



(12) 发明专利

(10) 授权公告号 CN 109829016 B

(45) 授权公告日 2021.06.01

(21) 申请号 201910071849.0

(22) 申请日 2019.01.25

(65) 同一申请的已公布的文献号
申请公布号 CN 109829016 A

(43) 申请公布日 2019.05.31

(73) 专利权人 网宿科技股份有限公司
地址 200030 上海市嘉定区环城路200号

(72) 发明人 汪建福

(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 黄志华

(51) Int. Cl.

G06F 16/27 (2019.01)

G06F 16/242 (2019.01)

G06F 16/2458 (2019.01)

(56) 对比文件

CN 101369878 A, 2009.02.18

CN 105450720 A, 2016.03.30

CN 106649632 A, 2017.05.10

CN 109033391 A, 2018.12.18

审查员 徐彬

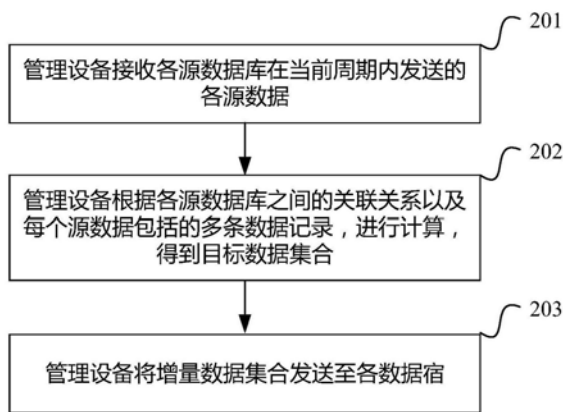
权利要求书3页 说明书11页 附图3页

(54) 发明名称

一种数据同步方法及装置

(57) 摘要

本发明公开了一种数据同步方法及装置。所述方法包括:管理设备接收各源数据库在当前周期内发送的各源数据后,根据各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集并将目标数据集合发送至各数据宿。本发明实施例中,只需要管理设备执行这一计算过程,在同步的过程进行目标数据集合的计算,使得数据宿能够直接获取目标数据集合,大大降低了整个系统的计算量,提升了数据同步的时效性。



1. 一种数据同步方法,其特征在于,所述方法应用于数据同步系统,所述系统包括管理设备、多个源数据库和至少一个数据宿;所述方法包括:

所述管理设备接收各源数据库在当前周期内发送的各源数据;所述各源数据为所述各源数据库在所述当前周期内对应的全量数据;

所述管理设备判断所述各源数据是否全部发送完成,若未全部发送完成,则判断各源数据中是否存在数据发送耗时大于预设阈值的目标源数据,若存在,则重新接收所述目标源数据,若不存在,则重复执行判断所述各源数据是否全部发送完成的操作,直至确定各源数据全部发送完成;

所述管理设备根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集合;所述目标数据集合中的一条数据记录在一个或多个源数据库的源数据中具有同一索引值;

所述管理设备根据各数据宿的数据需求,从所述目标数据集合中确定符合所述各数据宿中的数据需求的各目标数据,将所述各目标数据对应发送至所述各数据宿,所述各数据宿用于接收数据。

2. 根据权利要求1所述的方法,其特征在于,所述管理设备根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集合,包括:

所述管理设备根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定所述当前周期内的第一数据集合;

所述管理设备若确定存在所述当前周期的上一周期对应的第二数据集合,则确定所述第一数据集合相较于所述第二数据集合的增量数据集合,并将所述增量数据集合作为所述目标数据集合;所述第二数据集合是根据所述各源数据库在所述上一周期内发送的各源数据得到的。

3. 根据权利要求1所述的方法,其特征在于,每条数据记录包括索引值;

所述各源数据库之间的关联关系通过以下方式确定:

针对第一源数据中的第一数据记录,所述管理设备根据所述第一数据记录对应的索引值,若确定第二源数据中存在索引值与所述第一数据记录对应的索引值一致的第二数据记录,则根据所述第一数据记录在所述第一数据源中的位置和所述第二数据记录在所述第二数据源中的位置,确定所述第一数据记录与所述第二数据记录的关联关系;所述第一源数据为所述各源数据中任一源数据,所述第一数据记录为所述第一源数据包括的多条数据记录中任一数据记录;

根据所述每个源数据包括的多条数据记录之间的关联关系,确定所述各源数据库之间的关联关系。

4. 根据权利要求2所述的方法,其特征在于,所述各源数据库之间为异构数据库,所述异构数据库为数据库之间的结构不同的数据库;

所述各源数据的关联关系为基于结构化查询语言SQL的关联关系。

5. 根据权利要求2所述的方法,其特征在于,还包括:

所述管理设备若确定不存在所述第二数据集合,则将所述第一数据集合作为所述目标数据集合。

6. 根据权利要求1至5中任一项所述的方法,其特征在于,所述管理设备将所述目标数

据集合发送至各数据宿之后,还包括:

所述管理设备获取所述各数据宿的数据同步信息;所述数据同步信息包括数据增加记录、数据删除记录和数据修改记录中的至少一项;

所述管理设备根据所述各数据宿的数据同步信息,对所述各数据宿的同步情况进行监控。

7. 根据权利要求1至5中任一项所述的方法,其特征在于,所述管理设备接收各源数据库在当前周期内发送的源数据之后,还包括:

所述管理设备获取所述各源数据库的数据发送情况,所述数据发送情况是根据所述各源数据的数据发送耗时和/或数据发送量确定的;

所述管理设备根据所述各源数据库的数据发送情况,对所述各源数据库的同步情况进行监控。

8. 根据权利要求1至5中任一项所述的方法,其特征在于,所述管理设备将所述目标数据集合发送至各数据宿之后,还包括:

所述管理设备若确定所述目标数据集合发送失败,且发送失败的次数大于预设阈值,则向管理员发送告警消息,所述告警消息用于通知所述管理员所述目标数据集合发送失败。

9. 一种数据同步装置,其特征在于,所述装置应用于数据同步系统,所述系统包括管理设备、多个源数据库和至少一个数据宿;所述装置为管理设备,所述装置包括:

