



(12) 发明专利

(10) 授权公告号 CN 110083658 B

(45) 授权公告日 2021. 05. 25

(21) 申请号 201910181541.1

审查员 王辞

(22) 申请日 2019.03.11

(65) 同一申请的已公布的文献号

申请公布号 CN 110083658 A

(43) 申请公布日 2019.08.02

(73) 专利权人 北京达佳互联信息技术有限公司

地址 100085 北京市海淀区上地西路6号1

幢1层101D1-7

(72) 发明人 刘硕 李明 马彦兵 罗文楠

(74) 专利代理机构 北京润泽恒知识产权代理有

限公司 11319

代理人 莎日娜

(51) Int. Cl.

G06F 16/27 (2019.01)

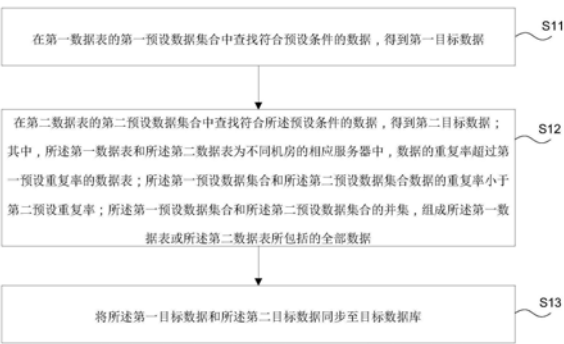
权利要求书2页 说明书17页 附图5页

(54) 发明名称

数据同步方法、装置、电子设备及存储介质

(57) 摘要

本申请是关于一种数据同步方法、装置、电子设备及存储介质,涉及数据处理技术领域,所述方法包括:在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;将所述第一目标数据和所述第二目标数据同步至目标数据库。各个机房分担了上述同步任务,缓解了各个机房的压力。各个服务器使用独立的线程池进行数据查找等,提升同步效率。



1. 一种数据同步方法,其特征在于,所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述方法包括:

在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;

将所述第一目标数据和所述第二目标数据同步至目标数据库。

2. 根据权利要求1所述的数据同步方法,其特征在于,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还包括:

将每个所述服务器的线程池初始化为预设数量的线程;

所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据,包括:

基于所述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

所述在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据,包括:

基于所述预设数量的线程,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据。

3. 根据权利要求1所述的数据同步方法,其特征在于,所述至少两个数据表分别存储于至少两个不同的源数据库中,所述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,所述至少两个不同的数据集群分别位于至少两个不同的机房中,所述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还包括:

将所述数据写入所述至少一个主数据集群;

将所述数据从任意一个所述主数据集群同步至所述至少一个从数据集群。

4. 根据权利要求3所述的数据同步方法,其特征在于,在所述至少两个不同的数据集群包括:至少两个从数据集群的情况下,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还包括:

在所述至少两个从数据集群中,确定所述第一数据表和所述第二数据表。

5. 根据权利要求1所述的数据同步方法,其特征在于,所述第一数据表的第一预设数据集合与所述第二数据表的第二预设数据集合的交集为0。

6. 根据权利要求1所述的数据同步方法,其特征在于,所述第一预设数据集合中数据的数量,与各个所述第二预设数据集合中数据的数量均相同。

7. 一种数据同步装置,其特征在于,所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述装置包括:

第一目标数据查找模块,被配置为在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

第二目标数据查找模块,被配置为在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;

同步模块,被配置为将所述第一目标数据和所述第二目标数据同步至目标数据库。

8. 根据权利要求7所述的数据同步装置,其特征在于,还包括:

线程初始化模块,被配置为将每个所述服务器的线程池初始化为预设数量的线程;

所述第一目标数据查找模块,包括:

第一目标数据查找单元,被配置为基于所述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

所述第二目标数据查找模块,包括:

第二目标数据查找单元,被配置为基于所述预设数量的线程,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据。

9. 根据权利要求7所述的数据同步装置,其特征在于,所述至少两个数据表分别存储于至少两个不同的源数据库中,所述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,所述至少两个不同的数据集群分别位于至少两个不同的机房中,所述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群,还包括:

主数据集群写入模块,配置为将所述数据写入所述至少一个主数据集群;

从数据集群写入模块,配置为将所述数据从任意一个所述主数据集群同步至所述至少一个从数据集群。

10. 根据权利要求9所述的数据同步装置,其特征在于,在所述至少两个不同的数据集群包括:至少两个从数据集群的情况下,还包括:

第一数据表和所述第二数据表确定模块,被配置为在所述至少两个从数据集群中,确定所述第一数据表和所述第二数据表。

11. 根据权利要求7所述的数据同步装置,其特征在于,所述第一数据表的第一预设数据集合与所述第二数据表的第二预设数据集合的交集为0。

12. 根据权利要求7所述的数据同步装置,其特征在于,所述第一预设数据集合中数据的数量,与各个所述第二预设数据集合中数据的数量均相同。

13. 一种电子设备,其特征在于,包括:

处理器;

用于存储处理器可执行指令的第一存储器;

其中,所述处理器被配置为执行如权利要求1至6中任一项所述的数据同步方法。

14. 一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行如权利要求1至6中任一项所述的数据同步方法。

数据同步方法、装置、电子设备及存储介质

技术领域

[0001] 本公开涉及数据处理技术领域,尤其涉及数据同步方法、装置、电子设备及存储介质。

背景技术

[0002] 相关技术中,许多业务对应较多的数据,通过对上述数据进行查找等,挑选出符合业务需求的数据,将上述挑选的数据同步至目标数据库中以进一步处理,可以获取更大的价值,因此,对数据进行查找、同步的需求也越来越多。

[0003] 通常针对存储于一个机房所有数据,进行处理,完成对数据的查找、同步等。

[0004] 然而,进行查找、同步的数据通常为海量数据,如数以亿计的数据,对一个机房中的所有数据进行查找、同步等操作,耗时较长。

发明内容

[0005] 为克服相关技术中存在的问题,本公开提供一种数据同步方法及装置。

[0006] 根据本公开实施例的第一方面,提供一种数据同步方法,所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述方法包括:

[0007] 在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

[0008] 在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;

[0009] 将所述第一目标数据和所述第二目标数据同步至目标数据库。

[0010] 可选的,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还包括:

[0011] 将每个所述服务器的线程池初始化为预设数量的线程;

[0012] 所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据,包括:

[0013] 基于所述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

[0014] 所述在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据,包括:

[0015] 基于所述预设数量的线程,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据。

[0016] 可选的,所述至少两个数据表分别存储于至少两个不同的源数据库中,所述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,所述至少两个不同的数据集群分别位于至少两个不同的机房中,所述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还包括:

[0017] 将所述数据写入所述至少一个主数据集群;

[0018] 将所述数据从任意一个所述主数据集群同步至所述至少一个从数据集群。

[0019] 可选的,在所述至少两个不同的数据集群包括:至少两个从数据集群的情况下,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还包括:

[0020] 在所述至少两个从数据集群中,确定所述第一数据表和所述第二数据表。

[0021] 可选的,所述第一数据表的第一预设数据集合与所述第二数据表的第二预设数据集合的交集为0。

[0022] 可选的,所述第一预设数据集合中数据的数量,与各个所述第二预设数据集合中数据的数量均相同。

[0023] 根据本公开实施例的第二方面,提供一种数据同步装置,所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述装置包括:

[0024] 第一目标数据查找模块,被配置为在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

[0025] 第二目标数据查找模块,被配置为在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;

[0026] 同步模块,被配置为将所述第一目标数据和所述第二目标数据同步至目标数据库。

[0027] 可选的,所述数据同步装置,还包括:

[0028] 线程初始化模块,被配置为将每个所述服务器的线程池初始化为预设数量的线程;

[0029] 所述第一目标数据查找模块,包括:

[0030] 第一目标数据查找单元,被配置为基于所述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

[0031] 所述第二目标数据查找模块,包括:

[0032] 第二目标数据查找单元,被配置为基于所述预设数量的线程,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据。

[0033] 可选的,所述至少两个数据表分别存储于至少两个不同的源数据库中,所述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,所述至少两个不同的数据集

群分别位于至少两个不同的机房中,所述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群,所述数据同步装置,还包括:

[0034] 主数据集群写入模块,配置为将所述数据写入所述至少一个主数据集群;

[0035] 从数据集群写入模块,配置为将所述数据从任意一个所述主数据集群同步至所述至少一个从数据集群。

[0036] 可选的,所述数据同步装置,在所述至少两个不同的数据集群包括:至少两个从数据集群的情况下,还包括:

[0037] 第一数据表和所述第二数据表确定模块,被配置为在所述至少两个从数据集群中,确定所述第一数据表和所述第二数据表。

[0038] 可选的,所述第一数据表的第一预设数据集合与所述第二数据表的第二预设数据集合的交集为0。

[0039] 可选的,所述第一预设数据集合中数据的数量,与各个所述第二预设数据集合中数据的数量均相同。

[0040] 根据本公开实施例的第三方面,提供一种电子设备,包括:

[0041] 处理器;

[0042] 用于存储处理器可执行指令的存储器;

[0043] 其中,所述处理器被配置为:执行上述的数据同步方法。

[0044] 根据本公开实施例的第四方面,提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行上述的数据同步方法。

[0045] 根据本公开实施例的第五方面,提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行上述的数据同步方法。

[0046] 本公开的实施例提供的技术方案可以包括以下有益效果:

[0047] 所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述方法包括:在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;将所述第一目标数据和所述第二目标数据同步至目标数据库。上述数据存储于各个不同机房中,在不同机房的相应的服务器中数据重复率较高的数据表中的不同预设数据集合中,查找符合预设条件的数据,也就是将查找操作在各个机房的相应服务器中进行了分解,每个机房的相应服务器中只查询了其存储数据表中的部分数据,并没有查询其存储数据表的全部数据,同时,第一预设数据集合和第二预设数据集合的并集,组成第一数据表或第二数据表所包括的全部数据,各个机房的相应服务器的查询结果形成了对该数据的查询结果,即查询结果是针对第一数据表的全部数据,或第二数据表的全部数据进行查询,查询全面没有遗漏;各个机房或各个机房的相应

服务器分担了上述查询,一方面缓解了各个机房的压力,同时缓解了各个机房相应服务器的压力,不会占用各个机房或各个机房相应服务器较多进程,可以提升数据同步效率,另一方面各个服务器都是位于不同机房的独立机器,各个服务器使用独立的线程池进行数据查找等,从很大程度上提升了数据查找等的效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0048] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0049] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。

[0050] 图1是根据一示例性实施例示出的一种数据同步方法的流程图。

[0051] 图2是根据一示例性实施例示出的一种数据分布示意图。

[0052] 图3是根据一示例性实施例示出的另一种数据同步方法的流程图。

[0053] 图4是根据一示例性实施例示出的一种数据同步装置框图。

[0054] 图5是根据一示例性实施例示出的另一种数据同步装置框图。

[0055] 图6是根据一示例性实施例示出的一种电子设备的框图。

[0056] 图7是根据一示例性实施例示出的另一种电子设备的框图。

具体实施方式

[0057] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

[0058] 图1是根据一示例性实施例示出的一种数据同步方法的流程图,所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中。

[0059] 具体的,可以参照图2所示,图2是根据一示例性实施例示出的一种数据分布示意图。上述图2中n、m均可以为大于等于2的整数。在本申请实施例中,上述数据分别存储于至少两个数据表中,上述至少两个数据表分别存储于至少两个服务器中,至少两个服务器分别位于至少两个不同的机房中。

[0060] 例如,参照图2所示,数据a可以存储在第1机房的服务器0中的数据表0中,该数据a还可以存储在第2机房的服务器0中的数据表0中,该数据a还可以存储在第m机房的服务器0中的数据表0中。m可以为大于等于2的整数。在本申请实施例中,对此不作具体限定。

[0061] 如图1所示,数据同步方法包括以下步骤:

[0062] 在步骤S11中,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据。

[0063] 在本申请实施例中,该第一数据表可以为上述任意机房中的任意数据表。在本申请实施例中,对此不作具体限定。

[0064] 例如,参照图2,该第一数据表可以为第1机房服务器0中的数据表0,或者,该第一数据表可以为第2机房服务器1中的数据表1等。

[0065] 在本申请实施例中,第一数据表的第一预设数据集合可以为该第一数据表中的任意条数据记录的集合,该第一预设数据集合中数据记录的条数可以为1条或多条等,在本申请实施例中,对此不作具体限定。

[0066] 例如,若第一数据表为第1机房服务器0中的数据表0,若第1机房服务器0中的数据表0中包括200万条数据记录,则第一数据表的第一预设数据集合可以为第1机房服务器0中的数据表0中包括200万条数据记录中的任意条数据记录的集合。如,第一数据表的第一预设数据集合可以为第1机房服务器0中的数据表0中包括的200万条数据记录中的前1000条数据记录。

[0067] 在本申请实施例中,该预设条件可以根据实际需要进行设定等,在本申请实施例中,对此不作具体限定。

[0068] 在本申请实施例中,可以在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据。该第一目标数据的数量可以大于等于0。

[0069] 例如,针对上述例子,若第一数据表的第一预设数据集合可以为第1机房服务器0中的数据表0中包括的200万条数据记录中的前1000条数据记录,则可以在第1机房服务器0中的数据表0中包括的200万条数据记录中的前1000条数据记录中,查找符合预设条件的数据,得到第一目标数据。

[0070] 在本申请实施例中,各个数据表中的每一条数据记录可以带有索引,如果第一预设集合为第一数据表中索引连续的数据记录,则可以通过起始索引等,完成从第一数据表的第一预设数据集合查找上述第一目标数据。

[0071] 例如,针对上述例子,可以采用的如下的SQL语句,从第1机房服务器0中的数据表0中包括的200万条数据记录中的前1000条数据记录中,便捷快速地查找上述第一目标数据:

[0072] `SELECT*FROM TABLE0 WHERE ID>=0AND LIMIT 1000。`

[0073] 或者,可以采用的如下的SQL语句,从第1机房服务器0中的数据表0中包括的200万条数据记录中的前1000条数据记录中,便捷快速地查找上述第一目标数据:

[0074] `SELECT*FROM TABLE0 WHERE ID>=0AND ID=<999LIMIT 1000。`

[0075] 在步骤S12中,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据。

[0076] 在本申请实施例中,可以提前设定第一预设重复率,该第一预设重复率可以根据实际需要进行设定。在本申请实施例中,对此不作具体限定。如,该第一预设重复率可以为99%。

[0077] 在本申请实施例中,上述第一数据表与上述第二数据表为:在不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表。该第二数据表的数量可以为1个或多个。在本申请实施例中,对此不作具体限定。

[0078] 在本申请实施例中,该第二数据表可以只存在于一个机房中,或者,该第二数据表

可以存在于多个机房中。在本申请实施例中,对此不作具体限定。

[0079] 例如,若第一数据表为第1机房服务器0中的数据表0,则第二数据表可以为第2机房相应服务器中,与第1机房中服务器0的数据表0中存储的200万条数据记录重复率超过该第一预设重复率的数据表0。或者,第二数据表可以为第2机房至第m机房相应服务器中,与第1机房服务器0中的数据表0中存储的200万条数据记录重复率超过该第一预设重复率的多个数据表0。如,该第二数据表可以为:第2机房的服务器0中的数据表0、第3机房的服务器0中的数据表0、第4机房的服务器0中的数据表0、……第m机房的服务器0中的数据表0,第2机房的服务器0中的数据表0、第3机房的服务器0中的数据表0、第4机房的服务器0中的数据表0、……第m机房的服务器0中的数据表0,分别与第一数据表第1机房服务器0中的数据表0中存储的200万条数据记录重复率,均超过该第一预设重复率。

[0080] 在本申请实施例中,可以提前设定该第二预设重复率,该第二预设重复率可以根据实际需要进行设定等。在本申请实施例中,对此不作具体限定。该第二预设重复率可以小于或远远小于上述第一预设重复率,或者,该第一预设重复率可以大于,或远远大于该第二预设重复率。

