例中,计算单元1001可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行数据库同步方法。

[0111] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0112] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0113] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0114] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0115] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)和互联网。

[0116] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务(“Virtual Private Server”,或简称“VPS”)中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0117] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0118] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

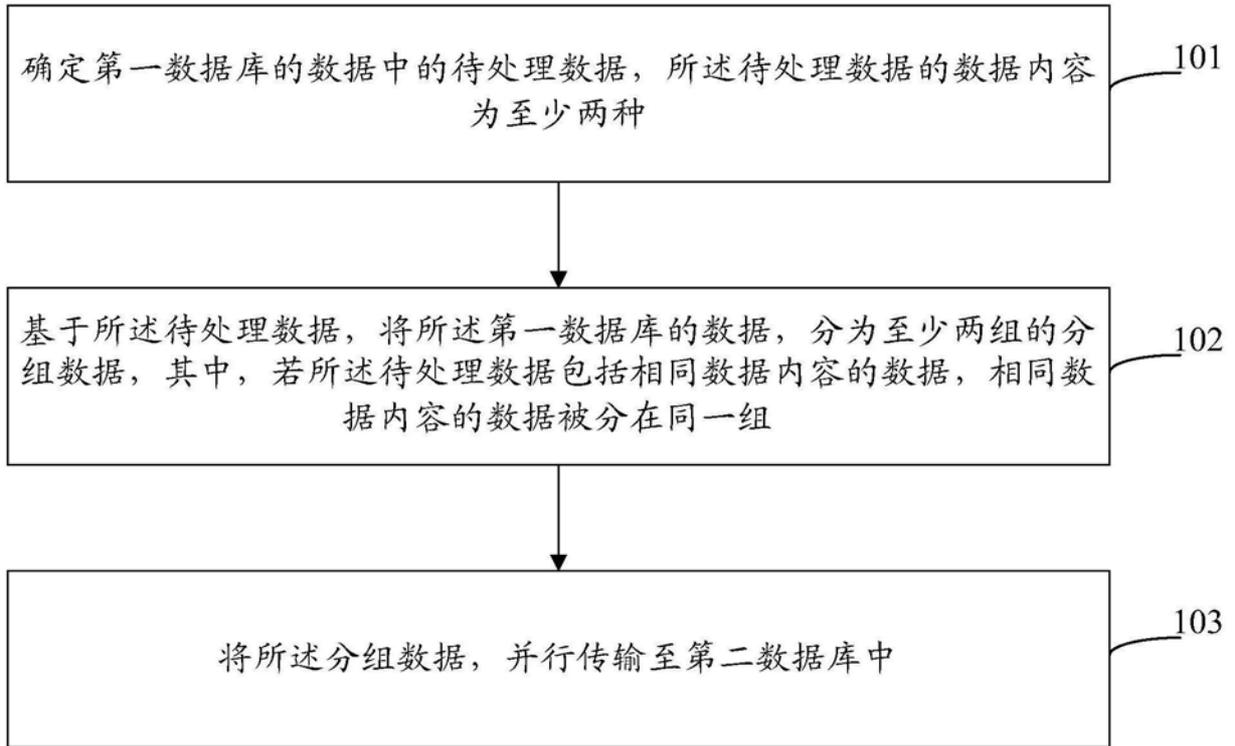


图1

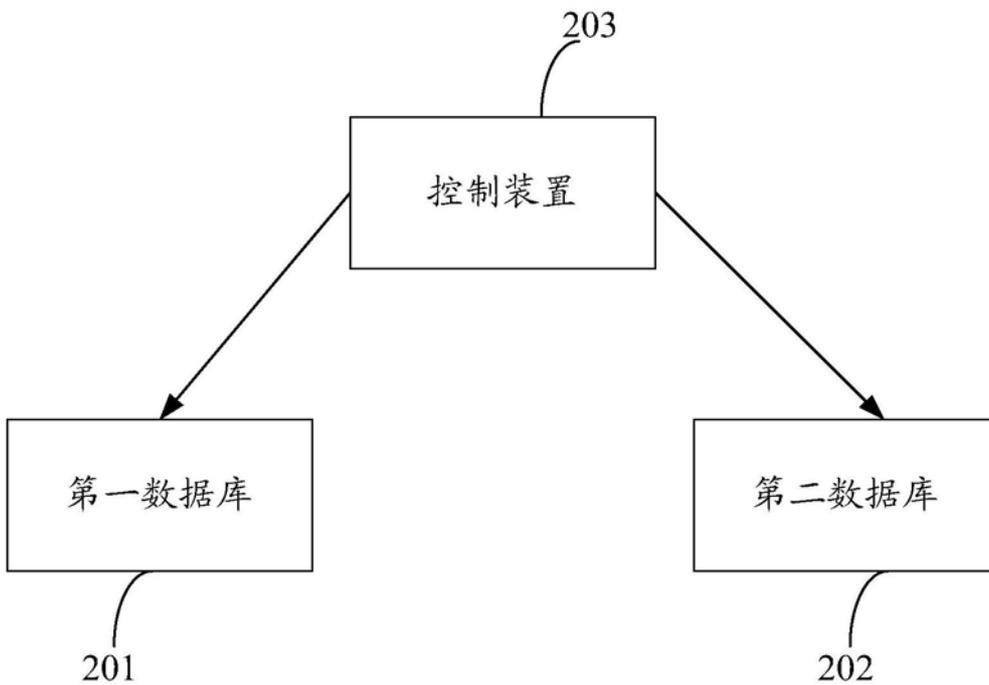


图2

A table with 5 rows and 4 columns. The columns are labeled col_a, col_b, col_c, and col_d. The rows contain the following data: (A市, A市, A市, A市), (A市, A市, B市, B市), (A市, B市, C市, C市), and (A市, B市, C市, D市). A dashed box encloses the entire fourth column (col_d), with an arrow pointing to it from the label '当前列' (Current Column) above.

col_a	col_b	col_c	col_d
A市	A市	A市	A市
A市	A市	B市	B市
A市	B市	C市	C市
A市	B市	C市	D市

图3

A table with 4 rows and 4 columns. The columns are labeled col_a, col_b, col_c, and col_d. The rows contain the following data: (A市, A市, A市, A市), (A市, A市, B市, B市), (A市, B市, C市, C市), and (A市, B市, C市, D市). A dashed box encloses the first two rows, with an arrow pointing to it from the label '分组_0' (Group 0) to its right. Another dashed box encloses the last two rows, with an arrow pointing to it from the label '分组_1' (Group 1) to its right.

