



(12) 发明专利申请

(10) 申请公布号 CN 114860821 A

(43) 申请公布日 2022. 08. 05

(21) 申请号 202110164536.7

(22) 申请日 2021.02.05

(71) 申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

(72) 发明人 李本利 于东海 孙仕杰 黄兴
林志恒 高品 王洋子豪 吴文强

(74) 专利代理机构 深圳市联鼎知识产权代理有
限公司 44232

专利代理师 叶虹

(51) Int. Cl.

G06F 16/25 (2019.01)

G06F 16/215 (2019.01)

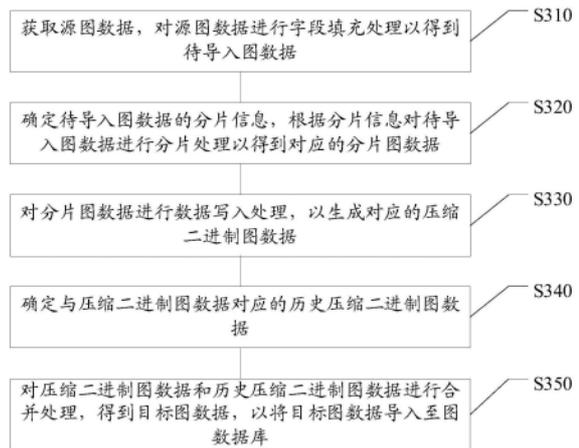
权利要求书3页 说明书17页 附图7页

(54) 发明名称

图数据库的数据导入方法及装置、存储介
质、电子设备

(57) 摘要

本公开提供一种图数据库的数据导入方法、
装置、电子设备和计算机可读存储介质;涉及数
据处理技术领域。该方法包括:获取源图数据,对
源图数据进行字段填充处理以得到待导入图数
据;确定待导入图数据的分片信息,根据分片信
息对待导入图数据进行分片处理以得到对应的
分片图数据;对分片图数据进行数据写入处理,
以生成对应的压缩二进制图数据;确定与压缩二
进制图数据对应的历史压缩二进制图数据;对压
缩二进制图数据和历史压缩二进制图数据进行
合并处理,得到目标图数据,以将目标图数据导
入至图数据库。本公开可以大幅降低图数据库处
理器的处理耗时,同时有效提高数据导入速度。



1. 一种图数据库的数据导入方法,其特征在于,包括:
 - 获取源图数据,对所述源图数据进行字段填充处理以得到待导入图数据;
 - 确定所述待导入图数据的分片信息,根据所述分片信息对所述待导入图数据进行分片处理以得到对应的分片图数据;
 - 对所述分片图数据进行数据写入处理,以生成对应的压缩二进制图数据;
 - 确定与所述压缩二进制图数据对应的历史压缩二进制图数据;
 - 对所述压缩二进制图数据和所述历史压缩二进制图数据进行合并处理,得到目标图数据,以将所述目标图数据导入至所述图数据库。
2. 根据权利要求1所述的方法,其特征在于,所述获取源图数据,包括:
 - 确定初始源数据,从所述初始源数据中确定顶点数据;
 - 从所述初始源数据中确定边数据;
 - 将所述顶点数据和所述边数据作为所述源图数据。
3. 根据权利要求1所述的方法,其特征在于,所述对所述源图数据进行字段填充处理以得到待导入图数据,包括:
 - 从图数据库中获取所述源图数据对应的元数据;
 - 根据所述元数据对所述源图数据进行字段校验处理,将通过所述字段校验处理的源图数据作为初始图数据;
 - 对所述初始图数据进行字段填充处理,以生成所述待导入图数据。
4. 根据权利要求3所述的方法,其特征在于,所述根据所述元数据对所述源图数据进行字段校验处理,包括:
 - 根据所述元数据确定参考字段类型和参考字段数量;
 - 确定所述源图数据的字段类型和所述源图数据的字段数量;
 - 确定所述源图数据的字段类型与所述参考字段类型是否一致,并确定所述源图数据的字段数量与所述参考字段数量是否一致。
5. 根据权利要求3所述的方法,其特征在于,所述初始图数据包括初始顶点数据和初始边数据,所述对所述初始图数据进行字段填充处理,以生成所述待导入图数据,包括:
 - 确定与所述初始顶点数据对应的补充顶点字段;
 - 确定与所述初始边数据对应的补充边字段;
 - 根据所述补充顶点字段和所述补充边字段对所述初始图数据进行所述字段填充处理,以生成所述待导入图数据。
6. 根据权利要求1所述的方法,其特征在于,所述待导入图数据包括待导入顶点数据和待导入边数据,所述分片信息包括顶点分片标识和边分片标识;
 - 所述确定所述待导入图数据的分片信息,包括:
 - 确定所述待导入图数据对应的目标顶点标识和目标起始顶点标识;
 - 获取所述图数据库对应的分片数;
 - 根据所述分片数和所述目标顶点标识确定所述待导入顶点数据所对应分片的标识,作为所述顶点分片标识;
 - 根据所述分片数和所述目标起始顶点标识确定所述待导入边数据所对应分片的标识,作为所述边分片标识。

7. 根据权利要求1所述的方法,其特征在于,所述对所述分片图数据进行数据写入处理,以生成对应的压缩二进制图数据,包括:

确定所述分片图数据中的分片顶点数据和分片边数据;

分别对所述分片顶点数据和所述分片边数据进行排序处理,以得到对应的分片排序顶点数据和分片排序边数据;

对所述分片排序顶点数据和所述分片排序边数据进行顺序写入处理,以生成所述压缩二进制图数据。

8. 根据权利要求7所述的方法,其特征在于,所述分别对所述分片顶点数据和所述分片边数据进行排序处理,以得到对应的分片排序顶点数据和分片排序边数据,包括:

确定所述分片顶点数据对应的顶点排序字段;

根据所述顶点排序字段对所述分片顶点数据进行排序处理,以生成所述分片排序顶点数据;

确定所述分片边数据对应的边排序字段;

根据所述边排序字段对所述分片边数据进行排序处理,以生成所述分片排序边数据。

9. 根据权利要求1所述的方法,其特征在于,所述对所述压缩二进制图数据和所述历史压缩二进制图数据进行合并处理,得到目标图数据,包括:

对所述压缩二进制图数据和所述历史压缩二进制图数据进行逻辑合并处理,以生成逻辑合并数据;

对所述逻辑合并数据进行物理合并去重处理,以生成所述目标图数据。

10. 根据权利要求9所述的方法,其特征在于,在所述对所述逻辑合并数据进行物理合并去重处理之前,所述方法还包括:

响应于数据查询请求,确定所述数据查询请求中的请求数据标识;

确定与所述请求数据标识对应的初始请求数据,并获取所述初始请求数据的时间戳信息;

根据所述时间戳信息从所述初始请求数据中确定目标请求数据,并返回所述目标请求数据。

11. 根据权利要求10所述的方法,其特征在于,所述请求数据标识包括请求顶点标识,所述初始请求数据包括初始请求顶点数据;

所述根据所述时间戳信息从所述初始请求数据中确定目标请求数据,包括:

根据所述请求顶点标识确定所述初始请求顶点数据,并获取所述初始请求顶点数据对应的顶点时间戳信息;

基于所述顶点时间戳信息从所述初始请求顶点数据确定最新顶点数据,将所述最新顶点数据作为所述目标请求数据。

12. 根据权利要求10所述的方法,其特征在于,所述请求数据标识包括请求边标识,所述初始请求数据包括初始请求边数据;

所述根据所述时间戳信息从所述初始请求数据中确定目标请求数据,包括:

根据所述请求边标识确定所述初始请求边数据,并获取所述初始请求边数据对应的边时间戳信息;

基于所述边时间戳信息从所述初始请求边数据确定最新边数据,将所述最新边数据作

为所述目标请求数据。

13. 一种图数据库的数据导入装置,其特征在于,包括:

数据获取模块,用于获取源图数据,对所述源图数据进行字段填充处理以得到待导入图数据;

数据分片模块,用于确定所述待导入图数据的分片信息,根据所述分片信息对所述待导入图数据进行分片处理以得到对应的分片图数据;

压缩数据生成模块,用于对所述分片图数据进行数据写入处理,以生成对应的压缩二进制图数据;