[0081] 例如,该第二预设重复率可以设定为1%,若第一预设重复率为99%,第一预设重复率为99%远远大于该第二预设重复率1%。

[0082] 在本申请实施例中,第一数据表和第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表,第二数据表的第二预设数据集合与第一数据表的第一预设数据集合数据的重复率可以小于第二预设重复率。第一预设数据集合和第二预设数据集合的并集,组成该第一数据表或第二数据表所包括的全部数据。

[0083] 具体的,可以在不同机房的相应服务器中,先找到数据的重复率超过第一预设重复率的第一数据表和第二数据表,然后可以在第一数据表中选择部分数据,作为该第一数据表的第一预设数据集合。如果第二数据表的数量为1,则再在第二数据表中,选择与上述第一预设数据集合数据的重复率小于第二预设重复的第二预设数据集合,上述第一预设数据集合与第二预设数据集合的并集组成第一数据表或第二数据表所包括的全部数据。如果,第二数据表的数量大于1,则分别在各个第二数据表中,选择与上述第一预设数据集合数据的重复率小于第二预设重复率的第二预设数据集合,可以得到1个或多个第二预设数据集合。若第二预设数据集合为1个,则第一预设数据集合与该第二预设数据集合的并集组成第一数据表或第二数据表所包括的全部数据。若第二预设数据集合为多个,则第一预设数据集合与多个第二预设数据集合的并集,组成第一数据表或第二数据表所包括的全部数据,进而上述对目标数据的查找,涵盖了第一数据表或第二数据表的全部数据,进而对目标数据的查找全面没有遗漏。

[0084] 例如,参照图2,若第一数据表为第1机房服务器0中的数据表0,第1机房服务器0中的数据表0中存储的200万条数据。若在第2机房至第m机房中,与第1机房服务器0中的数据表0中存储的200万条数据的数据重复率超过第一预设重复率的数据表为:第2机房的服务器0中的数据表0、第3机房的服务器0中的数据表0、第4机房的服务器0中的数据表0、……第m机房的服务器0中的数据表0。

[0085] 若从第1机房服务器0中的数据表0中存储的200万条数据中选择的第一预设数据集合为:第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录。参照

下表1,表1示出了本申请实施例中,第一预设数据集合和第二预设数据集合的分布情况。若在上述各个第二数据表中,选择与上述第一预设数据集合数据的重复率小于第二预设重复率的第二预设数据集合分别为:第2机房的服务器0中的数据表0的第1000至第1999条数据记录、第3机房的服务器0中的数据表0的第2000至第2999条数据记录、第4机房的服务器0中的数据表0的第3000至第3999条数据记录、……第m机房的服务器0中的数据表0的第19000至第19999条数据记录,上述第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录,分别与:第2机房的服务器0中的数据表0的第1000至第1999条数据记录、第3机房的服务器0中的数据表0的第2000至第2999条数据记录、第4机房的服务器0中的数据表0的第3000至第3999条数据记录、……第m机房的服务器0中的数据表0的第19000至第19999条数据记录的数据重复率,均小于该第二预设重复率。同时,第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录,与第2机房的服务器0中的数据表0的第1000至第1999条数据记录、第3机房的服务器0中的数据表0的第2000至第2999条数据记录、第4机房的服务器0中的数据表0的第3000至第3999条数据记录、……第m机房的服务器0中的数据表0的第19000至第19999条数据记录的并集,组成第一数据表第2机房的服务器0中的数据表0的全部数据,或,组成第二数据表第2机房的服务器0中的数据表0的全部数据,或,组成第二数据表第3机房的服务器0中的数据表0的全部数据,……,或,组成第二数据表第m机房的服务器0中的数据表0的全部数据。

[0086] 表1

[0087]

机房	第一预设数据集合	第二预设数据集合
第 1 机房	服务器 0 中的数据表 0 中的前 1000 条数据记录	
第 2 机房		服务器 0 中的数据表 0 的第 1000 至第 1999 条数据记录
第 3 机房		服务器 0 中的数据表 0 的第 2000 至第 2999 条数据记录
第 4 机房		服务器 0 中的数据表 0 的第 3000 至第 3999 条数据记录
.....	
第 m 机房		服务器 0 中的数据表 0 的第 19000 至第 19999 条数据记录

[0088] 在本申请实施例中,第一预设数据集合中数据的数量可以大于等于1,第二预设数据集合中数据的数量也可以大于等于1。第一预设数据集合中数据的数量与,第二预设数据集合中数据的数量可以相等或不等。若第二数据表为多个,则各个第二数据表的第二预设数据集合中数据的数量也可以相等或不等。第一数据表第二预设数据集合中数据的数量,与各个第二数据表的第二预设数据集合中数据的数量也可以相等或不等在本申请实施例中,对此不作具体限定。

[0089] 在本申请实施例中,可选的,上述第一数据表和第二数据表可以为不同机房的相应服务器中存储有相同数据的数据表。或者,上述第一数据表和第二数据表可以为不同机

房的相应服务器中存储有预设比例相同数据的数据表。在本申请实施例中,对此不作具体限定。

[0090] 在本申请实施例中,在上述第二数据表的第二预设集合中查找符合预设条件的数据,得到第二目标数据。

[0091] 例如,针对上述例子,则可以分别在第2机房的服务器0中的数据表0的第1000至第1999条数据记录中查找符合所述预设条件的数据,得到第二目标数据、第3机房的服务器0中的数据表0的第2000至第2999条数据记录中查找符合所述预设条件的数据,得到第二目标数据、第4机房的服务器0中的数据表0的第3000至第3999条数据记录中查找符合所述预设条件的数据,得到第二目标数据、……第m机房的服务器0中的数据表0的第19000至第19999条数据记录中查找符合所述预设条件的数据,得到第二目标数据。

[0092] 在本申请实施例中,各个数据表中的每一条数据记录可以带有索引,如果第二预设集合为第二数据表中索引连续的数据记录,则同样可以通过起始索引等,完成从第一数据表的第一预设数据集合查找上述第一目标数据。

[0093] 例如,针对上述例子,可以采用的如下的SQL语句,从第2机房服务器0中的数据表0中第1000至第1999条数据记录中,便捷快速地查找上述第二目标数据:

[0094] SELECT*FROM TABLE0 WHERE ID>=1000AND LIMIT 1000。

[0095] 或者,可以采用的如下的SQL语句,从第2机房服务器0中的数据表0中第1000至第1999条数据记录中,便捷快速地查找上述第一目标数据:

[0096] SELECT*FROM TABLE0 WHERE ID>=1000AND ID=<1999LIMIT 1000。

[0097] 在步骤S13中,将所述第一目标数据和所述第二目标数据同步至目标数据库。

[0098] 在本申请实施例中,同步可以包括:数据格式转换或数据格式兼容、无效数据过滤等。如,若从第一预设数据集合或第二预设数据集合中,查找的符合预设要求的数据,与目标数据库或目标数据库的数据格式不兼容等,则在将上述第一目标数据和第二目标数据同步至目标数据库的过程中,可以将上述第一目标数据和第二目标数据转换为与目标数据库或目标数据库的数据格式兼容的数据格式。如,将上述第一目标数据和第二目标数据同步至目标数据库的过程中,可以将上述第一目标数据和第二目标数据中的无效数据等,进行过滤,然后将过滤后的第一目标数据和第二目标数据,写入该目标数据库中。在本申请实施例中,对此不作具体限定。

[0099] 在本申请实施例中,目标数据库可以为存储目标数据的数据库,该目标数据可以为上述第一目标数据和第二目标数据的并集。该目标数据库可以存储于与上述第一数据表或第二数据表相同的机房中,或者,该目标数据库可以存储于与上述第一数据表或第二数据表不同的机房中。在本申请实施例中,对此不作具体限定。

[0100] 例如,针对上述例子,上述目标数据库可以存储于第1机房中,或者目标数据库可以存储于第m+1机房中。

[0101] 在本申请实施例中,将从第一数据表的第一预设数据集合中查找到的第一目标数据,和从第二数据表的第二预设数据集合中查找到的第二目标数据,都同步至上述目标数据库中。

[0102] 例如,针对上述例子,若从第一数据表的第一预设数据集合:第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录中查找到的第一目标数据为:第1机

