



(12) 发明专利申请

(10) 申请公布号 CN 113672624 A

(43) 申请公布日 2021. 11. 19

(21) 申请号 202110981161.3

(22) 申请日 2021.08.25

(71) 申请人 中汽创智科技有限公司

地址 211100 江苏省南京市江宁区秣陵街
道胜利路88号

(72) 发明人 何旭 杨彦召 薛信钊 郭超
曹阳

(74) 专利代理机构 广州三环专利商标代理有限
公司 44202

代理人 郝传鑫 贾允

(51) Int. Cl.

G06F 16/22 (2019.01)

G06F 16/2455 (2019.01)

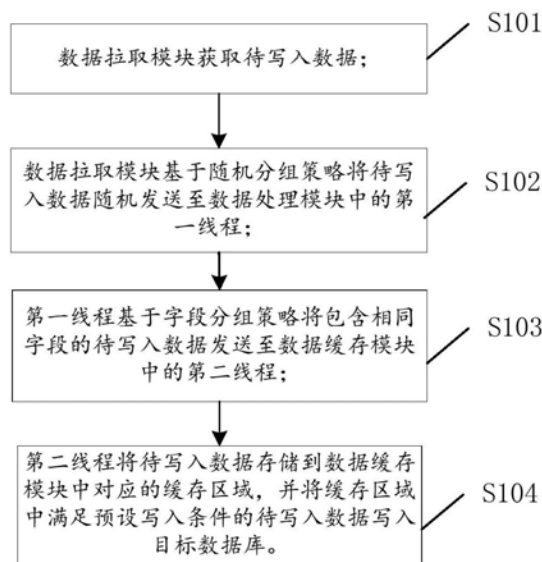
权利要求书2页 说明书6页 附图4页

(54) 发明名称

一种数据处理方法、系统及存储介质

(57) 摘要

本申请公开一种数据处理方法、系统及存储介质,该方法包括:由数据拉取模块获取待写入数据;所述数据拉取模块基于随机分组策略将所述待写入数据随机发送至数据处理模块中的第一线程;所述第一线程基于字段分组策略将包含相同字段的待写入数据发送至数据缓存模块中的第二线程;所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域,并将所述缓存区域中满足预设写入条件的待写入数据写入目标数据库,本申请采用缓存数据的预处理方式对待写入数据库的数据进行缓存,优化了写入数据库的流程,缓解了数据库写入数据的压力,提高了写入数据库的效率和稳定性。



1. 一种数据处理方法,其特征在于,应用于数据处理系统,所述系统包括数据拉取模块、数据处理模块和数据缓存模块,所述方法包括:

所述数据拉取模块获取待写入数据;

所述数据拉取模块基于随机分组策略将所述待写入数据随机发送至所述数据处理模块中的第一线程;

所述第一线程基于字段分组策略将包含相同字段的待写入数据发送至所述数据缓存模块中的第二线程;

所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域,并将所述缓存区域中满足预设写入条件的待写入数据写入目标数据库。

2. 根据权利要求1所述的一种数据处理方法,其特征在于,所述系统还包括线程监测模块,所述方法还包括:

所述线程监测模块周期性向所述第一线程和所述第二线程发送心跳数据包;

所述数据处理模块根据所述第一线程对所述心跳数据包的第一反馈,确定所述第一线程中待写入数据的处理状态;

所述数据缓存模块根据所述第二线程对所述心跳数据包的反馈,确定所述第二线程中待写入数据的处理状态。

3. 根据权利要求2所述的一种数据处理方法,其特征在于,所述线程监测模块周期性向所述第一线程和所述第二线程发送心跳数据包,包括:

所述线程监测模块基于全复制分组策略分别周期性的向所述数据处理模块中的各第一线程和所述数据缓存模块中各第二线程发送心跳数据包。

4. 根据权利要求3所述的一种数据处理方法,其特征在于,所述将所述缓存区域中满足预设写入条件的待写入数据写入目标数据库,包括:

判断所述缓存区域中与所述第二线程对应的待写入数据的数量是否大于等于存储量阈值;

若所述待写入数据的数量大于等于所述存储量阈值,则将所述数据缓存模块对应的第二线程中的所述待写入数据写入所述目标数据库。

5. 根据权利要求4所述的一种数据处理方法,其特征在于,在所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域之前,还包括:

获取当前待写入数据的目标约束字段;