历史数据确定模块,用于确定与所述压缩二进制图数据对应的历史压缩二进制图数据;

数据导入模块,用于对所述压缩二进制图数据和所述历史压缩二进制图数据进行合并处理,得到目标图数据,以将所述目标图数据导入至所述图数据库。

14. 一种电子设备,其特征在于,包括:

处理器;以及

存储器,所述存储器上存储有计算机可读指令,所述计算机可读指令被所述处理器执行时实现如权利要求1至12中任意一项所述的图数据库的数据导入方法。

15. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至12中任意一项所述的图数据库的数据导入方法。

图数据库的数据导入方法及装置、存储介质、电子设备

技术领域

[0001] 本公开涉及数据处理技术领域,具体而言,涉及一种图数据库的数据导入方法、图数据库的数据导入装置、电子设备以及计算机可读存储介质。

背景技术

[0002] 随着社交、电商、物联网等行业的快速发展,在不同的应用场景下,形成了一个庞大而复杂的关系网,采用传统数据库难以处理上述应用场景的关系运算;并且大数据行业需要处理的数据之间的关系随数据量呈几何级数增长,亟需一种支持海量复杂数据关系运算的数据库,因此图形数据库应运而生。

[0003] 图形数据库,简称图数据库,是一种非关系型数据库,它可以应用图形理论存储实体之间的关系信息。图数据库可以应用于知识图谱构建、安全风控、社交推荐以及链路跟踪等多种场景中。

[0004] 需要说明的是,在上述背景技术部分公开的信息仅用于加强对本公开的背景的理解,因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

发明内容

[0005] 本公开的目的在于提供一种图数据库的数据导入方法、图数据库的数据导入装置、电子设备以及计算机可读存储介质,进而在一定程度上克服现有技术中图数据库的数据导入速度较慢、图数据库处理器的耗时较高的问题。

[0006] 根据本公开的一个方面,提供一种图数据库的数据导入方法,包括:获取源图数据,对源图数据进行字段填充处理以得到待导入图数据;确定待导入图数据的分片信息,根据分片信息对待导入图数据进行分片处理以得到对应的分片图数据;对分片图数据进行数据写入处理,以生成对应的压缩二进制图数据;确定与压缩二进制图数据对应的历史压缩二进制图数据;对压缩二进制图数据和历史压缩二进制图数据进行合并处理,得到目标图数据,以将目标图数据导入至图数据库。

[0007] 根据本公开的一个方面,提供一种图数据库的数据导入装置,包括:数据获取模块,用于获取源图数据,对源图数据进行字段填充处理以得到待导入图数据;数据分片模块,用于确定待导入图数据的分片信息,根据分片信息对待导入图数据进行分片处理以得到对应的分片图数据;压缩数据生成模块,用于对分片图数据进行数据写入处理,以生成对应的压缩二进制图数据;历史数据确定模块,用于确定与压缩二进制图数据对应的历史压缩二进制图数据;数据导入模块,用于对压缩二进制图数据和历史压缩二进制图数据进行合并处理,得到目标图数据,以将目标图数据导入至图数据库。

[0008] 在本公开的一种示例性实施例中,数据获取模块包括源数据获取单元,上述源数据获取单元被配置为:确定初始源数据,从初始源数据中确定顶点数据;从初始源数据中确定边数据;将顶点数据和边数据作为源图数据。

[0009] 在本公开的一种示例性实施例中,数据获取模块还包括:元数据获取单元,用于从

图数据库中获取源图数据对应的元数据;校验处理单元,用于根据元数据对源图数据进行字段校验处理,将通过字段校验处理的源图数据作为初始图数据;填充处理单元,用于对初始图数据进行字段填充处理,以生成待导入图数据。

[0010] 在本公开的一种示例性实施例中,校验处理单元被配置为:根据元数据确定参考字段类型和参考字段数量;确定源图数据的字段类型和源图数据的字段数量;确定源图数据的字段类型与参考字段类型是否一致,并确定源图数据的字段数量与参考字段数量是否一致。

[0011] 在本公开的一种示例性实施例中,填充处理单元被配置为:确定与初始顶点数据对应的补充顶点字段;确定与初始边数据对应的补充边字段;根据补充顶点字段和补充边字段对初始图数据进行字段填充处理,以生成待导入图数据。

[0012] 在本公开的一种示例性实施例中,数据分片模块包括分片信息确定单元,上述分片信息确定单元被配置为:确定待导入图数据对应的目标顶点标识和目标起始顶点标识;获取图数据库对应的分片数;根据分片数和目标顶点标识确定待导入顶点数据所对应分片的标识,作为顶点分片标识;根据分片数和目标起始顶点标识确定待导入边数据所对应分片的标识,作为边分片标识。

[0013] 在本公开的一种示例性实施例中,压缩数据生成模块包括:分片数据确定单元,用于确定分片图数据中的分片顶点数据和分片边数据;数据排序单元,用于分别对分片顶点数据和分片边数据进行排序处理,以得到对应的分片排序顶点数据和分片排序边数据;压缩数据生成单元,用于对分片排序顶点数据和分片排序边数据进行顺序写入处理,以生成压缩二进制图数据。

[0014] 在本公开的一种示例性实施例中,数据排序单元被配置为:确定分片顶点数据对应的顶点排序字段;根据顶点排序字段对分片顶点数据进行排序处理,以生成分片排序顶点数据;确定分片边数据对应的边排序字段;根据边排序字段对分片边数据进行排序处理,以生成分片排序边数据。

[0015] 在本公开的一种示例性实施例中,数据导入模块包括目标数据生成单元,上述目标数据生成单元被配置为:对压缩二进制图数据和历史压缩二进制图数据进行逻辑合并处理,以生成逻辑合并数据;对逻辑合并数据进行物理合并去重处理,以生成目标图数据。

[0016] 在本公开的一种示例性实施例中,数据导入模块还包括:标识确定单元,用于响应于数据查询请求,确定数据查询请求中的请求数据标识;时间戳获取单元,用于确定与请求数据标识对应的初始请求数据,并获取初始请求数据的时间戳信息;请求数据确定单元,用于根据时间戳信息从初始请求数据中确定目标请求数据,并返回目标请求数据。

[0017] 在本公开的一种示例性实施例中,上述请求数据确定单元包括第一数据确定子单元,上述第一数据确定子单元被配置为:根据请求顶点标识确定初始请求顶点数据,并获取初始请求顶点数据对应的顶点时间戳信息;基于顶点时间戳信息从初始请求顶点数据确定最新顶点数据,将最新顶点数据作为目标请求数据。

[0018] 在本公开的一种示例性实施例中,上述请求数据确定单元还包括第二数据确定子单元,上述第二数据确定子单元被配置为:根据请求边标识确定初始请求边数据,并获取初始请求边数据对应的边时间戳信息;基于边时间戳信息从初始请求边数据确定最新边数据,将最新边数据作为目标请求数据。

[0019] 根据本公开的一个方面,提供一种电子设备,包括:处理器;以及存储器,用于存储所述处理器的可执行指令;其中,所述处理器配置为经由执行所述可执行指令来执行上述任意一项所述的方法。

[0020] 根据本公开的一个方面,提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述任意一项所述的方法。

[0021] 根据本公开的一个方面,提供一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质中读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述各个实施例中提供的图数据库的数据导入方法。

[0022] 本公开示例性实施例可以具有以下部分或全部有益效果:

[0023] 在本公开的一示例实施方式所提供的图数据库的数据导入方法中,对获取到的源图数据进行字段填充处理得到待导入图数据,对待导入图数据进行分片处理和数据写入处理后,生成对应的压缩二进制图数据,并将压缩二进制图数据与对应的历史压缩二进制图数据进行合并,得到目标图数据,将目标图数据导入至图数据库中。一方面,将对源图数据进行填充处理的过程从图数据库中分离出来,可以大量减少图数据库处理器的耗时,避免图数据库对文档服务产生影响。另一方面,通过对待导入图数据进行分片处理,可以使经过分片处理后的图数据符合图数据库的数据存储格式,提高数据导入速度。又一方面,生成待导入图数据对应的压缩二进制图数据,并导入至图数据库中,由于压缩二进制图数据为高压数据且对带宽挤占较小,因此,可以大幅提升数据导入速度。