房服务器0中的数据表0中的第30至60条数据,在第2机房的服务器0中的数据表0的第1000至第1999条数据记录中查找到的第二目标数据为:第2机房的服务器0中的数据表0的第1111至1200条数据、在第3机房的服务器0中的数据表0的第2000至第2999条数据记录中查找到的第二目标数据为:第3机房的服务器0中的数据表0的第2020至2040条数据、在第4机房的服务器0中的数据表0的第3000至第3999条数据记录……至第m机房的服务器0中的数据表0的第19000至第19999条数据记录中均没有查找到第二目标数据。则,上述第一目标数据可以为:第1机房服务器0中的数据表0中的第30至60条数据。上述第二目标数据可以为:第2机房的服务器0中的数据表0的第1111至1200条数据和第3机房的服务器0中的数据表0的第2020至2040条数据。则将,上述第一目标数据和上述第二目标数据同步至目标数据库中。

[0103] 相当于,分别查找了各个机房相应服务器的第一数据表和第二数据表中的部分数据,各个机房或各个机房相应服务器分别分担了对符合预设条件的数据的查找任务,针对上述例子,则是第1机房至第m机房分别分担了对符合预设条件的数据的查找任务,或者,第1机房的至第m机房分别分担了对符合预设条件的数据的查找任务可以从很大程度上缓解各个机房的压力,同时缓解了各个机房相应服务器的压力,不会占用各个机房或各个机房相应服务器较多进程,可以提升数据同步效率,而且,上述各个服务器为独立的机器,具有各个服务器使用独立的线程池进行数据查找等,从很大程度上提升了数据查找等的效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0104] 在本申请实施例中,可选的,机房数量越多,进而可能将同步任务分担至更多的机房中,可以从一定程度上提升了数据查找等的效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0105] 本公开的实施例提供的技术方案可以包括以下有益效果:所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述方法包括:在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;将所述第一目标数据和所述第二目标数据同步至目标数据库。上述数据存储于各个不同机房中,在不同机房的相应的服务器中数据重复率较高的数据表中的不同预设数据集合中,查找符合预设条件的数据,也就是将查找操作在各个机房的相应服务器中进行了分解,每个机房的相应服务器中只查询了其存储数据表中的部分数据,并没有查询其存储数据表的全部数据,同时,第一预设数据集合和第二预设数据集合的并集,组成第一数据表或第二数据表所包括的全部数据,各个机房的相应服务器的查询结果形成了对该数据的查询结果,即查询结果是针对第一数据表的全部数据,或第二数据表的全部数据进行查询,查询全面没有遗漏;各个机房或各个机房的相应服务器分担了上述查询,一方面缓解了各个机房的压力,同时缓解了各个机房相应服务器的压力,不会占用各个机房或各个机房相应服务器较多进程,可以提升数据同步效率,另一方面各个服务器都是位于不同

机房的独立机器,各个服务器使用独立的线程池进行数据查找等,从很大程度上提升了数据查找等的效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0106] 图3是根据一示例性实施例示出的一种数据同步方法的流程图,如图3所示,数据同步方法包括以下步骤。

[0107] 在步骤S21中,将每个所述服务器的线程池初始化为预设数量的线程。

[0108] 在本申请的实施例中,每个服务器通常为独立的机器,因此每个服务器均具有独立的线程池进行数据查找。可以将每个服务器的线程池初始化为预设数量的线程。

[0109] 例如,参照图2所示,若第1机房至第m机房的所有服务器均具有1个独立的线程池进行数据查找,则可以将上述每个服务器的上述独立的线程池初始化为预设数量的线程。

[0110] 在本申请的实施例中,上述预设数量,可以根据实际需要进行设定,该预设数量可以为大于等于1的整数。在本申请的实施例中,对此不作具体限定。

[0111] 例如,可以根据各个服务器存储的数据表的数量,确定各个服务器对应的线程的数量。如,若第1机房服务器0中存储有10000个数据表,可以将第1机房服务器0中处理数据查询的独立线程池初始化为10个线程,进而上述10个线程对应处理第1机房服务器0中的10000个数据表的查询任务。

[0112] 在本申请的实施例中,各个机房的各个服务器中,初始化后的线程的数量,可以相同或不同。同一机房的各个服务器中,初始化后的线程的数量也可以相同或不同,在本申请的实施例中,对此不作具体限定。

[0113] 在本申请的实施例中,可选的,所述至少两个数据表分别存储于至少两个不同的源数据库中,所述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,所述至少两个不同的数据集群分别位于至少两个不同的机房中,所述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还可以包括:将所述数据写入所述至少一个主数据集群;将所述数据从任意一个所述主数据集群同步至所述至少一个从数据集群。

[0114] 具体的,上述至少两个数据表分别存储于至少两个不同的源数据库中,上述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,上述至少两个不同的数据集群分别位于至少两个不同的机房中,上述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群。即,各个机房可以分别对应一个数据集群,各个数据集群可以包括至少一个主数据集群和至少一个从数据集群,上述主数据集群的数量、从数据集群的数量,不作具体限定。

[0115] 在本申请实施例中,可以先将数据写入上述至少一个主数据集群,然后从上述任一个主数据集群中,将数据同步至上述至少一个从数据集群。可以将主数据集群对应的机房性能设置为优于从数据集群的性能,进而可以从很大程度上提升先数据集群写入数据的效率。

[0116] 例如,参照图2,第1机房至第m机房可以分别对应第一数据集群至第m数据集群,该第1机房对应的第一数据集群可以为主数据集群,第2机房至第m机房可以为从数据集群。通常数据可以先写入上述该第1机房对应的主数据集群,然后,可以从上述第1机房对应的主数据集群同步至上述第2机房至第m机房对应的从数据集群中。也就是说,若忽略从上述第1

机房对应的主数据集群同步至上述第2机房至第m机房对应的从数据集群中的延迟的情况下,上述第1机房至第m机房中存储的数据可以认为是一致的,或相同的。

[0117] 在本申请实施例中,可选的,上述主数据集群对应的机房、或从数据集群对应的机房,均可以用作第一目标数据或第二目标数据的查询,或者,只是将主数据集群或从数据集群用作第一目标数据或第二目标数据的查询,在本申请实施例中,对此不作具体限定。

[0118] 例如,参照图2,若第1机房对应的第一数据集群为主数据集群,第2机房至第m机房为从数据集群。则第一数据表或第二数据表可以存储于第1机房至第m机房的任意一个或多个机房中。或者,第一数据表存储于第1机房中,第二数据表存储于第2机房至第m机房的任意一个或多个机房中。或者,第一数据表或第二数据表可以存储于第2机房至第m机房的任意一个或多个机房中。

[0119] 在本申请实施例中,参照图2,若上述第1机房至第m机房中,每个机房均可以存储有200亿条数据,每个机房均可以存储有10000个表,每个表中均可以存储200万条记录。每个机房对应一个数据集群,每个数据集群包括10个源数据库,每个源数据库可以对应一个服务器,每个源数据中可以存储1万个表,各个源数据库中存储的表的表名或编号,可以通过与10取余的方式进行快速确定。

[0120] 例如,参照图2,第1机房可以对应为1个数据集群,第1机房中服务器0至服务器9可以分别对应1个源数据库,每个源数据库分别存储1万个表。则,服务器0对应的源数据库中存储表名或编号,与10取余后为0的数据表,如数据表0、数据表10、数据表20……,数据表9990。服务器1对应的源数据库中存储表名或编号,与10取余后为1的数据表,如数据表1、数据表11、数据表21……,数据表9991。

[0121] 在本申请实施例中,各个源数据库中存储的表的表名或编号,通过与10取余的方式,可以快速确定,便于后续快速确定各个数据表对应的服务器编号或源数据库编号或标识等,进而可以根据表名或编号,即可快速确定其对应的服务器或源数据库。

[0122] 在本申请实施例中,可选的,在所述至少两个不同的数据集群包括:至少两个从数据集群的情况下,所述在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据之前,还可以包括:在所述至少两个从数据集群中,确定所述第一数据表和所述第二数据表。

[0123] 具体的,通常主数据集群具有一定的写数据压力,为了从较大程度上缓解主数据集群的处理压力,可以只在从数据集群中,确定第一数据表和第二数据表,进而后续对第一目标数据或第二目标数据的查找只是在从数据集群中进行,而不会在主数据集群中查找上述第一目标数据或第二目标数据,进而可以从很大程度上提升向主数据集群写入数据的效率。