判断所述缓存区域中的各待写入数据中是否存在所述目标约束字段对应的目标数据;

相应的,所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域包括:

若存在所述目标数据,所述第二线程基于所述当前待写入数据更新所述目标数据;

若不存在所述目标数据,所述第二线程将所述当前待写入数据存储至所述缓存区域。

6. 根据权利要求5所述的一种数据处理方法,其特征在于,在所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域之后,还包括:

所述数据缓存模块判断所述第二线程从接收到所述心跳数据包至得到所述第二反馈的时间间隔是否大于预设时间阈值;

若大于预设时间阈值,则将所述第二线程中的待写入数据写入所述目标数据库;

若不大于预设时间阈值,则执行判断所述缓存区域中与所述第二线程对应的待写入数据的数量是否大于等于存储量阈值的步骤。

7. 根据权利要求6所述的一种数据处理方法,其特征在于,在所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域之前,还包括:

所述数据缓存模块判断所述缓存区域中是否存在所述第二线程中的待写入数据对应的表信息;

若不存在所述表信息,则在所述缓存区域中生成与所述待写入数据相匹配的表信息;

若存在所述表信息,则基于所述待写入数据的表信息更新所述缓存区域中对应的表信息。

8. 根据权利要求7所述的一种数据处理方法,其特征在于,在所述数据缓存模块判断所述缓存区域中是否存在所述第二线程中的待写入数据对应的表信息之前,还包括:

所述数据缓存模块判断所述第二线程中的待写入数据的表信息是否为空;

若是,则丢弃所述第二线程中的待写入数据;

若否,则执行所述数据缓存模块判断所述缓存区域中是否存在所述第二线程中的待写入数据对应的表信息的步骤。

9. 一种数据处理系统,其特征在于,所述系统包括数据拉取模块、数据处理模块、数据缓存模块和线程监测模块;

所述数据拉取模块用于获取待写入数据,并基于随机分组策略将所述待写入数据随机发送至所述数据处理模块中的第一线程;

所述数据处理模块用于基于字段分组策略将包含相同字段的待写入数据发送至所述数据缓存模块中的第二线程;

所述数据缓存模块用于将所述待写入数据存储到所述数据缓存模块中对应的缓存区域,并将所述缓存区域中满足预设写入条件的待写入数据写入目标数据库;

所述线程监测模块用于周期性向所述第一线程和所述第二线程发送心跳数据包,以使所述数据处理模块根据所述第一线程对所述心跳数据包的第一反馈,确定所述第一线程中待写入数据的处理状态;所述数据缓存模块根据所述第二线程对所述心跳数据包的第二反馈,确定所述第二线程中待写入数据的处理状态。

10. 一种存储介质,其特征在于,所述存储介质中存储有至少一条指令或至少一段程序,所述至少一条指令或至少一段程序由处理器加载并执行以实现如权利要求1-8任一项所述的数据处理方法。

一种数据处理方法、系统及存储介质

技术领域

[0001] 本申请涉及数据库写入数据技术领域,特别涉及一种数据处理方法、系统及存储介质。

背景技术

[0002] 随着车联网技术应用不断深入,各类大数据业务系统对写入数据库的能力提出了更高的要求,一个坚强、可靠、快速的写入数据库的框架对于日益增长的数据处理业务显得尤为重要。

[0003] 由车联网系统的复杂性及业务要求决定,上送到系统的数据具有数量大、连续性高、数据量有突增现象、数据项分批不完整的可能性、数据有误、数据库字段类型不匹配,同时还可能存在由于网络通信的原因偶尔可能伴随有重复数据上送的现象,这些数据质量都加大了数据写入的难度。

[0004] 因此,需要一种数据处理的技术方案,解决现有技术中的数据库写入方法存在写入稳定性差、写入数据类型不匹配、难以适应数据由于复杂性而带来的写入数据库难,进而导致不能及时将待写入数据及时写入至目标数据库的问题。

发明内容

[0005] 为了解决现有技术的问题,本申请实施例提供了一种数据处理方法、系统及存储介质的技术方案,该技术方案如下所述:

[0006] 一方面,提供了一种数据处理方法,应用于数据处理系统,所述系统包括数据拉取模块、数据处理模块和数据缓存模块,所述方法包括:

[0007] 所述数据拉取模块获取待写入数据;