接收单元,用于接收各源数据库在当前周期内发送的各源数据;所述各源数据为所述各源数据库在所述当前周期内对应的全量数据;

处理单元,用于判断所述各源数据是否全部发送完成,若未全部发送完成,则判断各源数据中是否存在数据发送耗时大于预设阈值的目标源数据,若存在,则重新接收所述目标源数据,若不存在,则重复执行判断所述各源数据是否全部发送完成的操作,直至确定各源数据全部发送完成;在确定各源数据全部发送完成后,根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集合;所述目标数据集合中的一条数据记录在一个或多个源数据库的源数据中具有同一索引值;

发送单元,用于根据所述各数据宿的数据需求,从所述目标数据集合中确定符合所述各数据宿的数据需求的各目标数据;以及将所述各目标数据对应发送至所述各数据宿,所述各数据宿用于接收数据。

10. 根据权利要求9所述的装置,其特征在于,所述处理单元具体用于:

根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定所述当前周期内的第一数据集合;以及若确定存在所述当前周期的上一周期对应的第二数据集合,则确定所述第一数据集合相较于所述第二数据集合的增量数据集合,并将所述增量数据集合作为所述目标数据集合;所述第二数据集合是根据所述各源数据库在所述上一周期内发送的各源数据得到的。

11. 根据权利要求9所述的装置,其特征在于,每条数据记录包括索引值;

所述各源数据库之间的关联关系通过以下方式确定:

针对第一源数据中的第一数据记录,所述管理设备根据所述第一数据记录对应的索引值,若确定第二源数据中存在索引值与所述第一数据记录对应的索引值一致的第二数据记

录,则根据所述第一数据记录在所述第一数据源中的位置和所述第二数据记录在所述第二数据源中的位置,确定所述第一数据记录与所述第二数据记录的关联关系;所述第一源数据为所述各源数据中任一源数据,所述第一数据记录为所述第一源数据包括的多条数据记录中任一数据记录;

根据所述每个源数据包括的多条数据记录之间的关联关系,确定所述各源数据库之间的关联关系。

12. 根据权利要求10所述的装置,其特征在于,所述各源数据库之间为异构数据库,所述异构数据库为数据库之间的结构不同的数据库;

所述各源数据的关联关系为基于结构化查询语言SQL的关联关系。

13. 根据权利要求10所述的装置,其特征在于,所述处理单元还用于:

若确定不存在所述第二数据集合,则将所述第一数据集合作为所述目标数据集合。

14. 根据权利要求9至13中任一项所述的装置,其特征在于,所述发送单元将所述目标数据集合发送至各数据宿之后,所述处理单元还用于:

获取所述各数据宿的数据同步信息;所述数据同步信息包括数据增加记录、数据删除记录和数据修改记录中的至少一项;以及根据所述各数据宿的数据同步信息,对所述各数据宿的同步情况进行监控。

15. 根据权利要求9至13中任一项所述的装置,其特征在于,所述接收单元接收各源数据库在当前周期内发送的源数据之后,所述处理单元还用于:

获取所述各源数据库的数据发送情况,所述数据发送情况是根据所述各源数据的数据发送耗时和/或数据发送量确定的;以及根据所述各源数据库的数据发送情况,对所述各源数据库的同步情况进行监控。

16. 根据权利要求9至13中任一项所述的装置,其特征在于,所述发送单元将所述目标数据集合发送至各数据宿之后,所述处理单元还用于:

若确定所述目标数据集合发送失败,且发送失败的次数大于预设阈值,则向管理员发送告警消息,所述告警消息用于通知所述管理员所述目标数据集合发送失败。

17. 一种计算机可读存储介质,其特征在于,所述存储介质存储有指令,当所述指令在计算机上运行时,使得计算机实现执行权利要求1至8中任一项所述的方法。

18. 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如权利要求1至8中任一权利要求所述的方法。

一种数据同步方法及装置

技术领域

[0001] 本发明涉及数据处理技术领域,尤其涉及一种数据同步方法及装置。

背景技术

[0002] 数据同步一直是计算机系统之间进行交互的不可避免的问题,数据同步的时效性和准确性将直接影响计算机系统之间的交互。现有的数据同步技术,通常是先将各源数据库中的各源数据分别同步到数据宿,然后由数据宿进行计算,从而获取所需的数据。然而,这种方法对于每个数据宿而言,都需要进行这样的计算过程,过程比较复杂,比较浪费时间。

[0003] 基于此,目前亟需一种数据同步方法,用于解决现有技术中的数据同步过程比较复杂,耗时较长的问题。

发明内容

[0004] 本发明实施例提供一种数据同步方法及装置,以解决现有技术中数据同步过程比较复杂,耗时较长的技术问题。

[0005] 本发明实施例提供一种数据同步方法,所述方法应用于数据同步系统,所述系统包括管理设备、多个源数据库和至少一个数据宿;所述方法包括:

[0006] 所述管理设备接收各源数据库在当前周期内发送的各源数据;所述各源数据为所述各源数据库在所述当前周期内对应的全量数据;

[0007] 所述管理设备根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集合;

[0008] 所述管理设备将所述目标数据集合发送至各数据宿,所述数据宿用于接收数据。

[0009] 如此,本发明实施例设置管理设备,通过管理设备对各源数据库的各源数据进行计算,从而在管理设备层面确定目标数据集合,进而可以将目标增量数据集合发送给各数据宿。相比于现有技术中每个数据宿都需要分别计算目标数据集合的方式而言,本发明实施例中,只需要管理设备执行这一计算过程,在同步的过程进行目标数据集合的计算,使得数据宿能够直接获取目标数据集合,大大降低了整个系统的计算量,提升了数据同步的时效性;进一步地,采用管理设备向个数据宿发送目标数据集合,能够将一份目标数据集合同时发送到多个数据宿,从而能够提升数据同步的效率。