[0024] 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

[0025] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。显而易见地,下面描述中的附图仅仅是本公开的一些实施例,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。

[0026] 图1示出了可以应用本公开实施例的一种图数据库的数据导入方法及装置的示例性系统架构的示意图。

[0027] 图2示出了适于用来实现本公开实施例的电子设备的计算机系统的结构示意图。

[0028] 图3示意性示出了根据本公开的一个实施例的图数据库的数据导入方法的流程图。

[0029] 图4示意性示出了根据本公开的一个实施例的图数据库的数据导入方法的整体架构图。

[0030] 图5示意性示出了根据本公开的一个实施例的根据源图数据生成待导入图数据的流程图。

[0031] 图6示意性示出了根据本公开的一个实施例的对源图数据进行字段校验处理的流程图。

[0032] 图7示意性示出了根据本公开的一个实施例的对源图数据进行字段填充处理的流

程图。

[0033] 图8示意性示出了根据本公开的一个实施例的确定待导入图数据的分片信息的流程图。

[0034] 图9示意性示出了根据本公开的一个实施例的根据待导入图数据生成压缩二进制图数据的流程图。

[0035] 图10示意性示出了根据本公开的一个实施例的对分片数据进行排序处理以生成分片排序数据的流程图。

[0036] 图11示意性示出了根据本公开的一个实施例的对压缩二进制图数据和历史压缩二进制图数据进行合并处理的流程图。

[0037] 图12示意性示出了根据本公开的一个实施例的响应于数据查询请求以确定目标请求数据的流程图。

[0038] 图13示意性示出了根据本公开的一个实施例的图数据库的数据导入装置的结构图。

具体实施方式

[0039] 现在将参考附图更全面地描述示例实施方式。然而,示例实施方式能够以多种形式实施,且不应被理解为限于在此阐述的范例;相反,提供这些实施方式使得本公开将更加全面和完整,并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中,提供许多具体细节从而给出对本公开的实施方式的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而省略所述特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知技术方案以避免喧宾夺主而使得本公开的各方面变得模糊。

[0040] 此外,附图仅为本公开的示意性图解,并非一定是按比例绘制。图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体,不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0041] 区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层。

[0042] 区块链底层平台可以包括用户管理、基础服务、智能合约以及运营监控等处理模块。其中,用户管理模块负责所有区块链参与者的身份信息管理,包括维护公私钥生成(账户管理)、密钥管理以及用户真实身份和区块链地址对应关系维护(权限管理)等,并且在授权的情况下,监管和审计某些真实身份的交易情况,提供风险控制的规则配置(风控审计);基础服务模块部署在所有区块链节点设备上,用来验证业务请求的有效性,并对有效请求完成共识后记录到存储上,对于一个新的业务请求,基础服务先对接口适配解析和鉴权处

理(接口适配),然后通过共识算法将业务信息加密(共识管理),在加密之后完整一致的传输至共享账本上(网络通信),并进行记录存储;智能合约模块负责合约的注册发行以及合约触发和合约执行,开发人员可以通过某种编程语言定义合约逻辑,发布到区块链上(合约注册),根据合约条款的逻辑,调用密钥或者其它的事件触发执行,完成合约逻辑,同时还提供对合约升级注销的功能;运营监控模块主要负责产品发布过程中的部署、配置的修改、合约设置、云适配以及产品运行中的实时状态的可视化输出,例如:告警、监控网络情况、监控节点设备健康状态等。

[0043] 平台产品服务层提供典型应用的基本能力和实现框架,开发人员可以基于这些基本能力,叠加业务的特性,完成业务逻辑的区块链实现。应用服务层提供基于区块链方案的应用服务给业务参与方进行使用。

[0044] 示例性的,图1示出了可以应用本公开实施例的一种图数据库的数据导入方法及装置的示例性应用环境的系统架构的示意图。

[0045] 如图1所示,系统架构100可以包括终端设备101、102、103中的一个或多个,网络104,服务器105和图数据库106。其中,图数据库106可以设置于区块链中的区块链节点设备中。示例性的,通过本公开提供的图数据库的数据导入方法得到的目标图数据可以存储至区块链中。

[0046] 具体的,网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。终端设备101、102、103可以是具有显示屏的各种电子设备,包括但不限于台式计算机、便携式计算机、智能手机和平板电脑等等。应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器的数目。比如服务器105可以是多个服务器组成的服务器集群等。

[0047] 本公开实施例所提供的图数据库的数据导入方法一般由服务器105执行,相应地,图数据库的数据导入装置一般设置于服务器105中。举例而言,在一种示例性实施例中,可以是服务器105获取源图数据,并通过本公开实施例所提供的图数据库的数据导入方法将源图数据进行字段填充处理生成对应的待导入图数据,对待导入图数据进行分片处理、数据写入处理以及合并处理之后,生成目标图数据,并将目标图数据导入至图数据库106中,终端设备101、102、103可以通过网络104向图数据库106发送数据查询请求,以便图数据库106将目标请求数据返回至终端设备101、102、103。

[0048] 图2示出了适于用来实现本公开实施例的电子设备的计算机系统的结构示意图。

[0049] 需要说明的是,图2示出的电子设备的计算机系统200仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0050] 如图2所示,计算机系统200包括中央处理单元(CPU) 201,其可以根据存储在只读存储器(ROM) 202中的程序或者从存储部分208加载到随机访问存储器(RAM) 203中的程序而执行各种适当的动作和处理。在RAM 203中,还存储有系统操作所需的各种程序和数据。CPU 201、ROM 202以及RAM 203通过总线204彼此相连。输入/输出(I/O)接口205也连接至总线204。

[0051] 以下部件连接至I/O接口205:包括键盘、鼠标等的输入部分206;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分207;包括硬盘等的存储部分208;

以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分209。通信部分209经由诸如因特网的网络执行通信处理。驱动器210也根据需要连接至I/O接口205。可拆卸介质211,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器210上,以便于从其上读出的计算机程序根据需要被安装入存储部分208。

[0052] 特别地,根据本公开的实施例,下文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分209从网络上被下载和安装,和/或从可拆卸介质211被安装。在该计算机程序被中央处理单元(CPU)201执行时,执行本申请的方法和装置中限定的各种功能。在一些实施例中,计算机系统200还可以包括AI(Artificial Intelligence,人工智能)处理器,该AI处理器用于处理有关机器学习的计算操作。

[0053] 需要说明的是,本公开所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0054] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0055] 描述于本公开实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现,所描述的单元也可以设置在处理器中。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定。

[0056] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是

上述实施例中所述的电子设备中所包含的;也可以是单独存在,而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该电子设备执行时,使得该电子设备实现如下述实施例中所述的方法。例如,所述的电子设备可以实现如图3~图12所示的各个步骤等。

[0057] 目前,在进行图数据库的数据导入时,通常可以采用下述两种方式进行:一是批量写入;一是主动加载。

[0058] 在批量写入的技术方案中,可以由旁路模块解析数据,当通过数据解析处理生成解析数据后,获取解析后的数据,解析数据为文本格式数据。并通过批量batch请求获取到一定量的数据后,通过发请求的方式直接写入至图数据库中。通过多次batch请求将数据发送到图数据库中进行批写入。在主动加载的技术方案中,可以由图数据库主动加载文件,解析并写入本地数据。图数据库加载的文件同样为采用文本格式文件表示的图数据。

[0059] 然而,采用上述两种方案进行图数据库的数据导入时,由于导入图数据库中的数据格式为文本格式,在采用文本格式文件表示图数据时,文本格式数据将产生较为明显的膨胀,因此,采用现有技术方案时,数据导入速度较慢,如果面对百亿量级以上数据,将产生巨大的时间耗时。

[0060] 另外,采用现有技术方案,通常由图数据库处理相关的导入数据,导致图数据库所对应的中央处理器(central processing unit,CPU)耗时较高,容易影响图数据库CPU提供文档服务。

[0061] 基于上述一个或多个问题,本示例实施方式提供了一种图数据库的数据导入方法。该图数据库的数据导入方法可以应用于上述服务器105,也可以应用于上述终端设备101、102、103中的一个或多个,本示例性实施例中对此不做特殊限定。以下对本公开实施例的技术方案进行详细阐述:

[0062] 参考图3所示,该图数据库的数据导入方法可以包括以下步骤S310至步骤S350:

[0063] 步骤S310.获取源图数据,对源图数据进行字段填充处理以得到待导入图数据。

[0064] 步骤S320.确定待导入图数据的分片信息,根据分片信息对待导入图数据进行分片处理以得到对应的分片图数据。

[0065] 步骤S330.对分片图数据进行数据写入处理,以生成对应的压缩二进制图数据。

[0066] 步骤S340.确定与压缩二进制图数据对应的历史压缩二进制图数据。

[0067] 步骤S350.对压缩二进制图数据和历史压缩二进制图数据进行合并处理,得到目标图数据,以将目标图数据导入至图数据库。

[0068] 在本示例实施方式所提供的图数据库的数据导入方法中,对获取到的源图数据进行字段填充处理得到待导入图数据,对待导入图数据进行分片处理和数据写入处理后,生成对应的压缩二进制图数据,并将压缩二进制图数据与对应的历史压缩二进制图数据进行合并,得到目标图数据,将目标图数据导入至图数据库中。一方面,将对源图数据进行填充处理的过程从图数据库中分离出来,可以大量减少图数据库处理器的耗时,避免图数据库对文档服务产生影响。另一方面,通过对待导入图数据进行分片处理,可以使经过分片处理后的图数据符合图数据库的数据存储格式,提高数据导入速度。又一方面,生成待导入图数据对应的压缩二进制图数据,并导入至图数据库中,由于压缩二进制图数据为高压数据且对带宽挤占较小,因此,可以大幅提升数据导入速度。

[0069] 下面,对于本示例实施方式的上述步骤进行更加详细的说明。

[0070] 在步骤S310中,获取源图数据,对源图数据进行字段填充处理以得到待导入图数据。

[0071] 本示例实施方式中,图可以是由顶点集合以及顶点间的关系集合组成的一种数据结构,图数据通常包括顶点数据和边数据。顶点有时也称为节点或者交点,边有时也称为连接。每个节点可以代表一个实体,如人、地点、事物、类别或其他数据,每个边可以代表两个节点的关联方式。采用图结构可以对各种场景进行建模,例如,设备网络、道路系统、人的社交关系或由关系定义的任何其他事物等等。图数据库可以是以图这种数据结构存储和查询数据的数据库,图数据库具有处理图形数据模型的创建、读取、更新和删除操作。源图数据可以从初始源数据中确定出的可以表示图结构的数据。字段填充处理可以对源图数据进行字段填充的处理过程。待导入图数据可以对源图数据进行字段填充处理后生成的图数据,待导入图数据中可以包括图数据库中所存储数据的相关字段。

[0072] 由于源图数据中所包含的字段是相关用户在处理数据时所使用的字段,源图数据中未包括图数据库中存储数据的相关数据字段,因此,在在获取到源图数据后,可以对源图数据进行字段填充处理,以得到待导入图数据,以对待导入图数据进行相关处理并导入至图数据库中。

[0073] 在本公开的一种示例性实施例中,确定初始源数据,从初始源数据中确定顶点数据;从初始源数据中确定边数据;将顶点数据和边数据作为源图数据。

[0074] 其中,初始源数据可以从源数据平台中直接获取到的数据。初始源数据通常存放在大数据平台中,例如,初始源数据可以存储在Hive数据仓库工具、Hadoop分布式文件系统(Hadoop Distributed File System,HDFS)中等。图数据一般包括顶点数据和边数据,顶点数据可以包括顶点标识和顶点属性;其中,顶点标识(vertex id)可以图数据中节点的唯一标识符。顶点属性可以是图数据中的顶点所具有的属性。边数据可以包括起始顶点标识、终止顶点标识和边属性。在有向边构成的有向图中,起始顶点标识(source vertex id)可以是某条边作为起点的顶点标识;终止顶点标识(destination vertex id)可以是该条边指向的终点的顶点标识。边属性可以是图数据中边所具有的属性。在无向图中,则不区分起始顶点和终止顶点。

[0075] 参考图4,图4示意性示出了根据本公开的一个实施例的图数据库的数据导入方法的整体架构图。初始源数据可以存储在源数据平台410中,源数据平台410可以是任何提供数据的组件,例如Hive、HDFS、Hbase以及云服务提供的存储服务中。在源数据平台中获取到初始源数据后,可以从初始源数据中确定出源图数据。

[0076] 具体的,可以从初始源数据中确定出顶点数据,顶点数据可以包括顶点标识和顶点属性;顶点属性可以由用户进行自定义,根据具体场景进行配置。例如,在购物场景中,顶点可以表示物品、用户和货架等等,顶点标识可以是物品唯一标识、用户唯一标识和货架唯一标识。顶点属性可以表示物品的价格、物品分类等;还可以表示货架的尺寸、货架的材质等;也可以表示用户的年龄段、购买喜好等。在社交推荐场景中,顶点可以表示各个不同用户,顶点标识可以表示不同的用户标识。顶点属性边数据可以包括起始顶点标识、终止顶点标识和边属性;边属性可以表示起始顶点和终止顶点之间的关联关系。例如,在购物场景中,边属性可以表示用户购买某一物品的购买关系,还可以表示货架存放某一物品的存储

关系等；在社交推荐场景中，边属性可以表示用户1与用户2之间的好友关系等。在获取到顶点数据和边数据后，可以将顶点数据和边数据作为源图数据。

[0077] 在本公开的一种示例性实施例中，从图数据库中获取源图数据对应的元数据；根据元数据对源图数据进行字段校验处理，将通过字段校验处理的源图数据作为初始图数据；对初始图数据进行字段填充处理，以生成待导入图数据。

[0078] 其中，元数据可以描述源图数据的数据属性的信息。字段校验处理可以是确定源图数据的字段是否与元数据中定义字段的数据属性一致的校验过程。初始图数据可以通过字段校验处理的图数据，即初始图数据可以是数据字段与元数据中所定义字段的数据类型和数据数量一致的图数据。

[0079] 参考图5，图5示意性示出了根据本公开的一个实施例的根据源图数据生成待导入图数据的流程图。在步骤S510中，在基于初始源数据确定出源图数据之后，可以从图数据库中获取与源图数据对应的元数据。在步骤S520中，由于元数据中定义了源图数据的数据属性的相关信息，可以根据元数据对源图数据进行字段校验处理，将通过字段校验处理的源图数据作为初始图数据。在步骤S530中，对初始图数据进行字段填充处理，为初始图数据添加图数据库所需的字段，可以生成对应的待导入图数据。

[0080] 举例而言，可以通过图4中的计算引擎420对初始图数据进行处理，以生成待导入图数据。计算引擎420可以是任何一个分布式计算组件，例如，计算引擎420可以是Apache Flink数据处理平台，还可以是多点接口(MPI)并行计算框架，也可以是Hadoop映射规约(Hadoop MapReduce)计算框架等等。

[0081] 在本公开的一种示例性实施例中，根据元数据确定参考字段类型和参考字段数量；确定源图数据的字段类型和源图数据的字段数量；确定源图数据的字段类型与参考字段类型是否一致，并确定源图数据的字段数量与参考字段数量是否一致。

[0082] 其中，参考字段类型可以是元数据中定义的字段类型。参考字段数量可以是元数据中定义的字段数量。源图数据的字段类型是从初始源数据中确定出的源图数据对应的字段类型。源图数据的字段数量是从初始源数据中确定出的源图数据对应的字段数量。

[0083] 参考图6，图6示意性示出了根据本公开的一个实施例的对源图数据进行字段校验处理的流程图。在步骤S610中，从元数据中确定出参考字段类型和参考字段数量。在步骤S620中，可以分别确定出源图数据的字段类型和源图数据的字段数量。在步骤S630中，分别将源图数据的字段类型和参考字段类型，源图数据的字段数量与参考字段数量进行对比，当源图数据的字段类型和参考字段类型完全一致，且源图数据的字段数量与参考字段数量完全一致时，则认为源图数据通过字段校验处理过程。可以将通过字段校验处理的源图数据作为初始图数据。