[0008] 所述数据拉取模块基于随机分组策略将所述待写入数据随机发送至所述数据处理模块中的第一线程;

[0009] 所述第一线程基于字段分组策略将包含相同字段的待写入数据发送至所述数据缓存模块中的第二线程;

[0010] 所述第二线程将所述待写入数据存储到所述数据缓存模块中对应的缓存区域,并将所述缓存区域中满足预设写入条件的待写入数据写入目标数据库。

[0011] 另一方面,提供了一种数据处理系统,所述系统包括数据拉取模块、数据处理模块、数据缓存模块和线程监测模块;

[0012] 所述数据拉取模块用于获取待写入数据,并基于随机分组策略将所述待写入数据随机发送至所述数据处理模块中的第一线程;

[0013] 所述数据处理模块用于基于字段分组策略将包含相同字段的待写入数据发送至所述数据缓存模块中的第二线程;

[0014] 所述数据缓存模块用于将所述待写入数据存储到所述数据缓存模块中对应的缓存区域,并将所述缓存区域中满足预设写入条件的待写入数据写入目标数据库;

[0015] 所述线程监测模块用于周期性向所述第一线程和所述第二线程发送心跳数据包,以使所述数据处理模块根据所述第一线程对所述心跳数据包的第一反馈,确定所述第一线程中待写入数据的处理状态;所述数据缓存模块根据所述第二线程对所述心跳数据包的反馈,确定所述第二线程中待写入数据的处理状态。

[0016] 另一方面提供了一种存储介质,所述存储介质中存储有至少一条指令或至少一段程序,所述至少一条指令或所述至少一段程序由处理器加载并执行以实现如上所述的数据处理方法。

[0017] 本申请提供了一种数据处理方法、系统及存储介质,具有如下技术效果:

[0018] 本申请实施例通过数据拉取模块从数据源中获取待写入数据,并将获取的待写入数据基于随机分组策略将待写入数据随机发送至数据处理模块中的第一线程,使得待写入数据可以均匀的分配给第一线程,数据处理模块中的第一线程基于字段分组策略将包含相同字段的待写入数据发送至数据缓存模块中的第二线程,此步骤可以避免记录锁的问题发生,数据缓存模块中的第二线程将待写入数据存储到数据缓存模块中对应的缓存区域,并将缓存区域中满足预设写入条件的待写入数据写入目标数据库,基于上述技术方案,本申请优化了写入数据库的流程,缓解了数据库写入数据的压力,提高了写入数据库的效率和稳定性。

附图说明

[0019] 为了更清楚地说明本申请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单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,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。

[0020] 图1为本申请实施例提供的一种数据处理方法的流程示意图;

[0021] 图2为本申请实施例提供的另一种数据处理方法的流程示意图;

[0022] 图3为本申请实施例提供的数据缓存模块中缓存区域的示意图;

[0023] 图4为本申请实施例提供的数据缓存模块中数据入库的处理流程示意图;

[0024] 图5为本申请实施例提供的一种数据处理系统的拓扑结构示意图。

具体实施方式

[0025] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0026] 需要说明的是,本说明书提供了如实施例或流程图所述的方法操作步骤,但基于常规或者无创造性的劳动可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式,不代表唯一执行顺序。在实际中的系统或服务产品执行时,可以按照实施例或者附图所示的方法顺序执行或者并行执行(例如并行处理器或者多线程处理的环境)。

[0027] 请参阅图1,其为本申请实施例提供的一种数据处理方法的流程示意图,该方法应

用于数据处理系统,所述系统包括数据拉取模块、数据处理模块和数据缓存模块,下面结合图1对数据处理方法进行详细的说明,该方法具体包括以下步骤:

[0028] S101:数据拉取模块获取待写入数据;

[0029] 在本申请实施例中,数据拉取模块从消息中间件kafka中获取待写入数据,其中数据拉取模块中的消息源Spout,从消息中间件kafka中加载数据,数据拉取模块对获取的待写入数据进行封装成数据处理模块可识别的数据格式,该数据格式至少包括表名、操作类型、索引标识等,具体的数据格式可根据具体情况进行添加,在此不做说明,并将封装后的数据格式发送给数据处理模块。

[0030] S102:数据拉取模块基于随机分组策略将待写入数据随机发送至数据处理模块中的第一线程;