[0010] 在一种可能的实现方式中,所述管理设备根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集合,包括:

[0011] 所述管理设备根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定所述当前周期内的第一数据集合;

[0012] 所述管理设备若确定存在所述当前周期的上一周期对应的第二数据集合,则确定所述第一数据集合相较于所述第二数据集合的增量数据集合,并将所述增量数据集合作为所述目标数据集合;所述第二数据集合是根据所述各源数据库在所述上一周期内发送的各

源数据得到的。

[0013] 在一种可能的实现方式中,每条数据记录包括索引值;

[0014] 所述各源数据库之间的关联关系通过以下方式确定:

[0015] 针对第一源数据中的第一数据记录,所述管理设备根据所述第一数据记录对应的索引值,若确定第二源数据中存在索引值与所述第一数据记录对应的索引值一致的第二数据记录,则根据所述第一数据记录在所述第一数据源中的位置和所述第二数据记录在所述第二数据源中的位置,确定所述第一数据记录与所述第二数据记录的关联关系;所述第一源数据为所述各源数据中任一源数据,所述第一数据记录为所述第一源数据包括的多条数据记录中任一数据记录;

[0016] 根据所述每个源数据包括的多条数据记录之间的关联关系,确定所述各源数据库之间的关联关系。

[0017] 在一种可能的实现方式中,所述各源数据库之间为异构数据库,所述异构数据库为数据库之间的结构不同的数据库;

[0018] 所述各源数据的关联关系为基于结构化查询语言SQL的关联关系。

[0019] 如此,采用通用的SQL语言,能够降低各异构源数据计算的难度,方便各异构源数据的逻辑计算,也便于数据的维护。

[0020] 在一种可能的实现方式中,所述管理设备将所述目标数据集合发送至各数据宿,包括:

[0021] 所述管理设备根据所述各数据宿的数据需求,从所述目标数据集合中确定符合所述各数据宿的数据需求的各目标数据;

[0022] 所述管理设备将所述各目标数据对应发送至所述各数据宿。

[0023] 在一种可能的实现方式中,还包括:

[0024] 所述管理设备若确定不存在所述第二数据集合,则将所述第一数据集合作为所述目标数据集合。

[0025] 在一种可能的实现方式中,所述管理设备将所述目标数据集合发送至各数据宿之后,还包括:

[0026] 所述管理设备获取所述各数据宿的数据同步信息;所述数据同步信息包括数据增加记录、数据删除记录和数据修改记录中的至少一项;

[0027] 所述管理设备根据所述各数据宿的数据同步信息,对所述各数据宿的同步情况进行监控。

[0028] 如此,可以实时监测各数据宿的同步情况,便于管理设备统计分析。

[0029] 在一种可能的实现方式中,所述管理设备接收各源数据库在当前周期内发送的源数据之后,还包括:

[0030] 所述管理设备获取所述各源数据库的数据发送情况,所述数据发送情况是根据所述各源数据的数据发送耗时和/或数据发送量确定的;

[0031] 所述管理设备根据所述各源数据库的数据发送情况,对所述各源数据库的同步情况进行监控。

[0032] 如此,可以实时监测各源数据库的同步情况,便于管理设备统计分析。

[0033] 在一种可能的实现方式中,所述管理设备将所述目标数据集合发送至各数据宿之

后,还包括:

[0034] 所述管理设备若确定所述目标数据集合发送失败,且发送失败的次数大于预设阈值,则向管理员发送告警消息,所述告警消息用于通知所述管理员所述目标数据集合发送失败。

[0035] 如此,一旦目标数据集合多次出现发送失败的情况管理设备可以及时向管理员发送告警消息,以便管理员及时进行处理。

[0036] 本发明实施例提供一种数据同步装置,所述装置应用于数据同步系统,所述系统包括管理设备、多个源数据库和至少一个数据宿;所述装置为管理设备,所述装置包括:

[0037] 接收单元,用于接收各源数据库在当前周期内发送的各源数据;所述各源数据为所述各源数据库在所述当前周期内对应的全量数据;

[0038] 处理单元,用于根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,得到目标数据集合;

[0039] 发送单元,用于将所述目标数据集合发送至各数据宿,所述数据宿用于接收数据。

[0040] 在一种可能的实现方式中,所述处理单元具体用于:

[0041] 根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定所述当前周期内的第一数据集合;以及若确定存在所述当前周期的上一周期对应的第二数据集合,则确定所述第一数据集合相较于所述第二数据集合的增量数据集合,并将所述增量数据集合作为所述目标数据集合;所述第二数据集合是根据所述各源数据库在所述上一周期内发送的各源数据得到的。

[0042] 在一种可能的实现方式中,每条数据记录包括索引值;

[0043] 所述各源数据库之间的关联关系通过以下方式确定:

[0044] 针对第一源数据中的第一数据记录,所述管理设备根据所述第一数据记录对应的索引值,若确定第二源数据中存在索引值与所述第一数据记录对应的索引值一致的第二数据记录,则根据所述第一数据记录在所述第一数据源中的位置和所述第二数据记录在所述第二数据源中的位置,确定所述第一数据记录与所述第二数据记录的关联关系;所述第一源数据为所述各源数据中任一源数据,所述第一数据记录为所述第一源数据包括的多条数据记录中任一数据记录;

[0045] 根据所述每个源数据包括的多条数据记录之间的关联关系,确定所述各源数据库之间的关联关系。

[0046] 在一种可能的实现方式中,所述各源数据库之间为异构数据库,所述异构数据库为数据库之间的结构不同的数据库;

[0047] 所述各源数据的关联关系为基于结构化查询语言SQL的关联关系。

[0048] 在一种可能的实现方式中,所述发送单元具体用于:

[0049] 根据所述各数据宿的数据需求,从所述目标数据集合中确定符合所述各数据宿的数据需求的各目标数据;以及将所述各目标数据对应发送至所述各数据宿。

[0050] 在一种可能的实现方式中,所述处理单元还用于:

[0051] 若确定不存在所述第二数据集合,则将所述第一数据集合作为所述目标数据集合。

[0052] 在一种可能的实现方式中,所述发送单元将所述增量数据集合发送至各数据宿之

后,所述处理单元还用于:

[0053] 获取所述各数据宿的数据同步信息;所述数据同步信息包括数据增加记录、数据删除记录和数据修改记录中的至少一项;以及根据所述各数据宿的数据同步信息,对所述各数据宿的同步情况进行监控。

[0054] 在一种可能的实现方式中,所述接收单元接收各源数据库在当前周期内发送的源数据之后,所述处理单元还用于:

[0055] 获取所述各源数据库的数据发送情况,所述数据发送情况是根据所述各源数据的数据发送耗时和/或数据发送量确定的;以及根据所述各源数据库的数据发送情况,对所述各源数据库的同步情况进行监控。

[0056] 在一种可能的实现方式中,所述发送单元将所述目标数据集合发送至各数据宿之后,所述处理单元还用于:

[0057] 若确定所述目标数据集合发送失败,且发送失败的次数大于预设阈值,则向管理员发送告警消息,所述告警消息用于通知所述管理员所述目标数据集合发送失败。

[0058] 本申请实施例的还提供一种装置,该装置具有实现上文所描述的数据同步方法的功能。该功能可以通过硬件执行相应的软件实现,在一种可能的设计中,该装置包括:处理器、收发器、存储器;该存储器用于存储计算机执行指令,该收发器用于实现该装置与其他通信实体进行通信,该处理器与该存储器通过该总线连接,当该装置运行时,该处理器执行该存储器存储的该计算机执行指令,以使该装置执行上文所描述的数据同步方法。

[0059] 本发明实施例还提供一种计算机存储介质,所述存储介质中存储软件程序,该软件程序在被一个或多个处理器读取并执行时实现上述各种可能的实现方式中所描述的数据同步方法。

[0060] 本发明实施例还提供一种包含指令的计算机程序产品,当其在计算机上运行时,使得计算机执行上述各种可能的实现方式中所描述的数据同步方法。

附图说明

[0061] 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。

[0062] 图1为本发明实施例适用的系统架构图;

[0063] 图2为本发明实施例提供的一种数据同步方法所对应的流程示意图;

[0064] 图3为本发明实施例涉及到的整体性的流程示意图;

[0065] 图4为本发明实施例提供的一种数据同步装置的结构示意图。

具体实施方式

[0066] 下面结合说明书附图对本申请进行具体说明,方法实施例中的具体操作方法也可以应用于装置实施例中。

[0067] 图1示例性示出了本发明实施例适用的系统架构图。如图1所示,所述系统100包括管理设备101、多个源数据库(例如图1中示出的源数据库1021、源数据库1022和源数据库1023)以及至少一个数据宿(例如图1中示出的数据宿1031、数据宿1032和数据宿1033)。其中,管理设备101与任一源数据库可以通过网络连接,例如,管理设备101可以通过网络与源

数据库1021、源数据库1022和源数据库1023中任一源数据库连接；管理设备101与任一数据宿可以通过网络连接，例如，管理设备101可以通过网络与数据宿1031、数据宿1032和数据宿1033中任一数据宿连接。

[0068] 基于图1所示的系统架构，图2示例性示出了本发明实施例提供的一种数据同步方法所对应的流程示意图，具体包括如下步骤：

[0069] 步骤201，管理设备接收各源数据库在当前周期内发送的各源数据。

[0070] 步骤202，管理设备根据各源数据库之间的关联关系以及每个源数据包括的多条数据记录，进行计算，得到目标数据集合。

[0071] 步骤203，管理设备将目标数据集合发送至各数据宿，所述数据宿用于接收数据。

[0072] 如此，本发明实施例设置管理设备，通过管理设备对各源数据库的各源数据进行计算，从而在管理设备层面确定目标数据集合，进而可以将目标数据集合发送给各数据宿。相比于现有技术中每个数据宿都需要分别计算目标数据集合的方式而言，本发明实施例中，只需要管理设备执行这一计算过程，在同步的过程进行目标数据集合的计算，使得数据宿能够直接获取目标数据集合，大大降低了整个系统的计算量，提升了数据同步的时效性；进一步地，采用管理设备向个数据宿发送目标数据集合，能够将一份目标数据集合同时发送到多个数据宿，从而能够提升数据同步的效率。

[0073] 具体来说，步骤201中，各源数据可以为各源数据库在当前周期内对应的全量数据。举个例子，如表1所示，为各源数据的一种示例。其中，源数据库A对应的源数据为源数据库A在当前周期内对应的全量数据A，源数据库B对应的源数据为源数据库B在当前周期内对应的全量数据B，源数据库C对应的源数据为源数据库C在当前周期内对应的全量数据C。

[0074] 表1：各源数据的一种示例

源数据库	源数据
源数据库A	全量数据A
源数据库B	全量数据B
源数据库C	全量数据C

[0076] 进一步地，每个源数据可以包括多条数据记录。如表2所示，为源数据的一种示例。其中，源数据库A在当前周期内对应的全量数据A包括数据记录A1、数据记录A2和数据记录A3；源数据库B在当前周期内对应的全量数据B包括数据记录B1、数据记录B2和数据记录B3；源数据库C在当前周期内对应的全量数据C包括数据记录C1、数据记录C2和数据记录C3。

[0077] 表2：源数据的一种示例

[0078]	源数据	数据记录
	全量数据 A	数据记录 A1
		数据记录 A2
		数据记录 A3
	全量数据 B	数据记录 B1
		数据记录 B2
		数据记录 B3
	全量数据 C	数据记录 C1
		数据记录 C2
数据记录 C3		