col_a	col_b	col_c	col_d
A市	A市	A市	A市
A市	A市	B市	B市
A市	B市	C市	C市
A市	B市	C市	D市

图4

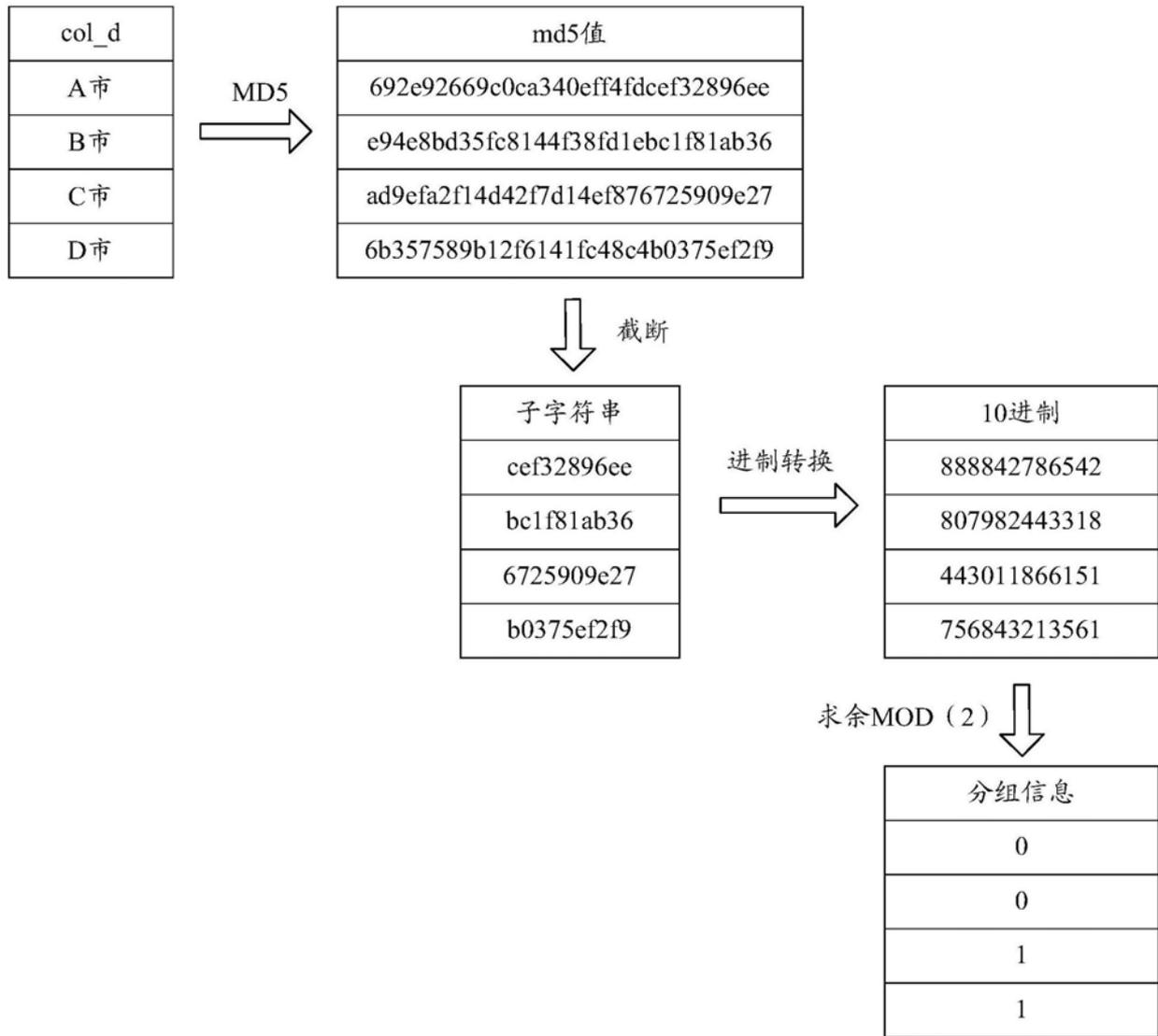


图5

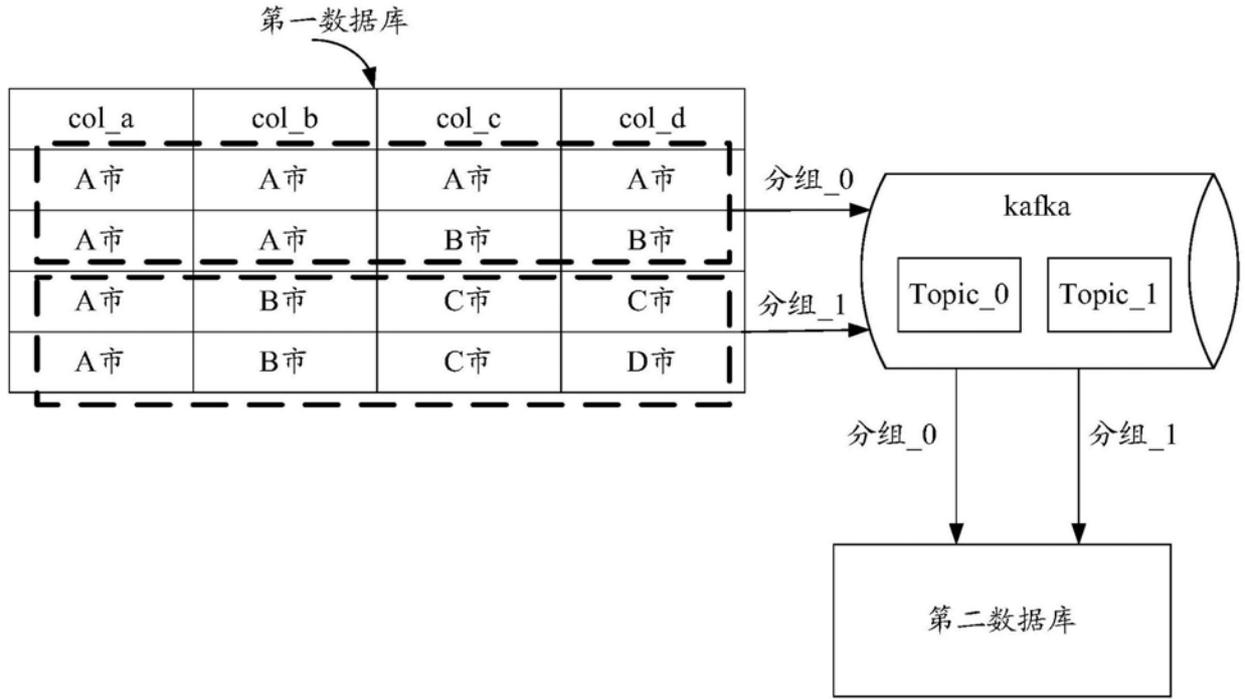


图6