[0031] 在本申请实施例中,数据拉取模块将封装后的待写入数据按照随机分组策略随机发送至数据处理模块中的第一线程,需要说明的是,随机分组策略可将待写入数据随机均等的分发至数据处理模块中的第一线程,使得每个线程中处理待写入数据的数量相同,其中,一个线程单独执行一个任务,这样使得第一线程对接收到的待写入数据可以最大化的利用并行度进行数据处理,进而减少了数据处理的时间。

[0032] S103:第一线程基于字段分组策略将包含相同字段的待写入数据发送至数据缓存模块中的第二线程;

[0033] 在本申请实施例中,数据处理模块将第一线程中的待写入数据进行处理得到数据缓存模块可识别的数据格式,该数据格式中所包含的关键信息如:表名、数据接收时间、操作类型、数据操作时间、表分区字段、索引标识和分组字段,第一线程基于字段分组策略将字段相同的待写入数据发送给数据缓存模块中的第二线程,在一个具体的实施例中,采用车辆ID号及时间联动为分组字段,将具有相同字段的车辆ID号及时间联动的分组字段发送给第二线程,这一步骤的实施能够避免不同线程处理同一车辆信息和同一时间联动的信息,进而解决了记录锁的问题,确保了数据的可靠传输。

[0034] 在一个可选的实施方式中,如图2所示,其为本申请实施例提供的另一种数据处理方法的流程示意图,数据处理系统还可以包括线程监测模块,该模块在数据处理方法中的详细步骤如下:

[0035] S201:线程监测模块周期性向第一线程和第二线程发送心跳数据包;

[0036] 其中,心跳数据包用于对第一线程和第二线程提供激励,避免一段时间内数据过少时,同样也可以将待写入数据及时写入至目标数据库。

[0037] S202:数据处理模块根据第一线程对心跳数据包的第一反馈,确定第一线程中待写入数据的处理状态;

[0038] S203:数据缓存模块根据第二线程对心跳数据包的第二反馈,确定第二线程中待写入数据的处理状态。

[0039] 在本申请实施例中,线程监测模块基于全复制分组策略分别周期性的向数据处理模块中的各第一线程和数据缓存模块中各第二线程发送心跳数据包,进而使得第一线程中的各第一线程均可以接收到心跳数据包,激励第一线程中的待写入数据进行处理,数据处理模块根据第一线程对心跳数据包的第一反馈,该第一反馈可以由数据格式中的操作时间以及接收时间来进行确定,确定第一线程中待写入数据的处理状态是传输状态,还是在预

设时间内都接收不到数据的传输,进而与缓存区域中的第二线程接收的心跳数据包反馈的第二反馈相结合,确定数据缓存模块中的待写入数据是否写入目标数据库。

[0040] 同样的,第二线程中的各第二线程也均可以接收到心跳数据包,激励第二线程中的待写入数据进行处理,数据缓存模块根据第二线程对心跳数据包的第二反馈,该第二反馈也可以由数据格式中的操作时间以及接收时间来确定,根据缓存区域中的第二线程接收的心跳数据包反馈的第二反馈确定数据缓存模块中的待写入数据是否写入目标数据库。

[0041] S104:第二线程将待写入数据存储到数据缓存模块中对应的缓存区域,并将缓存区域中满足预设写入条件的待写入数据写入目标数据库。

[0042] 在本申请实施例中,数据缓存模块与目标数据库具有类似的功能,该数据缓存模块具备数据库的特性,可对输入的待写入数据实现增加、删除、修改和索引等功能,通过数据缓存模块中的缓存区域实现对第二线程中的待写入数据进行缓存,在待写入数据进行写入数据库之前提前对待写入数据进行删除无效数据、更新待写入数据和检索是否有约束字段相同的数据等操作,以便于能够最大限度利用CPU,确保写入数据库中的待写入数据的完整性和有效性,进而提高待写入数据写入目标数据库的效率。

[0043] 具体的,数据缓存模块中的缓存区域如图3所示,其为本申请实施例提供的数据缓存模块中缓存区域的示意图,其中,TableInfoCache为缓存待写入数据中的表名、索引标识等数据格式信息的缓存区域,ColInfo为缓存待写入数据中的字段信息,ValueInfoCache为缓存待写入数据的缓存区域。