[0084] 例如，当参考字段类型中定义某一参考字段的数据类型为布尔类型时，如果源图数据中该字段类型也为布尔类型，则认为源图数据的字段类型与参考字段类型一致。当参考字段类型中定义另一参考字段的数据类型为整型类型时，如果源图数据中该字段类型为字符类型时，则认为源图数据的字段类型与参考字段类型不一致。当参考字段数量为2个时，如果源图数据的字段数量也为2个，则认为源图数据的字段数量与参考字段数量一致。当参考字段数量为3个时，如果源图数据的字段数量为2个，则认为源图数据的字段数量与

参考字段数量不一致。

[0085] 在本公开的一种示例性实施例中,确定与初始顶点数据对应的补充顶点字段;确定与初始边数据对应的补充边字段;根据补充顶点字段和补充边字段对初始图数据进行字段填充处理,以生成待导入图数据。

[0086] 其中,初始图数据可以包括初始顶点数据和初始边数据。补充顶点字段可以是对初始顶点数据进行字段填充的字段;补充顶点字段可以包括顶点类型标识、顶点时间戳和顶点属性规则标识。顶点类型标识(vertex type id)可以用于标识顶点对应的具体类型。顶点时间戳(vertex data timestamp)可以是顶点生成时所对应的时间戳。顶点属性规则标识(vertex schema version)可以是顶点属性所对应的不同版本规则的标识。补充边字段可以是对初始边数据进行字段填充的字段;补充边字段可以包括边类型标识、边时间戳和边属性规则标识。边类型标识(edge type id)可以用于标识边对应的具体类型。边时间戳(edge data timestamp)可以边生成时所对应的时间戳。边属性规则标识(edge schema version)可以是边属性所对应的不同版本规则的标识。

[0087] 在确定出初始图数据后,可以对初始图数据进行字段填充处理,补齐初始图数据所需的完整信息,使得经过字段填充处理后的图数据具备图数据库中数据存储的数据格式,即生成待导入图数据。参考图7,图7示意性示出了根据本公开的一个实施例的对源图数据进行字段填充处理的流程图。在步骤S710中,确定与初始顶点数据对应的补充顶点字段。补充顶点数据可以包括顶点类型标识、顶点时间戳和顶点属性规则标识。不同的顶点类型可以对应一个顶点类型标识,举例而言,在购物场景中,物品、用户和货架可以分别对应的不同的顶点类型,可以将三者对应的顶点类型标识分别标记为vt0001、vt0002、vt0003等。某一用户类型的顶点对应的顶点时间戳可以是“2020-10-10 20:00:00”。由于顶点属性规则记录了顶点数据的属性列表具体包含哪些字段,每个字段的字段类型。当顶点数据的字段数量发生变化或字段类型发生变化时,均可以形成一个新的顶点属性规则,因此,可以采用顶点属性规则标识区分不同版本的顶点属性规则。例如,顶点属性规则标识可以是vertex schema version1、vertex schema version2、vertex schema version3等等。

[0088] 在步骤S720中,确定与初始边数据对应的补充边字段。补充边字段可以包括边类型标识、边时间戳和边属性规则标识。两个顶点数据之间可以对应多种关联关系,例如,在购物场景下,当用户购买某一物品时,用户顶点与物品顶点之间的关联关系可以是购买,可以将购买关系标识为et0001;当货架上存放某一物品时,货架顶点与物品顶点之间的关联关系可以为储物,可以将储物关系标识为et002,同理,可以标记该场景下的其他顶点之间的关联关系。与顶点时间戳类似,用户购买某一物品的购买关系对应的边时间戳可以为“2020-10-10 21:00:00”。与顶点属性规则类似,边属性规则同样记录了边数据的属性列表,由于边属性规则可能发生变化,因此,可以采用边属性规则标识区分不同版本的边属性规则。例如,边属性规则标识可以是edge schema version1、edge schema version2等等。

[0089] 在步骤S730中,可分别根据补充顶点字段和补充边字段对初始图数据进行字段填充处理,以生成待导入图数据。

[0090] 在步骤S320中,确定待导入图数据的分片信息,根据分片信息对待导入图数据进行分片处理以得到对应的分片图数据。

[0091] 本示例实施方式中,分片信息可以是待导入图数据所处的数据分片的信息,例如,

分片信息可以包括数据分片的编号等。分片图数据可以是经过分片处理后具有不同数据分片的分片信息的图数据。

[0092] 由于在大数据背景下,图数据库可以是分布式数据库,图数据库可以采用分片存储的方式进行数据存储。在进行图数据库的数据导入之前,可以先根据图数据库的分片规则对待导入图数据进行分片处理,将待导入图数据划分到不同的存储目录下,因此,在对待导入图数据进行分片处理时,分片处理的分片规则与图数据库进行分片存储的规则一致。

[0093] 参考图4,在生成待导入图数据后,可以通过计算引擎420确定出待导入图数据对应的分片信息,以根据分片信息对待导入图数据进行分片处理,生成分片图数据。

[0094] 在本公开的一种示例性实施例中,确定待导入图数据对应的目标顶点标识和目标起始顶点标识;获取图数据库对应的分片数;根据分片数和目标顶点标识确定待导入顶点数据所对应分片的标识,作为顶点分片标识;根据分片数和目标起始顶点标识确定待导入边数据所对应分片的标识,作为边分片标识。

[0095] 其中,待导入图数据中包括待导入顶点数据和待导入边数据。目标顶点标识可以是待导入图数据中包含的待导入顶点数据对应的标识。目标起始顶点标识可以是待导入图数据中包含的待导入边数据中起始顶点对应的标识。分片数可以是数据分片的数量,分片数与图数据库对应的数据分片的数量相同。顶点分片标识可以是根据分片数和目标顶点标识确定出的待导入顶点数据所对应数据分片的标识,顶点分片标识的取值范围为 $[0, \text{分片数})$ 。边分片标识可以是根据分片数和目标起始顶点标识确定出的待导入边数据所对应数据分片的标识,边分片标识的取值范围为 $[0, \text{分片数})$ 。

[0096] 参考图8,图8示意性示出了根据本公开的一个实施例的确定待导入图数据的分片信息的流程图。在步骤S810中,确定与待导入图数据对应的目标顶点标识和目标起始顶点标识。在步骤S820中,获取图数据库对应的分片数。例如,参考图4,图4中所示的分片数为4。在步骤S830中,通过对分片数和目标顶点标识进行计算,可以确定出待导入顶点数据所对应的顶点分片标识。具体的,可以采用公式1所示的计算方式计算顶点分片标识;其中,“hash”可以为哈希运算,“%”可以是取余运算。

[0097] 顶点分片标识 $=\text{hash}(\text{vertex id})\% \text{分片数}$ (公式1)

[0098] 在步骤S840中,通过对分片数和目标起始顶点标识进行计算,可以确定出待导入边数据所对应的边分片标识,可以采用公式2所示的计算方式计算边分片标识。

[0099] 边分片标识 $=\text{hash}(\text{source vertex id})\% \text{分片数}$ (公式2)

[0100] 通过上述计算方式,可以确定出待导入图数据中所有待导入顶点数据和待导入边数据所对应的数据分片。

[0101] 在步骤S330中,对分片图数据进行数据写入处理,以生成对应的压缩二进制图数据。

[0102] 本示例实施方式中,数据写入处理可以是利用数据库引擎进行数据写入,以生成压缩二进制图数据的处理过程。压缩二进制图数据可以是采用二进制表示的高压缩图数据,例如,压缩二进制图数据可以是排序字符串表(Sorted Strings Table, SST)文件, SST文件可以特指RocksDB的磁盘文件,RocksDB是基于C++语言编写的嵌入式键值对(Key/Value, KV)存储引擎,它是一个高效、高性能、单点的数据库引擎。压缩二进制图数据还可以是任何一个高性能KV存储引擎的磁盘文件,比如InnoDB存储引擎、LevelDB存储引擎以及

HBase存储引擎等。

[0103] 在确定出分片图数据后,可以通过高性能KV存储引擎对分片图数据进行数据写入处理,并生成对应的压缩二进制图数据。由于现有的图数据库数据导入方法中,通常采用文本格式表示待导入图数据,采用文本格式表示图数据,将产生图数据较大的数据膨胀的现象,因此,本公开在进行数据导入之前,可以将待导入图数据转换为压缩二进制图数据,由于压缩二进制图数据是一种高压缩数据,且对宽带的基站较小,从而可以大幅加快图数据库的数据导入速度。