[0079] 举个例子,如表3a所示,为源数据库A在当前周期内发送的各源数据的一种示例。如表3b所示,为源数据库B在当前周期内发送的各源数据的一种示例。根据表3a和表3b示出的内容可知,源数据库A在当前周期内发送的源数据(即源数据库A在当前周期内对应的全量数据)中包括:数据记录A1,张三语文分数为80分;数据记录A2,李四数学分数90分;数据记录A3,赵五外语分数为100分。源数据库B在当前周期内发送的源数据(即源数据库B在当前周期内对应的全量数据)中包括:数据记录B1,张三语文指导老师为王老师;数据记录B2,李四数学指导老师为钱老师;数据记录B3,赵五外语指导老师为孙老师。

[0080] 表3a:源数据库A在当前周期内发送的源数据的一种示例

[0081]	编号	源数据		
		姓名	科目	分数
	A1	张三	语文	80
	A2	李四	数学	90
	A3	赵五	外语	100

[0082] 表3b:源数据库B在当前周期内发送的源数据的一种示例

[0083]	编号	源数据		
		姓名	科目	指导老师
	B1	张三	语文	王老师
	B2	李四	数学	钱老师
	B3	赵五	外语	孙老师

[0084] 更进一步地,每条数据记录可以包括索引值。以表3a示出的内容为例,姓名对应的内容可以为每条数据记录的索引值,即,数据记录A1的索引值为“张三”,数据记录A2的索引值为“李四”,数据记录A3的索引值为“赵五”。类似地,表3b中,数据记录B1的索引值也为“张三”,数据记录B2的索引值也为“李四”,数据记录B3的索引值也为“赵五”。

[0085] 进一步地,在执行步骤201之后,管理设备还可以获取各源数据库的数据发送情况,进而可以根据各源数据库的数据发送情况,对各源数据库的同步情况进行监控。其中,数据发送情况可以是根据述各源数据的数据发送耗时和/或数据发送量确定的。如此,可以实时监测各源数据库的同步情况,便于管理设备统计分析。

[0086] 步骤202中,根据各源数据库在当前周期内发送的各源数据,可以确定各源数据库之间的关联关系。具体地,针对第一源数据中的第一数据记录,管理设备可以根据一数据记

录对应的索引值,若确定第二源数据中存在索引值与第一数据记录对应的索引值一致的第二数据记录,则根据第一数据记录在第一数据源中的位置和第二数据记录在第二数据源中的位置,确定第一数据记录与第二数据记录的关联关系;其中,第一源数据为各源数据中任一源数据,第一数据记录为第一源数据包括的多条数据记录中任一数据记录。进而,管理设备可以根据每个源数据包括的多条数据记录之间的关联关系,确定各源数据库之间的关联关系。

[0087] 以表3a和表3b示出的内容为例,针对源数据库A发送的源数据中的数据记录A1,该数据记录A1的索引值为“张三”,并且在源数据库B发送的源数据中可以确定存在数据记录B1,该数据记录B1的索引值也为“张三”,进而,可以确定数据记录A1和数据记录B1之间存在关联关系,且这一关联关系可以根据数据记录A1在源数据库A发送的源数据中的位置(即第一行)和数据记录B1在源数据库B发送的源数据中的位置(即第一行)来确定,即,源数据库A发送的源数据中第一行的数据记录A1与源数据库B发送的源数据中第一行的数据记录B1存在关联关系。类似地,可以确定源数据库A发送的源数据中第二行的数据记录A2与源数据库B发送的源数据中第二行的数据记录B2存在关联关系,源数据库A发送的源数据中第三行的数据记录A3与源数据库B发送的源数据中第三行的数据记录B3存在关联关系。

[0088] 进而,根据数据记录A1、数据记录A2和数据记录A3与数据记录B1、数据记录B2和数据记录B3之前关联关系,可以确定源数据库A和源数据库B之间的关联关系,即源数据库A的第一行与源数据库B第一行关联,源数据库A的第二行与源数据库B第二行关联,源数据库A的第三行与源数据库B第三行关联。

[0089] 本发明实施例中,管理设备可以根据各源数据库之间的关联关系以及每个源数据包括的多条数据记录,进行计算,可以得到目标数据集合。

[0090] 具体地,管理设备可以根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定当前周期内的第一数据集合。其中,第一数据集合可以是指各源数据库发送的各源数据的集合。以表3a和表3b示出的内容为例,根据上述所描述的源数据库A和源数据库B之间的关联关系以及表3a和表3b中包括的各数据记录,可以得到如表4所示的第一数据集合的一种示例。

[0091] 表4: 第一数据集合的一种示例

编号	第一数据集合			
	姓名	科目	分数	指导老师
1	张三	语文	80	王老师
2	李四	数学	90	钱老师
3	赵五	外语	100	孙老师

[0093] 需要说明的是,上文所举的例子中,各源数据库之间可以为同构数据库。在其它可能的示例中,各源数据库之间也可以为异构数据库,异构数据库为数据库之间的结构不同的数据库。进而,各源数据的关联关系为基于结构化查询语言SQL的关联关系。具体来说,可以分析各异构源数据库,映射成对应的数据表,并对各个异构源数据的关系进行分析,输出符合要求的SQL语句。如此,采用通用的SQL语言,能够降低各异构源数据计算的难度,方便各异构源数据的逻辑计算,也便于数据的维护。

[0094] 进一步地,管理设备首先可以判断是否存在当前周期的上一周期对应的第二数据

集合,若存在,则可以确定第一数据集合相较于第二数据集合的增量数据集合,并将增量数据集合作为目标数据集合;若不存在,则可以将第一数据集合作为目标数据集合。其中,第二数据集合可以是根据各源数据库在上一周期内发送的各源数据得到的,具体的获得方式与第一数据集合的获得方式类型,此处不再详细描述。

[0095] 本发明实施例中,只需要管理设备在同步的过程进行增量数据集合的计算,使得数据宿能够直接获取增量数据集合,大大降低了整个系统的计算量,提升了数据同步的时效性;进一步地,采用管理设备向个数据宿发送增量数据集合,能够将一份增量数据集合同时发送到多个数据宿,从而能够提升数据同步的效率。