[0044] 在一个可选的实施方式中,请参阅图4,其为本申请实施例提供的数据缓存模块中数据入库的处理流程示意图,通过判断缓存区域中与第二线程对应的待写入数据的数量是否大于等于存储量阈值,来确定是否将缓冲区域中的待写入数据写入数据库,若待写入数据的数量大于等于存储量阈值,则将数据缓存模块对应的线程中的待写入数据写入目标数据库,若待写入数据的数量小于存储量阈值,则继续对缓存区域中的待写入数据进行缓存,直至满足预设的写入条件,再将缓存区域中的待写入数据写入目标数据库,这在一定程度上缓解了大量数据同时写入数据的压力,提高了写入数据库的效率。

[0045] 在一个可选的实施方式中,继续参见图4,在第二线程将待写入数据存储到数据缓存模块中对应的缓存区域之前,还可以包括:

[0046] 获取当前待写入数据的目标约束字段,通过判断缓存区域中的各待写入数据中是否存在目标约束字段对应的目标数据,来确定第二线程将待写入数据是否直接存储到数据缓存模块中对应的缓存区域,若存在目标数据,第二线程基于当前待写入数据更新目标数据,若不存在目标数据,第二线程将当前待写入数据存储至缓存区域,该执行步骤可以避免将重复数据写入缓存区域,同时对同一目标约束字段的目标数据添加在缓存区域,这在一定程度上提高了数据入库的可靠性。

[0047] 在一个可选的实施方式中,继续参见图4,在第二线程将待写入数据存储到数据缓存模块中对应的缓存区域之后,还可以包括:

[0048] 数据缓存模块通过判断缓存区域中的第二线程从接收到心跳数据包至得到第二反馈的时间间隔是否大于预设时间阈值,若大于预设时间阈值,则将第二线程中的待写入数据写入目标数据库,若不大于预设时间阈值,则执行判断缓存区域中与第二线程对应的待写入数据的数量是否大于等于存储量阈值的步骤,该执行步骤避免了当在缓存区域中的

待写入数据数量过少时,不能将缓存区域中的待写入数据写入目标数据库的情况,这在一定程度上使得数据入库具有完备性和通用性。

[0049] 在一个可选的实施方式中,继续参见图4,在第二线程将待写入数据存储到数据缓存模块中对应的缓存区域之前,还可以包括:

[0050] 通过数据缓存模块判断缓存区域中是否存在第二线程中的待写入目标数据对应的表信息,若不存在表信息,则在缓存区域中生成与第二线程中的待写入数据相匹配的表信息,若存在表信息,则基于第二线程中的待写入数据的表信息更新缓存区域中对应的表信息,具体的,在数据缓存模块中,对各个第二线程中的待写入目标数据对应的表信息都需要进行识别,若判断缓存区域中不存在第二线程中的待写入目标数据,则在缓存区域中生成与第二线程中待写入数据相匹配的表信息,以便于当下一个待写入目标数据的表信息与存储的表信息相一致时,则不需要对该表信息进行生成,可直接基于第二线程中的待写入数据的表信息更新缓存区域中对应的表信息,该执行步骤可避免当待写入数据写入数据库时进行多次连接查询,而导致数据写入的性能降低等情况。

[0051] 在一个可选的实施方式中,继续参见图4,在数据缓存模块判断缓存区域中是否存在第二线程中的待写入目标数据对应的表信息之前,还可以包括:

[0052] 数据缓存模块判断第二线程中的待写入目标数据的表信息是否为空,若是,则丢弃第二线程中的待写入目标数据,若否,则执行数据缓存模块判断缓存区域中是否存在第二线程中的待写入目标数据对应的表信息的步骤,该执行步骤可在缓存区域中对无效数据的删除,这在一定程度上提高了写入数据库的效率。

[0053] 由本申请实施例的上述技术方案可见,本申请实施例通过数据拉取模块从数据源中获取待写入数据,并将获取的待写入数据基于随机分组策略将待写入数据随机发送至数据处理模块中的第一线程,使得待写入数据可以均匀的分配给第一线程,数据处理模块中的第一线程基于字段分组策略将包含相同字段的待写入数据发送至数据缓存模块中的第二线程,此步骤可以避免记录锁的问题发生,数据缓存模块中的第二线程将待写入数据存储到数据缓存模块中对应的缓存区域,并将缓存区域中满足预设写入条件的待写入数据写入目标数据库,基于上述技术方案,本申请优化了写入数据库的流程,缓解了数据库写入数据的压力,提高了写入数据库的效率和稳定性。