[0104] 在本公开的一种示例性实施例中,确定分片图数据中的分片顶点数据和分片边数据;分别对分片顶点数据和分片边数据进行排序处理,以得到对应的分片排序顶点数据和分片排序边数据;对分片排序顶点数据和分片排序边数据进行顺序写入处理,以生成压缩二进制图数据。

[0105] 其中,分片顶点数据可以是经过分片处理后的顶点数据。分片边数据可以是经过分片处理后的边数据。分片排序顶点数据可以是经过排序处理的分片顶点数据。分片排序边数据可以是经过排序处理的分片边数据。

[0106] 在根据分片图数据生成压缩二进制图数据之前,可以先对各个数据分片中的分片图数据进行排序处理。参考图9,图9示意性示出了根据本公开的一个实施例的根据待导入图数据生成压缩二进制图数据的流程图。在步骤S910中,确定分片图数据中的分片顶点数据和分片边数据。在步骤S920中,在对分片顶点数据进行排序处理时,可以按照分片顶点数据的相关字段进行排序处理,生成分片排序顶点数据。同样的,在对分片边数据进行排序处理时,可以按照分片边数据的相关字段进行排序处理,生成分片排序边数据。在步骤S930中,对于生成的分片排序顶点数据和分片排序边数据可以采用高性能KV存储引擎进行顺序写入处理,以生成压缩二进制图数据。

[0107] 举例而言,在采用RocksDB存储引擎进行数据写入处理时,可以利用RocksDB库(RocksDB lib)进行顺序写入处理,RocksDB lib可以相当于一个格式转换器,将分片图数据生成对应的SST文件。对分片图数据进行排序处理,可以使得生成的分片排序数据符合图数据库的数据格式需求,同时提高后续的数据导入速度。在根据分片图数据生成对应的SST文件后,SST文件将按照“分片ID/子分片id.sst”的存储路径规则写入至Hadoop分布式文件系统430中。

[0108] 在本公开的一种示例性实施例中,确定分片顶点数据对应的顶点排序字段;根据顶点排序字段对分片顶点数据进行排序处理,以生成分片排序顶点数据;确定分片边数据对应的边排序字段;根据边排序字段对分片边数据进行排序处理,以生成分片排序边数据。

[0109] 其中,顶点排序字段可以是对分片顶点数据进行排序处理时所使用的排序字段,例如,顶点排序字段可以包括顶点标识、顶点类型标识、顶点时间戳和顶点属性规则标识等。边排序字段可以是对分片边数据进行排序处理所使用的排序字段,例如,边排序字段可以包括起始顶点标识、终止顶点标识、边类型标识、边时间戳和边属性规则标识等。

[0110] 参考图10,图10示意性示出了根据本公开的一个实施例的对分片数据进行排序处理以生成分片排序数据的流程图。在对分片顶点数据进行排序处理之前,在步骤S1010中,可以确定与分片顶点数据对应的顶点排序字段,顶点排序字段可以包括排序优先级,例如,顶点排序字段的排序优先级由高到低为:顶点标识、顶点类型标识、顶点时间戳、顶点属性

规则标识。在步骤S1020中,按照顶点排序字段将分片顶点数据按照数字大小(如0-9)或字符排序先后(如“a-z”)进行排序处理。由于图数据库中同一顶点在不同应用场景下可能产生不同的相关数据。在进行排序处理时,如果两个顶点数据的顶点标识相同,则继续对比这两个顶点数据的顶点类型标识,以此类推,直至完成对所有分片顶点数据的排序处理,生成对应的分片排序顶点数据。

[0111] 类似地,在步骤S1030中,在对分片边数据进行排序处理之前,可以确定与分片边数据对应的边排序字段,边排序字段同样包括排序优先级,例如,边排序字段的排序优先级由高到低为:起始顶点标识、终止顶点标识、边类型标识、边时间戳、边属性规则标识。在步骤S1040中,按照边排序字段将分片边数据按照数字大小或字符排序先后进行排序处理,并生成对应的分片排序边数据。

[0112] 在步骤S340中,确定与压缩二进制图数据对应的历史压缩二进制图数据。

[0113] 本示例实施方式中,历史压缩二进制图数据可以是与压缩二进制图数据处于同一数据分片中,且正在图数据库中提供服务的压缩二进制图数据。

[0114] 对于生成的压缩二进制图数据,由于生成的压缩二进制图数据具有对应的路由信息,可以根据路由信息查找与压缩二进制图数据处于同一数据分片的分片路径下的所有压缩二进制图数据,即历史压缩二进制图数据,并对二者进行处理。

[0115] 在步骤S350中,对压缩二进制图数据和历史压缩二进制图数据进行合并处理,得到目标图数据,以将目标图数据导入至图数据库。

[0116] 本示例实施方式中,合并处理可以是对压缩二进制图数据和历史压缩二进制图数据进行合并处理的过程,可以包括逻辑合并处理和物理合并去重处理等。目标图数据可以是经过合并处理后的压缩二进制图数据,是可以直接导入至图数据库中的图数据。

[0117] 在确定出历史压缩二进制图数据后,可以对压缩二进制图数据和历史压缩二进制图数据进行合并处理。具体的,可以包括两个阶段的合并处理,一个是逻辑合并处理,另一个是物理合并去重处理。将经过合并处理生成目标图数据导入至图数据库440中。

[0118] 在本公开的一种示例性实施例中,对压缩二进制图数据和历史压缩二进制图数据进行逻辑合并处理,以生成逻辑合并数据;对逻辑合并数据进行物理合并去重处理,以生成目标图数据。

[0119] 其中,逻辑合并处理可以是同时保留压缩二进制图数据和历史压缩二进制图数据的合并处理操作。逻辑合并数据可以是包括压缩二进制图数据和历史压缩二进制图数据的数据。合并去重处理可以是对逻辑合并数据中的重复数据进行物理删除的合并处理操作,经过合并去重处理后的图数据中仅保留时间戳信息最近的数据。

[0120] 参考图11,图11示意性示出了根据本公开的一个实施例的对压缩二进制图数据和历史压缩二进制图数据进行合并处理的流程图。在确定出压缩二进制图数据和历史压缩二进制图数据后,在步骤S1110中,可以对压缩二进制图数据和历史压缩二进制图数据进行逻辑合并处理,生成逻辑合并数据。举例而言,可以采用RocksDB的ingest函数进行逻辑合并处理。在逻辑合并处理过程结束后,可以通过RocksDB的接口访问图数据库中原有的历史压缩二进制图数据,以及新生成的压缩二进制图数据。

[0121] 在步骤S1120中,对于生成的逻辑合并数据,可以对其进行物理合并去重处理,以生成目标图数据。举例而言,可以启动一个RocksDB的后台线程进行合并去重处理,RocksDB

的后台线程可以对同一个顶点标识的顶点数据进行合并去重处理,仅保留具有该顶点标识的最新顶点数据。由于一个边可以对应起始顶点和终止顶点,起始顶点数据和终止顶点数据同样属于顶点数据,因此可以根据对顶点数据进行处理过程进行处理,当删除一个顶点时。与该顶点对应的边数据也将被删除。在经过物理合并去重处理生成目标图数据后,可以将目标图数据导入至图数据库中。

[0122] 在某些场景下,采用上述图数据库的数据导入方法,在处理100亿规模的数据时,可以比Hugegraph图数据库的数据导入速度快10倍,比Neo4j图形数据库的数据导入速度快3倍。

[0123] 在本公开的一种示例性实施例中,响应于数据查询请求,确定数据查询请求中的请求数据标识;确定与请求数据标识对应的初始请求数据,并获取初始请求数据的时间戳信息;根据时间戳信息从初始请求数据中确定目标请求数据,并返回目标请求数据。

[0124] 其中,数据查询请求可以是请求端从图数据库中获取目标请求数据的请求。请求数据标识可以是数据查询请求中包括的数据标识。初始请求数据可以是直接根据请求数据标识确定出的请求数据,根据同一个请求数据标识可以确定多个相关的初始请求数据。时间戳信息可以是初始请求数据对应的时间戳的相关信息。目标请求数据可以是根据时间戳信息从初始请求数据中确定出的请求数据。