[0096] 举个例子,如5所示,为第二数据集合的一种示例。其中,第二数据集合中的第1条数据记录为,张三语文分数为80分,指导老师为王老师;李四数学分数为90分,指导老师为钱老师。

[0097] 表5:第二数据集合的一种示例

编号	第一数据集合			
	姓名	科目	分数	指导老师
1	张三	语文	80	王老师
2	李四	数学	90	钱老师

[0099] 进一步地,根据表4和表5示出的内容,可以得到如表6所示的增量数据集合的一种示例即可以确定增量数据集合为:赵五外语100分,指导老师为孙老师。

[0100] 表6:增量数据集合的一种示例

编号	增量数据集合			
	姓名	科目	分数	指导老师
3	赵五	外语	100	孙老师

[0102] 步骤203中,管理设备可以直接将目标数据集合发送至各数据宿,其中,数据宿可以用于接收数据。

[0103] 进一步地,管理设备也可以根据各数据宿的数据需求,从目标数据集合中确定符合各数据宿的数据需求的各目标数据,进而可以将各目标数据对应发送至各数据宿。举个例子,以表6示出的增量数据集合为例,若数据宿A的数据需求为需要科目和分数,则可以确定数据宿A对应的目标数据为:赵五外语分数为100分;若数据宿B的数据需求为需要科目和指导老师,则可以确定数据宿B对应的目标数据为:赵五外语指导老师为孙老师。

[0104] 在执行步骤203之后,管理设备可以获取各数据宿的数据同步信息,进而可以根据各数据宿的数据同步信息,对各数据宿的同步情况进行监控。其中,数据同步信息可以包括数据增加记录、数据删除记录和数据修改记录中的至少一项。如此,可以实时监测目标数据集合写入到各数据宿的过程,监控各数据宿的同步情况,比如监控数据增加记录、数据删除记录和数据修改记录等,便于管理设备统计分析;进而,本发明实施例中,通过管理设备监测数据同步过程,使得整个数据同步过程对于管理员而言是透明化的,从而便于管理员实时掌控数据公布情况。

[0105] 进一步地,在执行步骤203之后,管理设备若确定目标数据集合发送失败,且发送失败的次数大于预设阈值,则向管理员发送告警消息,其中,告警消息用于通知管理员目标数据集合发送失败。如此,一旦目标数据集合多次出现发送失败的情况管理设备可以及时

向管理员发送告警消息,以便管理员及时进行处理。

[0106] 为了更清楚地介绍上述数据同步方法,下面结合图3,对本发明实施例中所涉及到的流程进行整体性说明。如图3所示,可以包括以下步骤:

[0107] 步骤301,各源数据库在当前周期内向管理设备发送各源数据。

[0108] 步骤302,管理设备判断各源数据是否全部发送完成,若未全部发送完成,则执行步骤303;否则,执行步骤305。

[0109] 步骤303,管理设备判断各源数据的数据发送耗时是否大于预设阈值,若大于,则执行步骤304;否则,返回步骤302。

[0110] 步骤304,管理设备尝试重新接收数据发送耗时大于预设阈值的源数据。

[0111] 步骤305,管理设备根据各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定当前周期内的第一数据集合。

[0112] 步骤306,管理设备判断是否存在当前周期的上一周期对应的第二数据集合,若存在,执行步骤307;否则,执行步骤310。

[0113] 步骤307,管理设备确定第一数据集合相较于第二数据集合的增量数据集合,并将增量数据集合作为目标数据集合。

[0114] 步骤308,管理设备根据各数据宿的数据需求,从目标数据集合中确定符合所述各数据宿的数据需求的各目标数据。

[0115] 步骤309,管理设备将各目标数据对应发送至各数据宿。

[0116] 步骤310,管理设备将第一数据集合作为所述目标数据集合,并根据各数据宿的数据需求,从目标数据集合中确定符合所述各数据宿的数据需求的各目标数据。

[0117] 步骤311,管理设备将各目标数据发送至各数据宿。

[0118] 针对上述步骤301至步骤311所描述的内容,下面以异构源数据库为例,进行举例说明。首先,管理设备可以分析各异构源数据库,映射成对应的源数据表,并对各个异构源数据的关系进行分析,输出符合要求的SQL语句;然后,管理设备可以选择一种稳定的内存数据库方案(例如h2),根据异构源数据对应的源数据表,创建内存数据库;再然后,管理设备可以执行定时任务,将各个异构源数据分别同步到内存数据库中;进而,管理设备可以检查各个异构源数据是否已全部同步到内存数据库相应表中,如果全部同步,则开始计算,具体地,根据上文输出的SQL语句,在内存数据库中执行查询,生成当前周期的全量数据,然后判断是否存在上一周期对应的全量数据,如果不存在,则可以直接推送当前周期的全量数据到各数据宿中;如果存在,则可以将当前周期的全量数据与上一周期的全量数据做比较,生成增量数据集合,并将增量数据集合推送到各数据宿中。进一步地,管理设备还可以检查各数据宿更新的数据于内存数据库中生成的数据在数量和内容上是否一致,如果一致,则表明同步成功。

[0119] 基于同样的发明构思,图4示例性示出了本发明实施例提供的一种数据同步装置的结构示意图,该装置可以应用于数据同步系统,该系统包括管理设备、多个源数据库和至少一个数据宿;该装置可以为管理设备,如图4所示,该装置包括接收单元401、处理单元402、发送单元403;其中,

[0120] 接收单元401,用于接收各源数据库在当前周期内发送的各源数据;所述各源数据为所述各源数据库在所述当前周期内对应的全量数据;

[0121] 处理单元402,用于根据所述各源数据库之间的关联关系以及每个源数据包括的多条数据记录,确定所述当前周期内的第一数据集合;若确定存在所述当前周期的上一周期对应的第二数据集合,则比较所述第二数据集合和所述第一数据集合,得到增量数据集合;所述第二数据集合是根据所述各源数据库在所述上一周期内发送的各源数据得到的;