[0054] 本申请实施例中还提供了一种数据处理系统,如图5所示,其为本申请实施例提供了一种数据处理系统的拓扑结构示意图,其中箭头方向为数据的流向,该系统包括数据拉取模块10、数据处理模块20、数据缓存模块30和线程监测模块40;

[0055] 数据拉取模块10用于获取待写入数据,并基于随机分组策略将待写入数据随机发送至数据处理模块20中的第一线程;

[0056] 数据处理模块20用于基于字段分组策略将包含相同字段的待写入数据发送至数据缓存模块30中的第二线程;

[0057] 数据缓存模块30用于将待写入数据存储到数据缓存模块30中对应的缓存区域,并将缓存区域中满足预设写入条件的待写入数据写入目标数据库;

[0058] 线程监测模块40用于周期性向第一线程和第二线程发送心跳数据包,以使数据处理模块20根据第一线程对心跳数据包的第一反馈,确定第一线程中待写入数据的处理状态,数据缓存模块30根据第二线程对心跳数据包的第二反馈,确定第二线程中待写入数据

的处理状态。

[0059] 关于上述实施例中的系统,其中各个模块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。

[0060] 本申请实施例还提供了一种存储介质,该存储介质中存储有至少一条指令或至少一段程序,该至少一条指令或至少一段程序由处理器加载并执行以实现上述方法实施例提供的数据处理方法。

[0061] 可选地,在本实施例中,上述存储介质可以位于网络的多个网络服务器中的至少一个网络服务器。可选地,在本实施例中,上述存储介质可以包括但不限于:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0062] 需要说明的是:上述本申请实施例先后顺序仅仅为了描述,不代表实施例的优劣。且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。