[0125] 参考图12,图12示意性示出了根据本公开的一个实施例的响应于数据查询请求以确定目标请求数据的流程图。在步骤S1210中,当接收到数据查询请求时,可以响应于数据查询请求,从数据查询请求中确定出请求数据标识,例如,数据查询请求数据中可以包括顶点标识和边标识等。在步骤S1220中,根据请求数据标识可以从图数据库中确定出与请求数据标识对应的初始请求数据,并获取初始请求数据的时间戳信息。由于在未进行物理合并去重处理之前,对于同一个数据标识的请求数据,图数据库中可能包含多份数据。当确定出的初始请求数据包含多份数据时,可以获取初始请求数据中各份数据的时间戳信息。在步骤S1230中,根据时间戳信息从初始请求数据中确定出时间信息为最新的数据,作为目标请求数据,并向请求端返回目标请求数据。

[0126] 由于物理合并去重处理可能耗费较多的时间,因此,在物理合并去重处理完成之前,如果接收到数据查询请求,可以根据数据查询请求中的请求数据标识从图数据库中确定初始请求数据,并对初始请求数据进行逻辑去重处理,即从初始请求数据中获取时间戳信息为最新的数据作为目标请求数据,并向请求端返回目标请求数据。

[0127] 在本公开的一种示例性实施例中,根据请求顶点标识确定初始请求顶点数据,并获取初始请求顶点数据对应的顶点时间戳信息;基于顶点时间戳信息从初始请求顶点数据确定最新顶点数据,将最新顶点数据作为目标请求数据。

[0128] 其中,请求顶点标识可以是数据查询请求中所包含的顶点数据所对应的数据标识。初始请求顶点数据可以是直接根据请求顶点标识确定出的顶点数据。顶点时间戳信息可以是初始请求顶点数据对应的时间戳的相关信息。最新顶点数据可以是时间戳信息为最新生成时间,即最新生成的顶点数据。

[0129] 在接收到数据查询请求时,确定数据查询请求中包含的请求顶点标识。根据请求顶点标识可以从图数据库中确定出与该请求顶点标识对应的所有顶点数据,作为初始请求顶点数据。在确定出初始请求顶点数据后,可以获取初始请求顶点数据对应的顶点时间戳

信息,根据顶点时间戳信息可以从初始请求顶点数据中确定出时间戳为最近的最新顶点数据,并将最新顶点数据作为目标请求数据。

[0130] 在本公开的一种示例性实施例中,根据请求边标识确定初始请求边数据,并获取初始请求边数据对应的边时间戳信息;基于边时间戳信息从初始请求边数据确定最新边数据,将最新边数据作为目标请求数据。

[0131] 其中,请求边标识可以是数据查询请求中所包含的边数据所对应的数据标识。初始请求边数据可以是直接根据请求边标识确定出的边数据。边时间戳信息可以是初始请求边数据对应的时间戳的相关信息。最新边数据可以是时间戳信息为最新生成时间,即最新生成的边数据。

[0132] 在接收到数据查询请求时,确定数据查询请求中包含的请求边标识。根据请求边标识可以从图数据库中确定出与该请求边标识对应的所有边数据,作为初始请求边数据。在确定出初始请求边数据后,可以获取初始请求边数据对应的边时间戳信息,根据边时间戳信息可以从初始请求边数据中确定出时间戳为最近的最新边数据,并将最新边数据作为目标请求数据。

[0133] 需要说明的是,在一些示例性实施例中,在生成压缩二进制图数据后,图数据库440可以从HDFS中下载压缩二进制图数据到本地,并获取本地中已有的历史压缩二进制图数据,以在本地对压缩二进制图数据和历史压缩二进制图数据进行数据合并处理,本公开对此不作任何特殊限定。

[0134] 应当注意,尽管在附图中以特定顺序描述了本公开中方法的各个步骤,但是,这并非要求或者暗示必须按照该特定顺序来执行这些步骤,或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的,可以省略某些步骤,将多个步骤合并为一个步骤执行,以及/或者将一个步骤分解为多个步骤执行等。

[0135] 进一步的,本示例实施方式中,还提供了一种图数据库的数据导入装置。该图数据库的数据导入装置可以应用于一服务器或终端设备。参考图13所示,该图数据库的数据导入装置1300可以包括数据获取模块1310、数据分片模块1320、压缩数据生成模块1330、历史数据确定模块1340以及数据导入模块1350。其中:

[0136] 数据获取模块1310,用于获取源图数据,对源图数据进行字段填充处理以得到待导入图数据;

[0137] 数据分片模块1320,用于确定待导入图数据的分片信息,根据分片信息对待导入图数据进行分片处理以得到对应的分片图数据;

[0138] 压缩数据生成模块1330,用于对分片图数据进行数据写入处理,以生成对应的压缩二进制图数据;

[0139] 历史数据确定模块1340,用于确定与压缩二进制图数据对应的历史压缩二进制图数据;

[0140] 数据导入模块1350,用于对压缩二进制图数据和历史压缩二进制图数据进行合并处理,得到目标图数据,以将目标图数据导入至图数据库。

[0141] 在本公开的一种示例性实施例中,数据获取模块1310包括源数据获取单元,上述源数据获取单元被配置为:确定初始源数据,从初始源数据中确定顶点数据;从初始源数据中确定边数据;将顶点数据和边数据作为源图数据。

[0142] 在本公开的一种示例性实施例中,数据获取模块1310还包括:元数据获取单元,用于从图数据库中获取源图数据对应的元数据;校验处理单元,用于根据元数据对源图数据进行字段校验处理,将通过字段校验处理的源图数据作为初始图数据;填充处理单元,用于对初始图数据进行字段填充处理,以生成待导入图数据。

[0143] 在本公开的一种示例性实施例中,校验处理单元被配置为:根据元数据确定参考字段类型和参考字段数量;确定源图数据的字段类型和源图数据的字段数量;确定源图数据的字段类型与参考字段类型是否一致,并确定源图数据的字段数量与参考字段数量是否一致。

[0144] 在本公开的一种示例性实施例中,填充处理单元被配置为:确定与初始顶点数据对应的补充顶点字段;确定与初始边数据对应的补充边字段;根据补充顶点字段和补充边字段对初始图数据进行字段填充处理,以生成待导入图数据。

[0145] 在本公开的一种示例性实施例中,数据分片模块1320包括分片信息确定单元,上述分片信息确定单元被配置为:确定待导入图数据对应的目标顶点标识和目标起始顶点标识;获取图数据库对应的分片数;根据分片数和目标顶点标识确定待导入顶点数据所对应分片的标识,作为顶点分片标识;根据分片数和目标起始顶点标识确定待导入边数据所对应分片的标识,作为边分片标识。

[0146] 在本公开的一种示例性实施例中,压缩数据生成模块1330包括:分片数据确定单元,用于确定分片图数据中的分片顶点数据和分片边数据;数据排序单元,用于分别对分片顶点数据和分片边数据进行排序处理,以得到对应的分片排序顶点数据和分片排序边数据;压缩数据生成单元,用于对分片排序顶点数据和分片排序边数据进行顺序写入处理,以生成压缩二进制图数据。

[0147] 在本公开的一种示例性实施例中,数据排序单元被配置为:确定分片顶点数据对应的顶点排序字段;根据顶点排序字段对分片顶点数据进行排序处理,以生成分片排序顶点数据;确定分片边数据对应的边排序字段;根据边排序字段对分片边数据进行排序处理,以生成分片排序边数据。

[0148] 在本公开的一种示例性实施例中,数据导入模块1350包括目标数据生成单元,上述目标数据生成单元被配置为:对压缩二进制图数据和历史压缩二进制图数据进行逻辑合并处理,以生成逻辑合并数据;对逻辑合并数据进行物理合并去重处理,以生成目标图数据。

[0149] 在本公开的一种示例性实施例中,数据导入模块1350还包括:标识确定单元,用于响应于数据查询请求,确定数据查询请求中的请求数据标识;时间戳获取单元,用于确定与请求数据标识对应的初始请求数据,并获取初始请求数据的时间戳信息;请求数据确定单元,用于根据时间戳信息从初始请求数据中确定目标请求数据,并返回目标请求数据。