[0124] 例如,参照图2所示,若第1机房对应的第一数据集群为主数据集群,第2机房至第m机房为从数据集群。从数据集群的数量可以大于等于2。通常主数据集群对应的第1机房具有一定的写数据压力,为了从较大程度上缓解主数据集群或第1机房的处理压力,可以只在第2机房至第m机房对应的从数据集群中,确定第一数据表和第二数据表,也就是说第一数据表或第二数据表可以为第2机房至第m机房的任意一个或多个机房中的数据表,并不在第1机房存储的多个数据表中确定第一数据表或第二数据表。进而后续对第一目标数据或第二目标数据的查找只是在第2机房至第m机房对应的从数据集群中进行,而不会在第1机房

对应的主数据集群中查找上述第一目标数据或第二目标数据,进而可以从很大程度上提升向第1机房对应的主数据集群写入数据的效率。

[0125] 如,若第1机房对应的第一数据集群为主数据集群,第2机房至第m机房为从数据集群,则确定的第一数据表可以为第2机房服务器0中的数据表0,第二数据表可以为第3机房服务器0中的数据表0……,第m机房服务器0中的数据表0。进而在第2机房服务器0中的数据表0的第一预设数据集合中,查找符合预设条件的第一目标数据,分别在第3机房服务器0中的数据表0……,第m机房服务器0中的数据表0的第二预设数据集合中,查找符合预设条件的第二目标数据。

[0126] 在本申请实施例中,可选的,若上述第一目标数据或第二目标数据对实时性要求较高,或,主数据集群的性能较优等,同样可以在上述主数据集群中确定第一数据表或第二数据表。在本申请实施例中,对此不作具体限定。

[0127] 在步骤S22中,基于所述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据。

[0128] 在本申请的实施例中,可以基于上述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据。如,可以同时基于上述预设数量的线程,在第一数据表的第一预设集合中查找符合预设条件的数据,得到第一目标数据,或者,从上述预设数量的线程中,选择部分线程,在第一数据表的第一预设集合中查找符合预设条件的数据,得到第一目标数据。在本申请实施例中,对此不作具体限定。

[0129] 例如,参照图2,若第一数据表的第一预设数据集合为:第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录。若第1机房服务器0中,进行数据查找独立的线程池被初始化为10个线程,则可以采用上述10个线程同时从第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录中查找第一目标数据。或者,可以从上述10个线程中挑选部分线程,从第1机房服务器0中的数据表0中存储的200万条数据中的前1000条数据记录中查找第一目标数据。

[0130] 在步骤S23中,基于所述预设数量的线程,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据。

[0131] 在本申请的实施例中,步骤S23可以参照上述的步骤S12的相关记载,为了避免重复此处不再赘述。需要说明的是,可以基于上述预设数量的线程,在第二数据表的第二预设数据集合中查找符合上述预设条件的数据,得到第二目标数据。

[0132] 例如,参照图2,针对上述例子,若第二数据表的第二预设数据集合为:第2机房的服务器0中的数据表0的第1000至第1999条数据记录、第3机房的服务器0中的数据表0的第2000至第2999条数据记录、第4机房的服务器0中的数据表0的第3000至第3999条数据记录……至第m机房的服务器0中的数据表0的第19000至第19999条数据记录,则第2机房的服务器0可以基于第2机房的服务器0的预设数量的线程,在第2机房的服务器0中的数据表0的第1000至第1999条数据记录中,查找符合预设条件的第二目标数据。第3机房的服务器0,可

以基于第3机房的服务器0的预设数量的线程,在第3机房的服务器0中的数据表0的第2000至第2999条数据记录中查找符合预设条件的第二目标数据。第4机房的服务器0,可以基于第4机房的服务器0的预设数量的线程,在第4机房的服务器0中的数据表0的第3000至第3999条数据记录中,查找符合预设条件的第二目标数据……,第m机房的服务器0,可以基于第m机房的服务器0的预设数量的线程,在第m机房的服务器0中的数据表0的第19000至第19999条数据记录中,查找符合预设条件的第二目标数据。也就是说,各个服务器可以基于预设数量的线程在第二数据表的第二预设集合,查找符合预设条件的第二目标数据,进而可以从很大程度上提升对第二目标数据的查找效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0133] 在本申请的实施例中,将每个服务器的线程池初始化为预设数量的线程,基于上述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据,基于上述预设数量的线程,在第二数据表的第二预设数据集合中查找符合预设条件的数据,得到第二目标数据,在每个服务器中,采用预设数量的线程查找第一目标数据或第二目标数据,进而可以从很大程度上提升对第一目标数据或第二目标数据的查找效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0134] 在本申请实施例中,为了避免重复查询,可选的,所述第一数据表的第一预设数据集合与所述第二数据表的第二预设数据集合的交集为0。

[0135] 具体的,第一数据表的第一预设数据集合与第二数据表的第二预设数据集合的交集为0,即,第一数据表的第一预设数据集合与第二数据表的第二预设数据集合中没有重复的数据记录。

[0136] 例如,针对上述例子,若从第一数据表的第一预设数据集合为:第1机房服务器0中的数据表0中前1000条数据记录。第二数据表的第二预设数据集合分别为:第2机房的服务器0中的数据表0的第1000至第1999条数据记录、第3机房的服务器0中的数据表0的第2000至第2999条数据记录、第4机房的服务器0中的数据表0的第3000至第3999条数据记录、……第m机房的服务器0中的数据表0的第19000至第19999条数据记录,则第1机房服务器0中的数据表0中前1000条数据记录、第2机房的服务器0中的数据表0的第1000至第1999条数据记录、第3机房的服务器0中的数据表0的第2000至第2999条数据记录、第4机房的服务器0中的数据表0的第3000至第3999条数据记录、……第m机房的服务器0中的数据表0的第19000至第19999条数据记录中,没有重复的数据记录,或各个数据均只被记录或存储一次。进而在上述第一预设数据集合和第二数据集合中,没有重复的数据记录,进而不会对重复的数据记录进行查询,可以提升第一目标数据和第二目标数据的同步效率。

[0137] 在本申请实施例中,第一数据表的第一预设数据集合与第二数据表的第二预设数据集合的交集为0,进而在上述第一预设数据集合和第二数据集合中,没有重复的数据记录,进而不会对重复的数据记录进行查询,可以提升第一目标数据和第二目标数据的同步效率。

[0138] 在本申请实施例中,可选的,所述第一预设数据集合中数据的数量,与各个所述第二预设数据集合中数据的数量均相同。

[0139] 具体的,在各个机房相应服务器中相同数量的数据记录中,查找第一目标数据或第二目标数据,进而各个机房或各个机房的服务器的查询任务相同,即,各个机房或各个机

房的服务器均分了查询任务或同步任务,不会导致某个机房或某个机房的服务器查询同步任务过重等,进而可以从一定程度上提升同步效率。

[0140] 例如,参照上述表1,在第1机房中,确定的第一预设数据集合中数据的数量为1000条,在第2机房至第m机房中确定的各个第二预设数据集合中数据的数量也为1000条,即,分别在第1机房至第m机房的各个机房中的1000条数据记录中,查找上述第一目标数据或第二目标数据,进而第1机房至第m机房的各个机房的查询任务相同,即,第1机房至第m机房的各个机房中均分了查询任务或同步任务,不会导致某个机房或某个机房的服务器查询同步任务过重等,进而可以从一定程度上提升同步效率。

[0141] 在本申请实施例中,可选的,若某些机房的性能更优等,该机房中第一预设数据集合中数据的数量,或第二预设数据集合中数据的数量可以相应设置的多一些。即,在性能较好的机房中较多的数据记录中,查询上述第一目标数据或第二目标数据,以缓解其它性能较差机房的同步压力,可以从一定程度上提升数据同步效率。在本申请实施例中,对此不作具体限定。