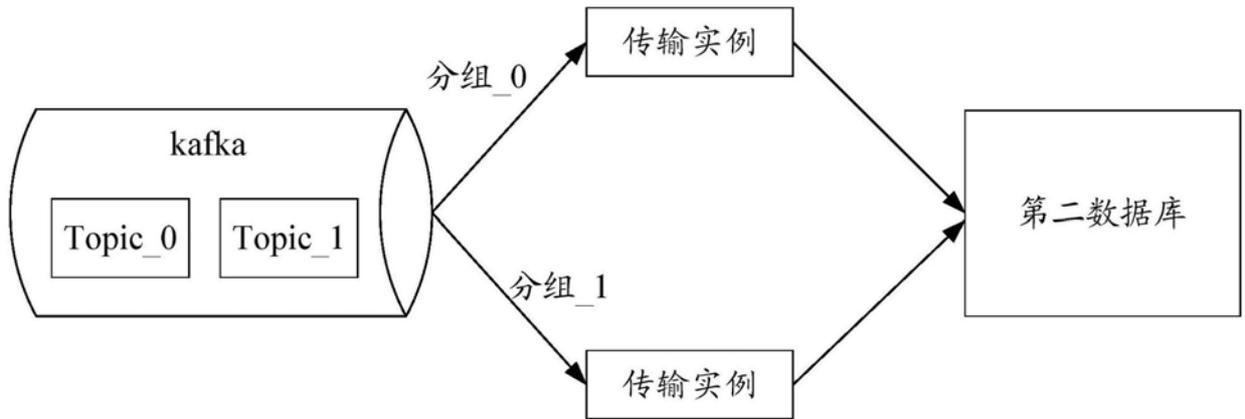


图7

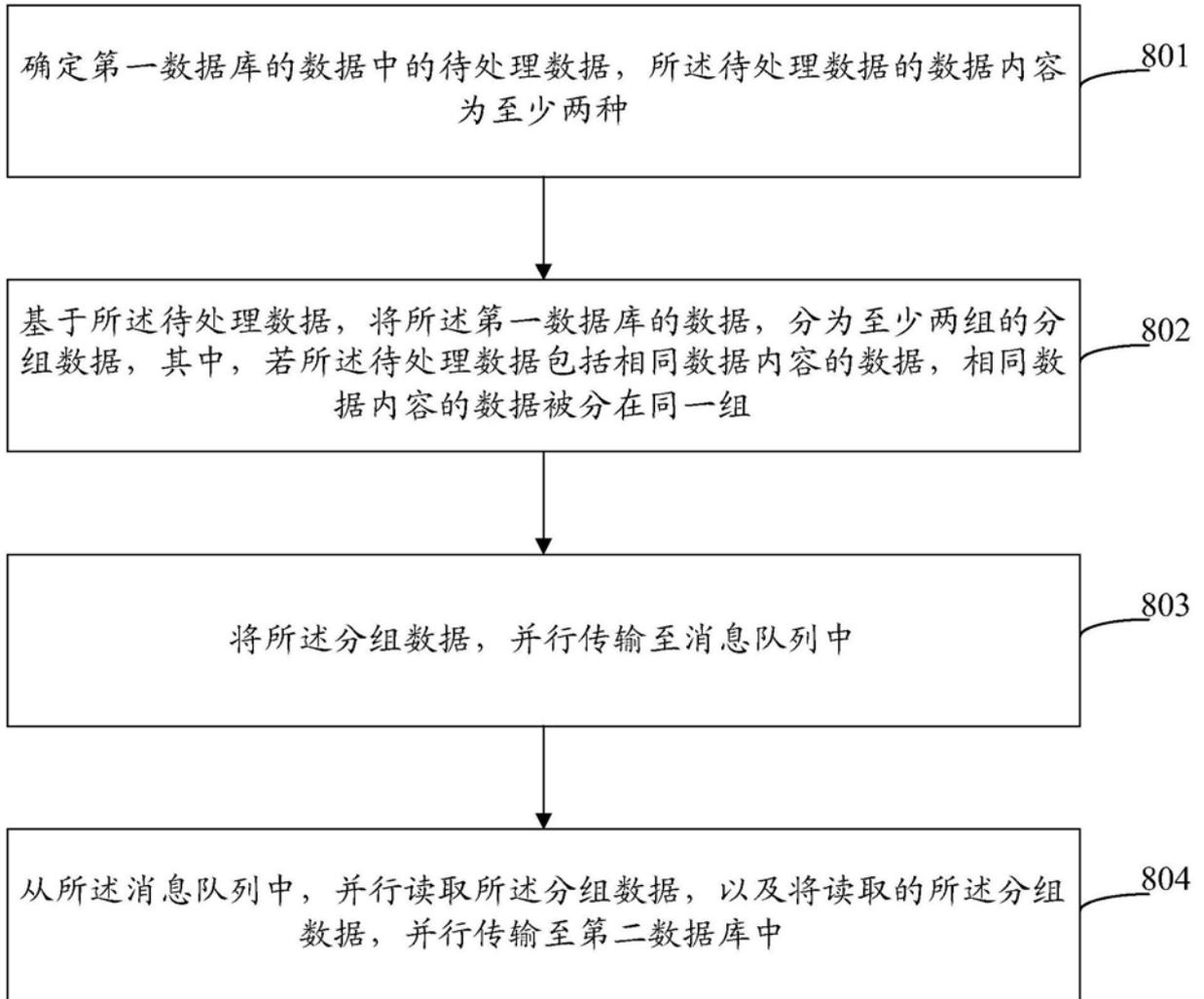


图8