[0063] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应所述以权利要求的保护范围为准。

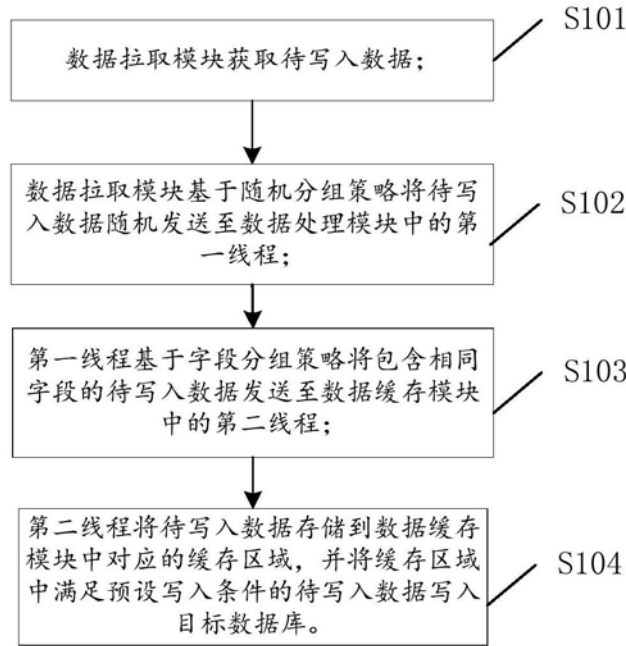


图1

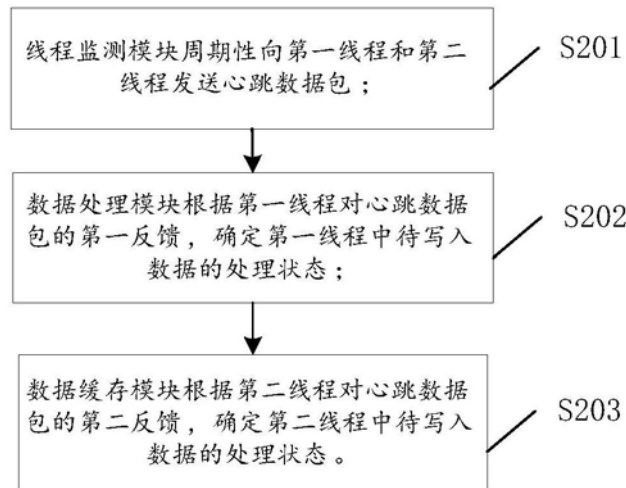