[0142] 在步骤S24中,将所述第一目标数据和所述第二目标数据同步至目标数据库。

[0143] 在本申请的实施例中,上述步骤S24可以参照上述的步骤S13的相关记载,为了避免重复此处不再赘述。

[0144] 本公开的实施例提供的技术方案可以包括以下有益效果:所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述方法包括:在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;将所述第一目标数据和所述第二目标数据同步至目标数据库。上述数据存储于各个不同机房中,在不同机房的相应的服务器中数据重复率较高的数据表中的不同预设数据集合中,查找符合预设条件的数据,也就是将查找操作在各个机房的相应服务器中进行了分解,每个机房的相应服务器中只查询了其存储数据表中的部分数据,并没有查询其存储数据表的全部数据,同时,第一预设数据集合和第二预设数据集合的并集,组成第一数据表或第二数据表所包括的全部数据,各个机房的相应服务器的查询结果形成了对该数据的查询结果,即查询结果是针对第一数据表的全部数据,或第二数据表的全部数据进行查询,查询全面没有遗漏;各个机房或各个机房的相应服务器分担了上述查询,一方面缓解了各个机房的压力,同时缓解了各个机房相应服务器的压力,不会占用各个机房或各个机房相应服务器较多进程,可以提升数据同步效率,另一方面各个服务器都是位于不同机房的独立机器,各个服务器使用独立的线程池进行数据查找等,从很大程度上提升了数据查找等的效率,进而可以大幅度减少数据查询、同步的时间,提升同步效率。

[0145] 图4是根据一示例性实施例示出的一种数据同步装置400,所述数据分别存储于至少两个数据表中,所述至少两个数据表分别存储于至少两个不同的服务器中,所述至少两个不同的服务器分别位于至少两个不同的机房中,所述装置400可以包括:

[0146] 第一目标数据查找模块402,被配置为在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

[0147] 第二目标数据查找模块403,被配置为在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据;其中,所述第一数据表和所述第二数据表为不同机房的相应服务器中,数据的重复率超过第一预设重复率的数据表;所述第一预设数据集合和所述第二预设数据集合数据的重复率小于第二预设重复率;所述第一预设数据集合和所述第二预设数据集合的并集,组成所述第一数据表或所述第二数据表所包括的全部数据;

[0148] 同步模块404,被配置为将所述第一目标数据和所述第二目标数据同步至目标数据库。

[0149] 可选的,在上述图4的基础上,参照图5,所述数据同步装置,还可以包括:

[0150] 线程初始化模块401,被配置为将每个所述服务器的线程池初始化为预设数量的线程;

[0151] 所述第一目标数据查找模块,可以包括:

[0152] 第一目标数据查找单元4021,被配置为基于所述预设数量的线程,在第一数据表的第一预设数据集合中查找符合预设条件的数据,得到第一目标数据;

[0153] 所述第二目标数据查找模块,可以包括:

[0154] 第二目标数据查找单元4031,被配置为基于所述预设数量的线程,在第二数据表的第二预设数据集合中查找符合所述预设条件的数据,得到第二目标数据。

[0155] 可选的,所述至少两个数据表分别存储于至少两个不同的源数据库中,所述至少两个不同的源数据库分别存储于至少两个不同的数据集群中,所述至少两个不同的数据集群分别位于至少两个不同的机房中,所述至少两个不同的数据集群分别包括:至少一个主数据集群和至少一个从数据集群,所述数据同步装置,还可以包括:

[0156] 主数据集群写入模块,配置为将所述数据写入所述至少一个主数据集群;

[0157] 从数据集群写入模块,配置为将所述数据从任意一个所述主数据集群同步至所述至少一个从数据集群。

[0158] 可选的,所述数据同步装置,在所述至少两个不同的数据集群包括:至少两个从数据集群的情况下,还可以包括:

[0159] 第一数据表和所述第二数据表确定模块,被配置为在所述至少两个从数据集群中,确定所述第一数据表和所述第二数据表。

[0160] 可选的,所述第一数据表的第一预设数据集合与所述第二数据表的第二预设数据集合的交集为0。

[0161] 可选的,所述第一预设数据集合中数据的数量,与各个所述第二预设数据集合中数据的数量均相同。

[0162] 关于上述实施例中的装置,其中各个单元执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0163] 图6是根据一示例性实施例示出的一种电子设备600的框图。例如,电子设备600可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。

[0164] 参照图6,电子设备600可以包括以下一个或多个组件:处理组件602,存储器604,电源组件606,多媒体组件608,音频组件610,输入/输出(I/O)的接口612,传感器组件614,以及通信组件616。

[0165] 处理组件602通常控制电子设备600的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件602可以包括一个或多个处理器620来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件602可以包括一个或多个模块,便于处理组件602和其他组件之间的交互。例如,处理组件602可以包括多媒体模块,以方便多媒体组件608和处理组件602之间的交互。

[0166] 存储器604被配置为存储各种类型的数据以支持在设备600的操作。这些数据的示例包括用于在电子设备600上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器604可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0167] 电源组件606为电子设备600的各种组件提供电力。电源组件606可以包括电源管理系统,一个或多个电源,及其他与为电子设备600生成、管理和分配电力相关联的组件。

[0168] 多媒体组件608包括在所述电子设备600和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件608包括一个前置摄像头和/或后置摄像头。当设备600处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0169] 音频组件610被配置为输出和/或输入音频信号。例如,音频组件610包括一个麦克风(MIC),当电子设备600处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器604或经由通信组件616发送。在一些实施例中,音频组件610还包括一个扬声器,用于输出音频信号。

[0170] I/O接口612为处理组件602和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0171] 传感器组件614包括一个或多个传感器,用于为电子设备600提供各个方面的状态评估。例如,传感器组件614可以检测到设备600的打开/关闭状态,组件的相对定位,例如所述组件为电子设备600的显示器和小键盘,传感器组件614还可以检测电子设备600或电子设备600一个组件的位置改变,用户与电子设备600接触的存在或不存在,电子设备600方位或加速/减速和电子设备600的温度变化。传感器组件614可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件614还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件614还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0172] 通信组件616被配置为便于电子设备600和其他设备之间有线或无线方式的通信。电子设备600可以接入基于通信标准的无线网络,如WiFi,运营商网络(如2G、3G、4G或5G),或它们的组合。在一个示例性实施例中,通信组件616经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件616还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RF标识)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0173] 在示例性实施例中,电子设备600可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述图1至图3的数据同步方法,且能达到相同的技术效果,为了避免重复此处不再赘述。

[0174] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器604,上述指令可由装置600的处理器620执行以完成上述图1至图3的数据同步方法,且能达到相同的技术效果,为了避免重复此处不再赘述。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0175] 图7是根据一示例性实施例示出的另一种电子设备700的框图。例如,电子设备700可以被提供为一服务器。参照图7,电子设备700包括处理组件722,其进一步包括一个或多个处理器,以及由存储器732所代表的存储器资源,用于存储可由处理组件722的执行的指令,例如应用程序。存储器732中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外,处理组件722被配置为执行指令,以执行上述图1至图3的数据同步方法,且能达到相同的技术效果,为了避免重复此处不再赘述。

[0176] 电子设备700还可以包括一个电源组件726被配置为执行电子设备700的电源管理,一个有线或无线网络接口770被配置为将电子设备700连接到网络,和一个输入输出(I/O)接口758。电子设备700可以操作基于存储在存储器732的操作系统,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™或类似。

[0177] 根据本公开实施例的一个方面,提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行上述图1至图3的数据同步方法,且能达到相同的技术效果,为了避免重复此处不再赘述。

[0178] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本申请的真正范围和精神由下面的权利要求指出。