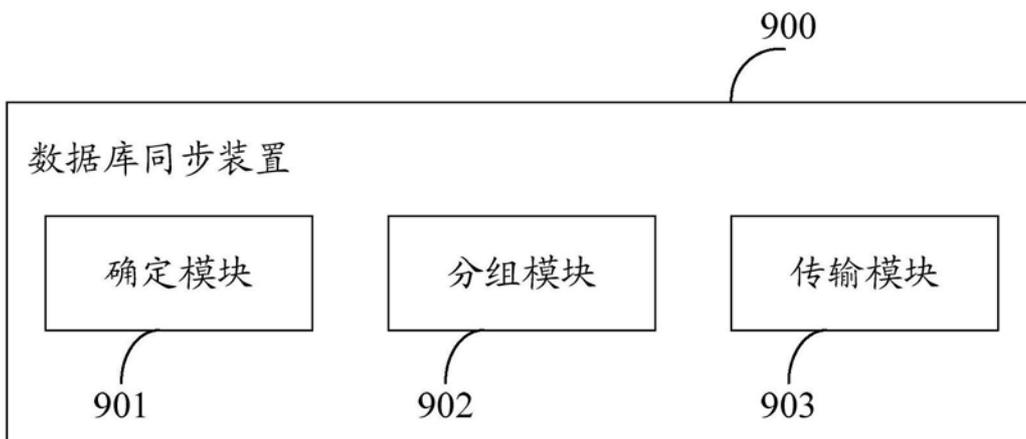


图9

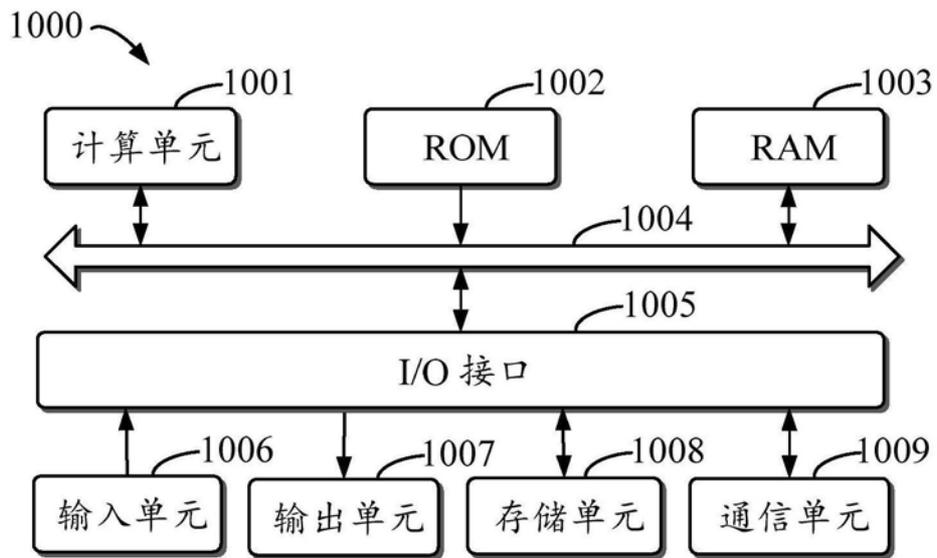


图10