图2

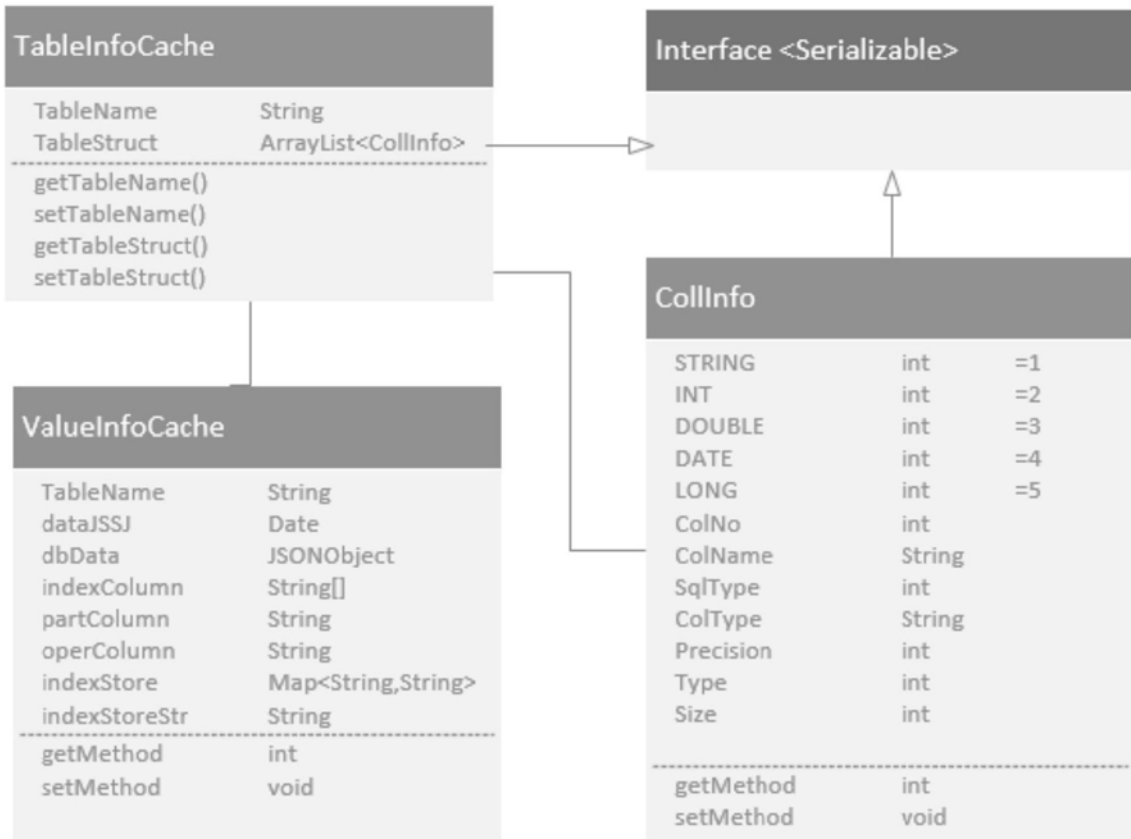


图3

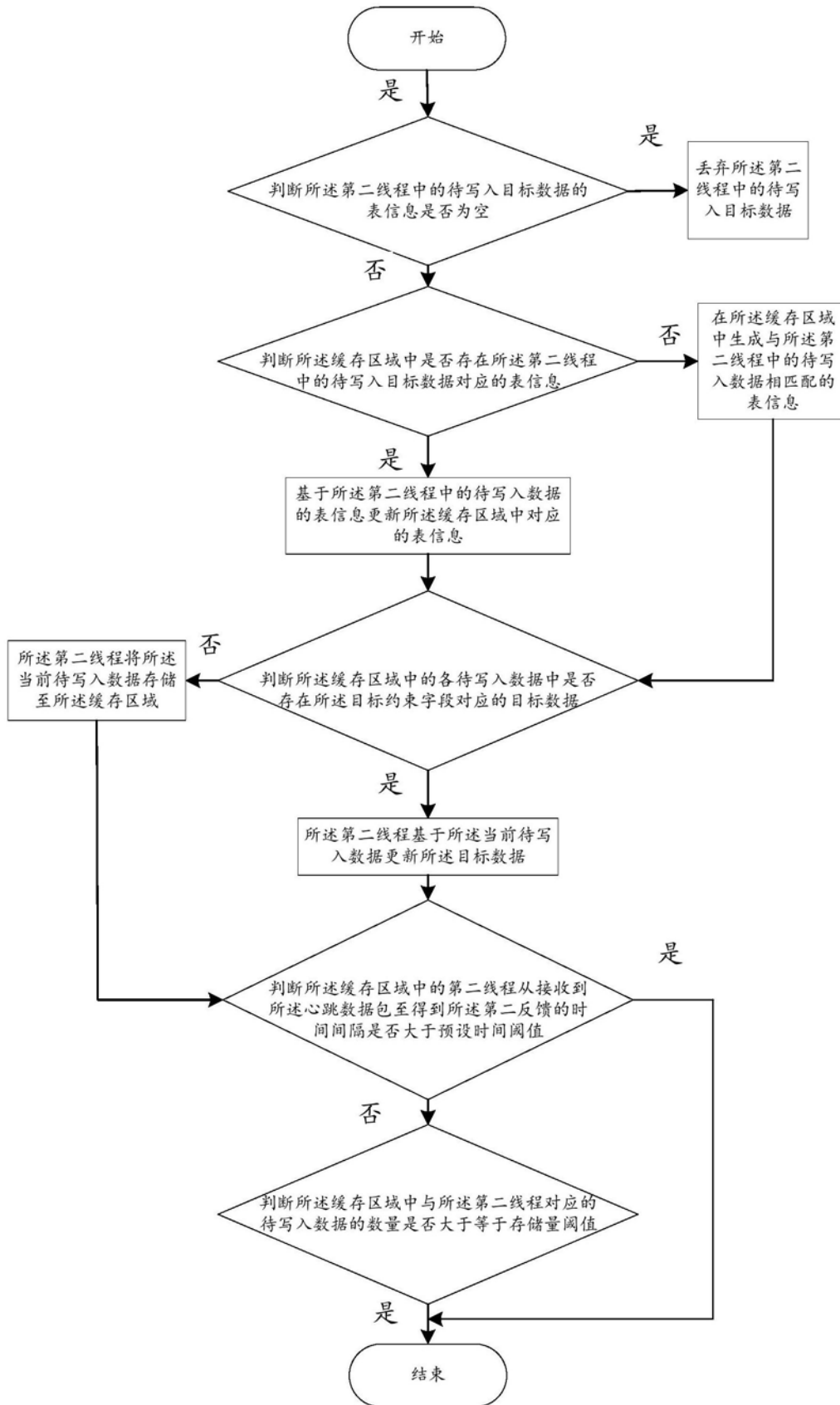


图4

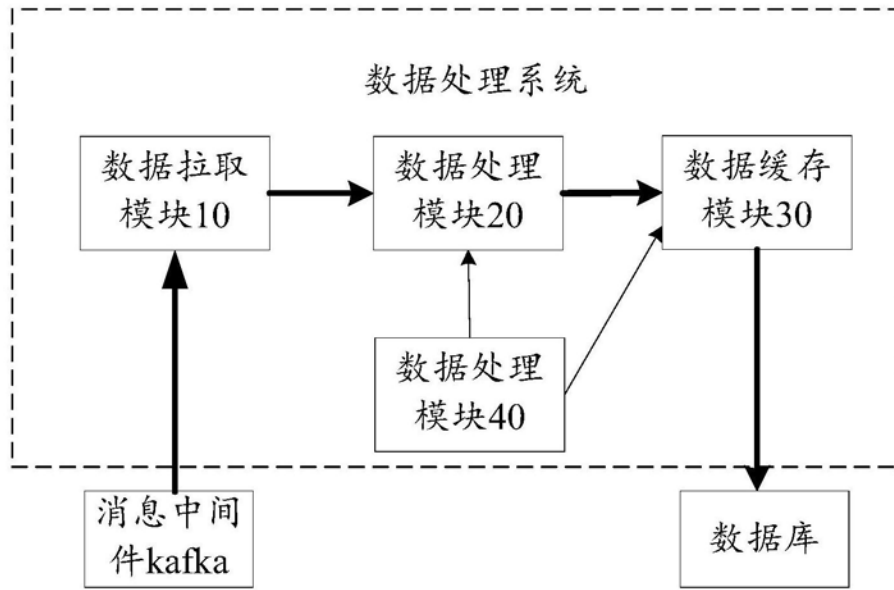


图5