[0179] 应当理解的是,本申请并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

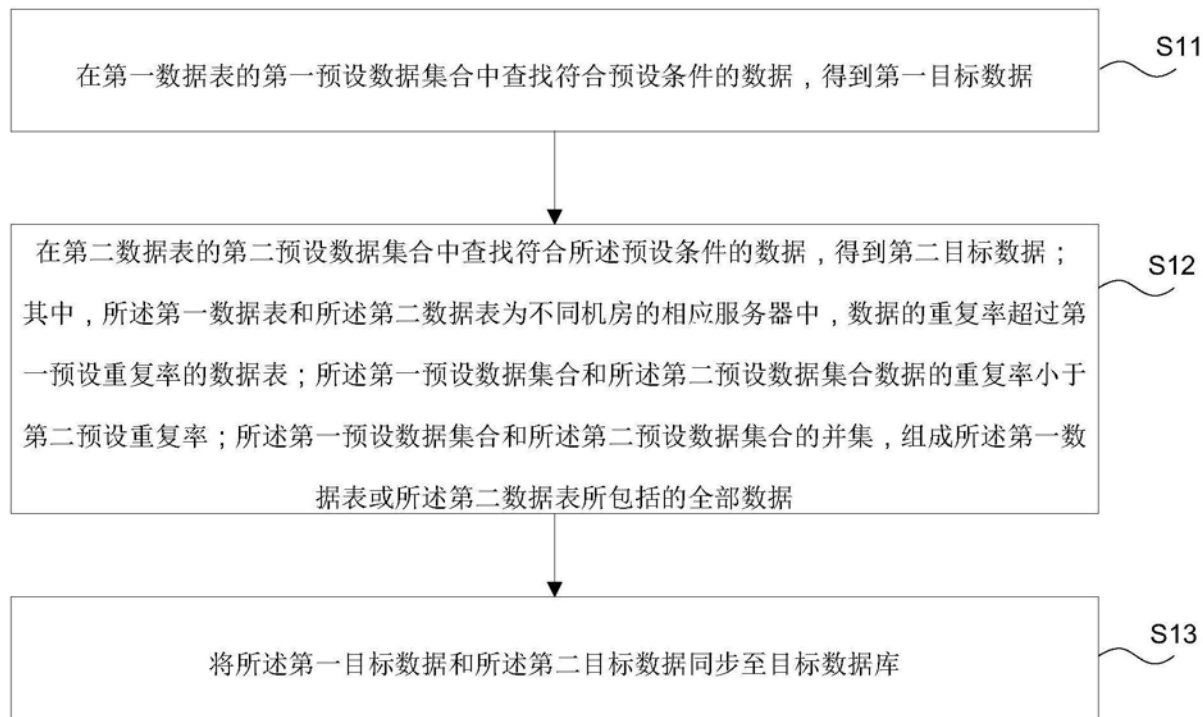


图1

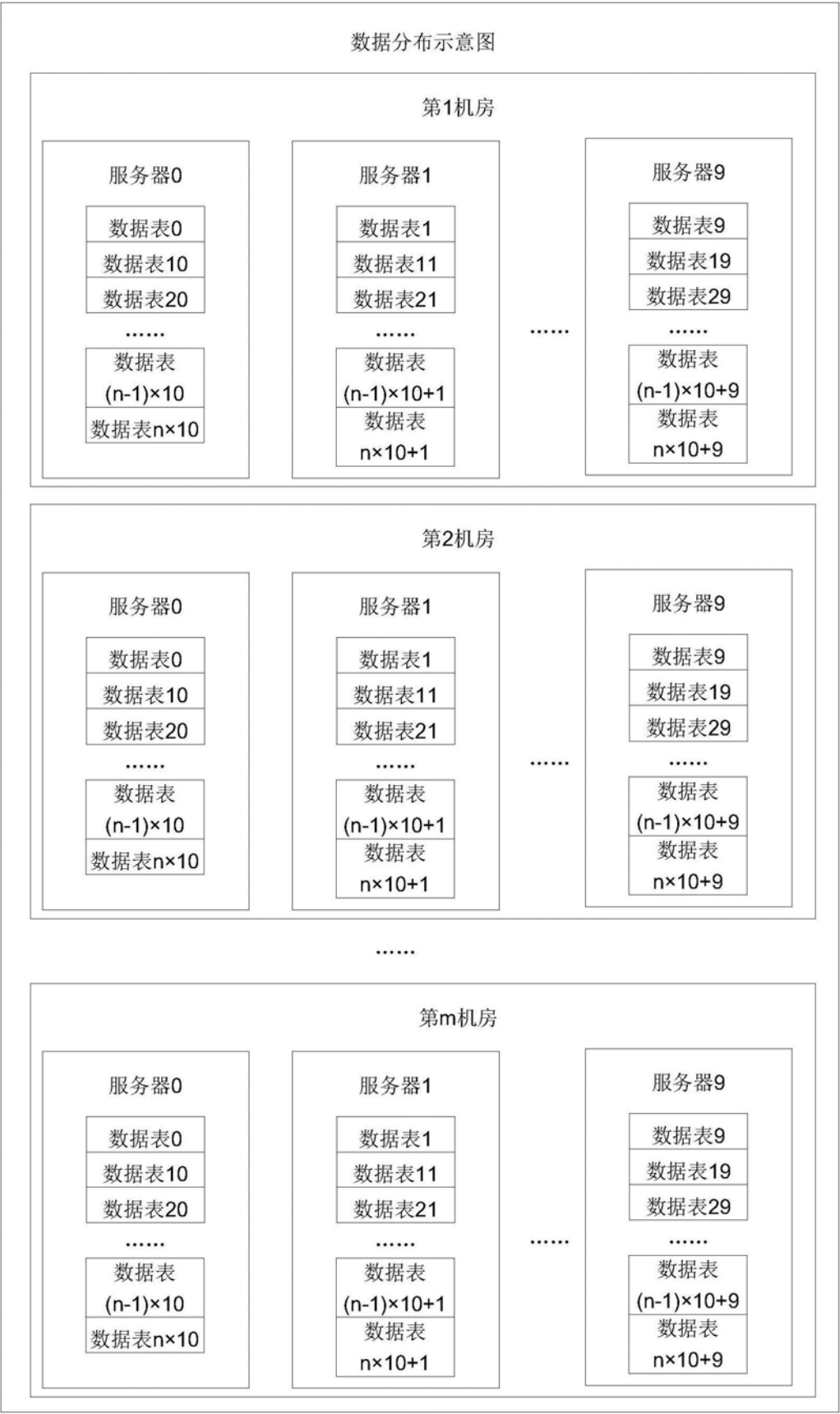


图2

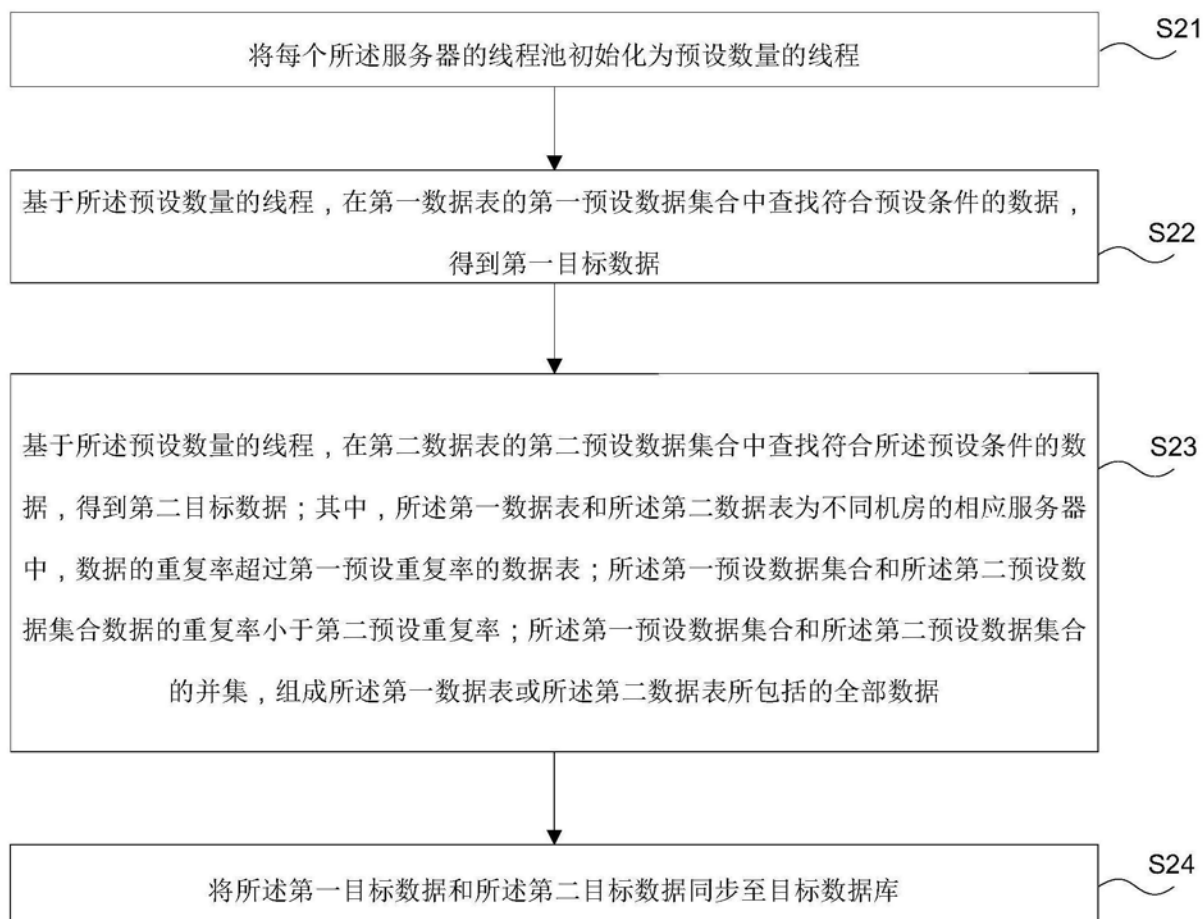