[0122] 发送单元403,用于将所述增量数据集合发送至各数据宿。

[0123] 在一种可能的实现方式中,每条数据记录包括索引值;

[0124] 所述各源数据库之间的关联关系通过以下方式确定:

[0125] 针对第一源数据中的第一数据记录,所述管理设备根据所述第一数据记录对应的索引值,若确定第二源数据中存在索引值与所述第一数据记录对应的索引值一致的第二数据记录,则根据所述第一数据记录在所述第一数据源中的位置和所述第二数据记录在所述第二数据源中的位置,确定所述第一数据记录与所述第二数据记录的关联关系;所述第一源数据为所述各源数据中任一源数据,所述第一数据记录为所述第一源数据包括的多条数据记录中任一数据记录;

[0126] 根据所述每个源数据包括的多条数据记录之间的关联关系,确定所述各源数据库之间的关联关系。

[0127] 在一种可能的实现方式中,所述各源数据库之间为异构数据库;

[0128] 所述各源数据的关联关系为基于结构化查询语言SQL的关联关系。

[0129] 在一种可能的实现方式中,所述发送单元403具体用于:

[0130] 根据所述各数据宿的数据需求,从所述增量数据集合中确定符合所述各数据宿的数据需求的各目标增量数据;以及将所述各目标增量数据对应发送至所述各数据宿。

[0131] 在一种可能的实现方式中,所述处理单元402还用于:

[0132] 若确定不存在所述第二数据集合,则将所述第一数据集合发送至各数据宿。

[0133] 在一种可能的实现方式中,所述发送单元403将所述增量数据集合发送至各数据宿之后,所述处理单元402还用于:

[0134] 获取所述各数据宿的数据同步信息;所述数据同步信息包括数据增加记录、数据删除记录和数据修改记录中的至少一项;以及根据所述各数据宿的数据同步信息,对所述各数据宿的同步情况进行监控。

[0135] 在一种可能的实现方式中,所述接收单元401接收各源数据库在当前周期内发送的源数据之后,所述处理单元402还用于:

[0136] 获取所述各源数据库的数据发送情况,所述数据发送情况是根据所述各源数据的数据发送耗时和/或数据发送量确定的;以及根据所述各源数据库的数据发送情况,对所述各源数据库的同步情况进行监控。

[0137] 本申请实施例的还提供一种装置,该装置具有实现上文所描述的数据同步方法的功能。该功能可以通过硬件执行相应的软件实现,在一种可能的设计中,该装置包括:处理器、收发器、存储器;该存储器用于存储计算机执行指令,该收发器用于实现该装置与其他通信实体进行通信,该处理器与该存储器通过该总线连接,当该装置运行时,该处理器执行该存储器存储的该计算机执行指令,以使该装置执行上文所描述的数据同步方法。

[0138] 本发明实施例还提供一种计算机存储介质,所述存储介质中存储软件程序,该软件程序在被一个或多个处理器读取并执行时实现上述各种可能的实现方式中所描述的数

据同步方法。

[0139] 本发明实施例还提供一种包含指令的计算机程序产品,当其在计算机上运行时,使得计算机执行上述各种可能的实现方式中所描述的数据同步方法。

[0140] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明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,本发明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。

[0141] 本发明是参照根据本发明实施例的方法、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。

[0142] 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。

[0143] 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。

[0144] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0145] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