图3

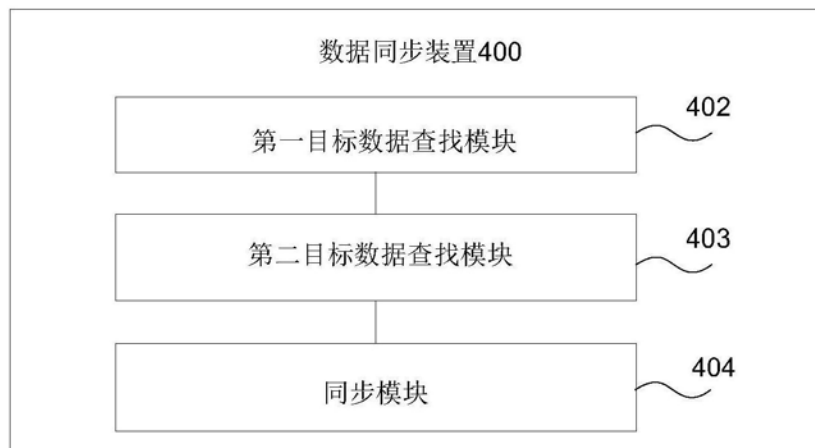


图4

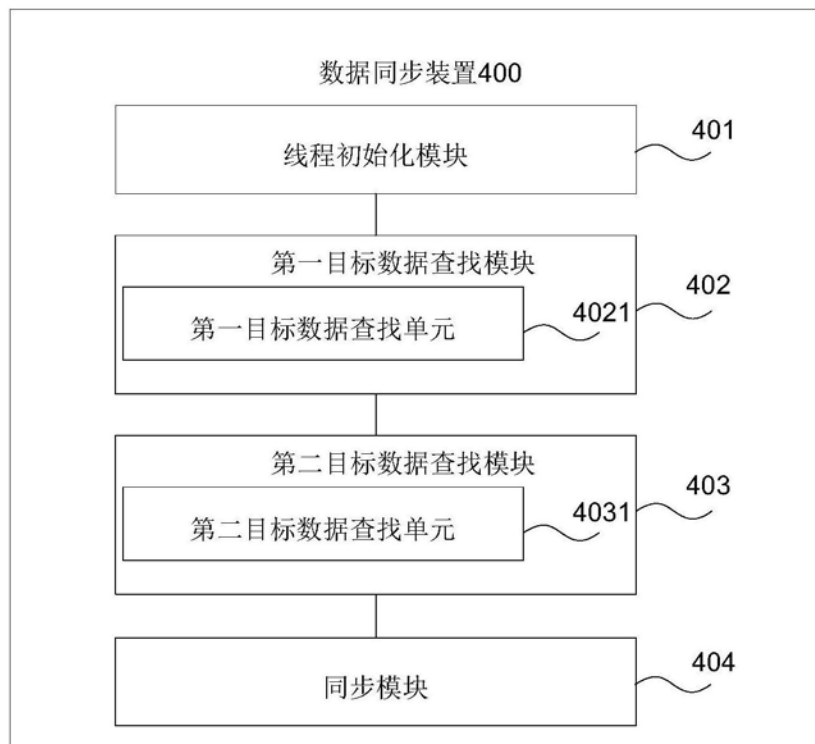


图5

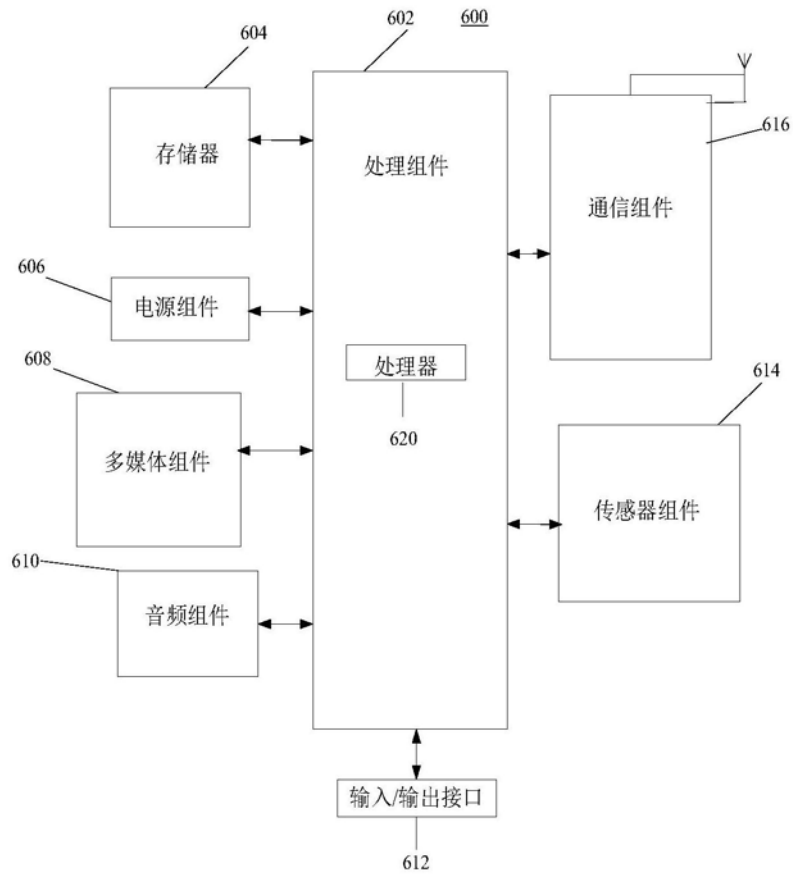


图6

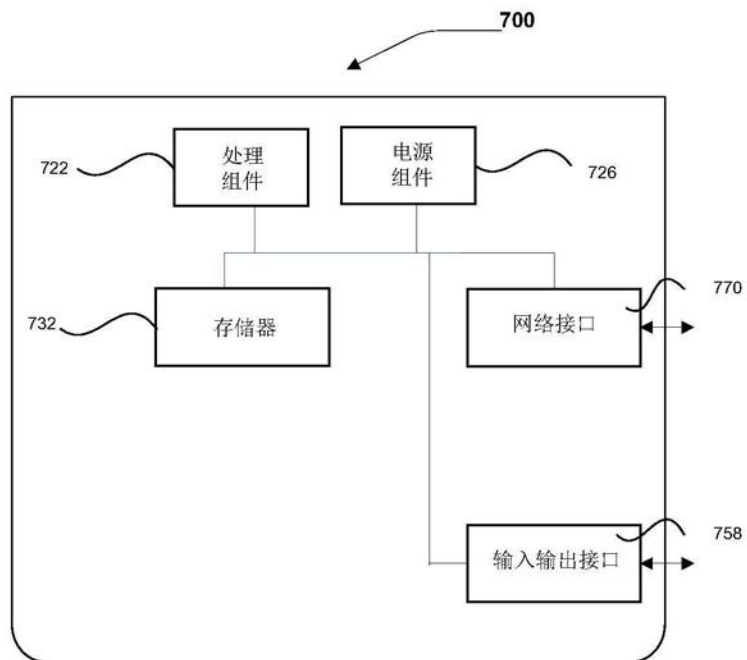


图7