100

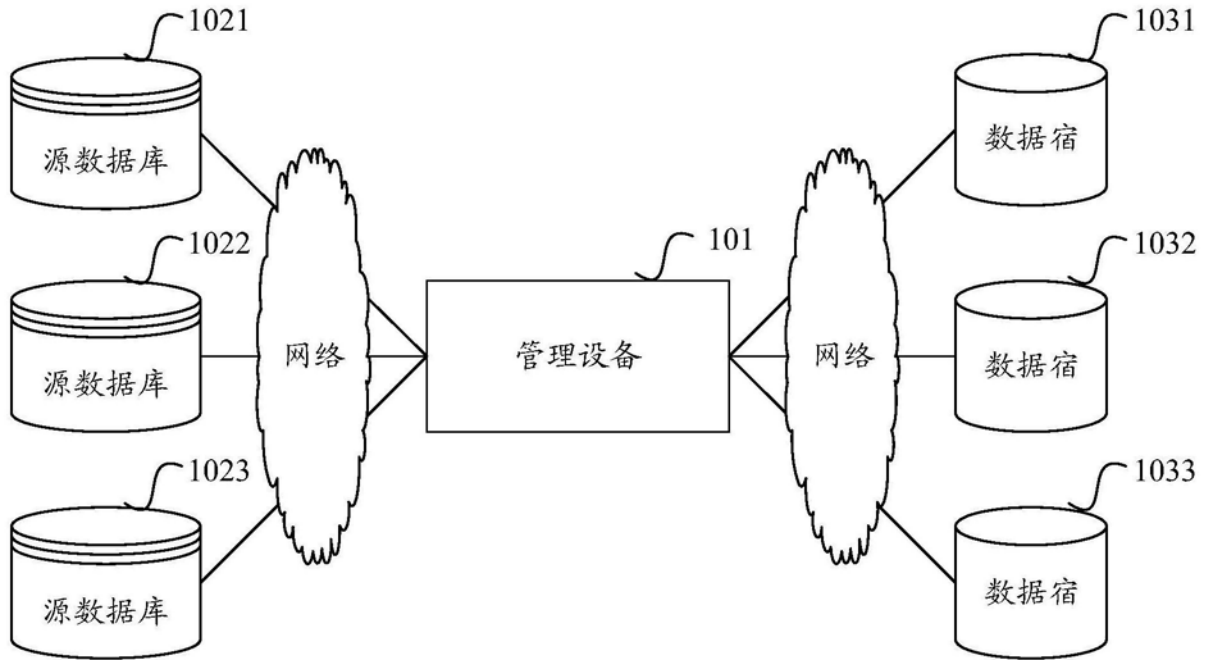


图1

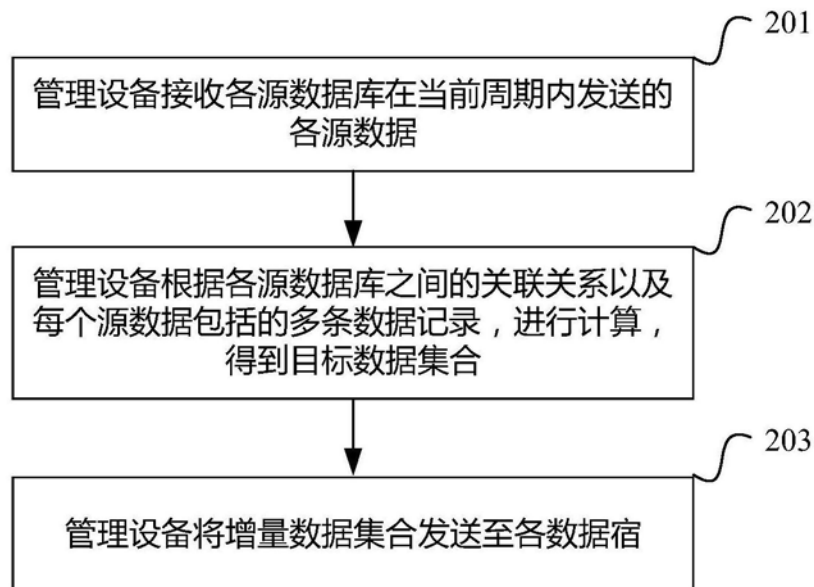


图2

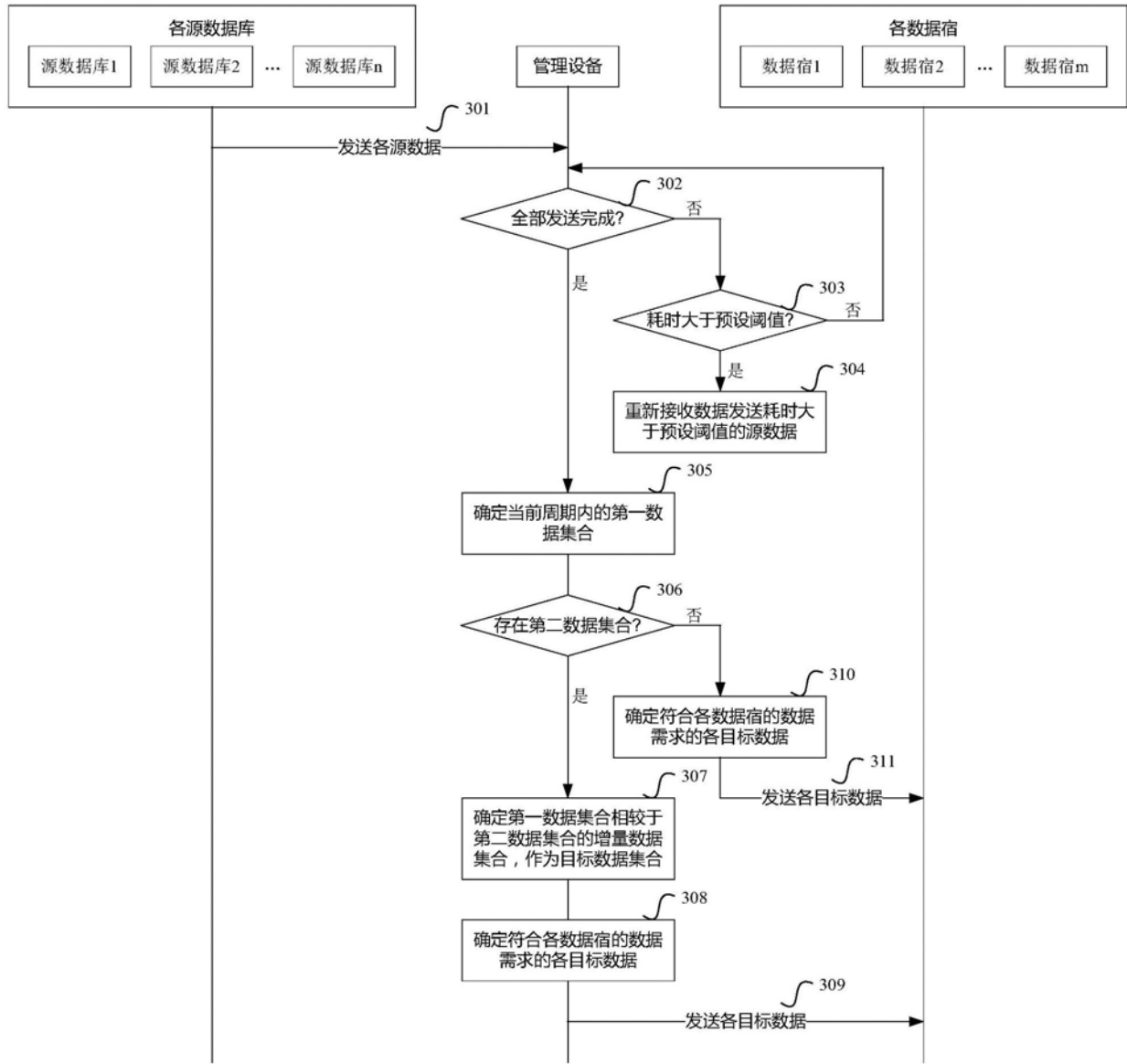


图3

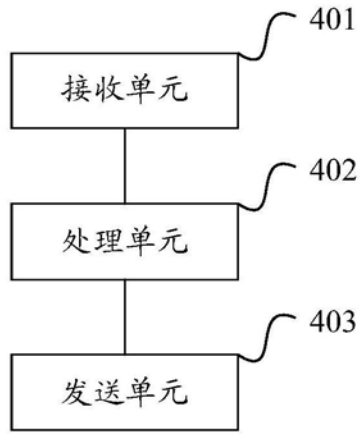


图4