[0150] 在本公开的一种示例性实施例中,上述请求数据确定单元包括第一数据确定子单元,上述第一数据确定子单元被配置为:根据请求顶点标识确定初始请求顶点数据,并获取初始请求顶点数据对应的顶点时间戳信息;基于顶点时间戳信息从初始请求顶点数据确定最新顶点数据,将最新顶点数据作为目标请求数据。

[0151] 在本公开的一种示例性实施例中,上述请求数据确定单元还包括第二数据确定子单元,上述第二数据确定子单元被配置为:根据请求边标识确定初始请求边数据,并获取初

始请求边数据对应的边时间戳信息；基于边时间戳信息从初始请求边数据确定最新边数据，将最新边数据作为目标请求数据。

[0152] 上述图数据库的数据导入装置中各模块或单元的具体细节已经在对应的图数据库的数据导入方法中进行了详细的描述，因此此处不再赘述。

[0153] 应当注意，尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元，但是这种划分并非强制性的。实际上，根据本公开的实施方式，上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之，上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0154] 本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本公开的其他实施方案。本申请旨在涵盖本公开的任何变型、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公开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，本公开的真正范围和精神由下面的权利要求指出。

[0155] 应当理解的是，本公开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。本公开的范围仅由所附的权利要求来限制。

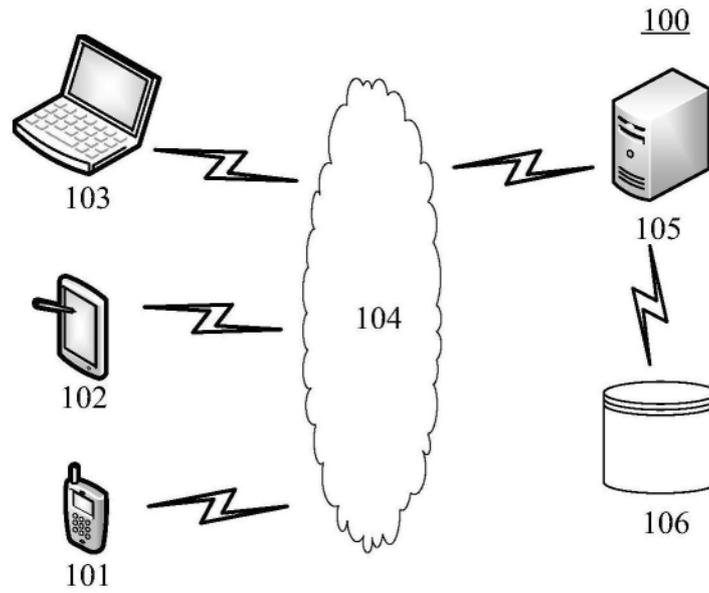


图1

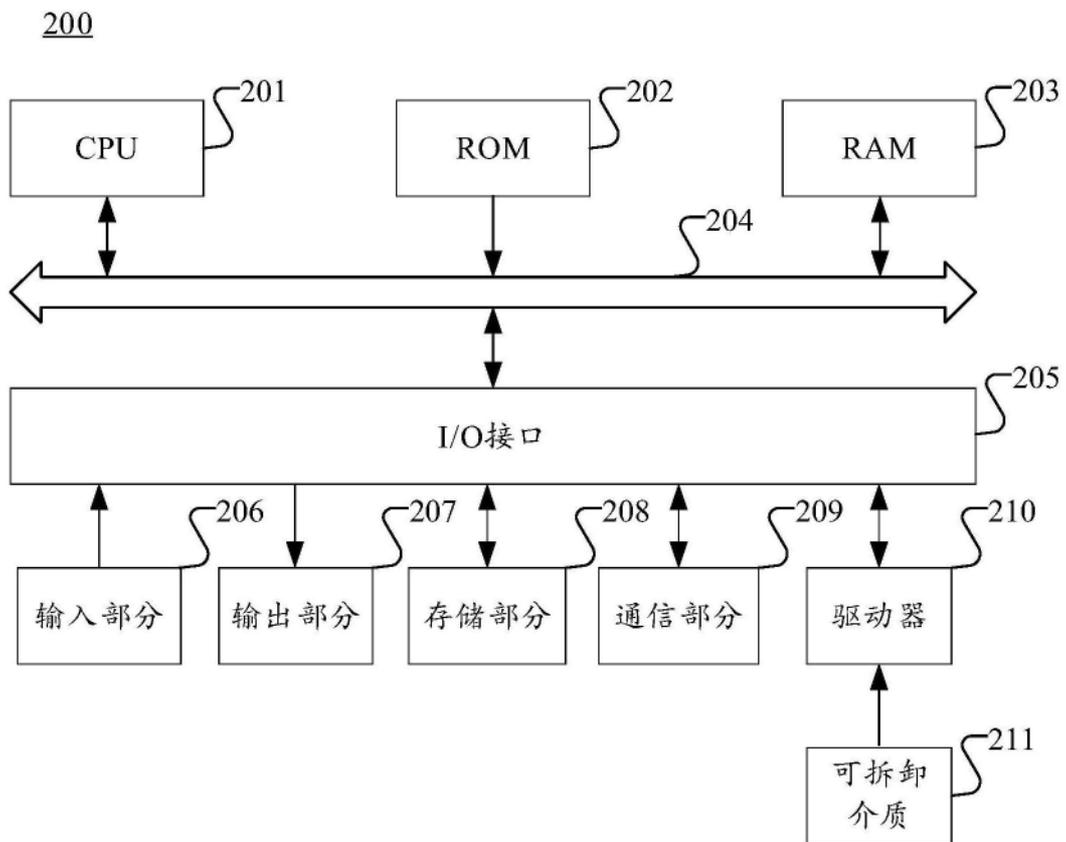


图2

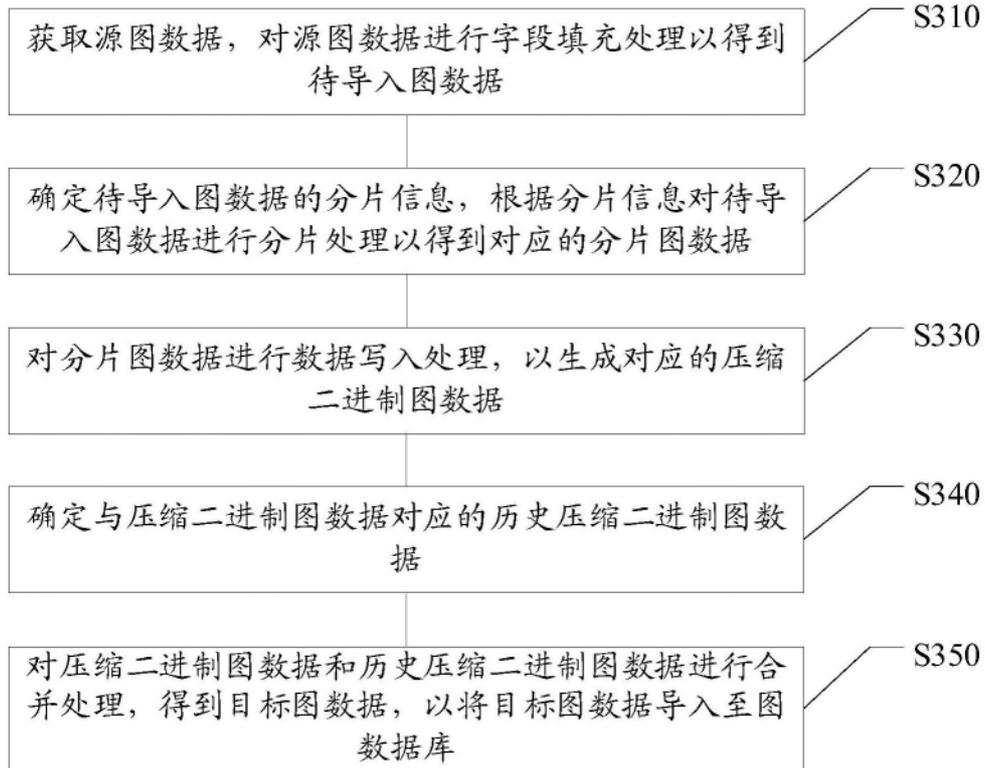


图3

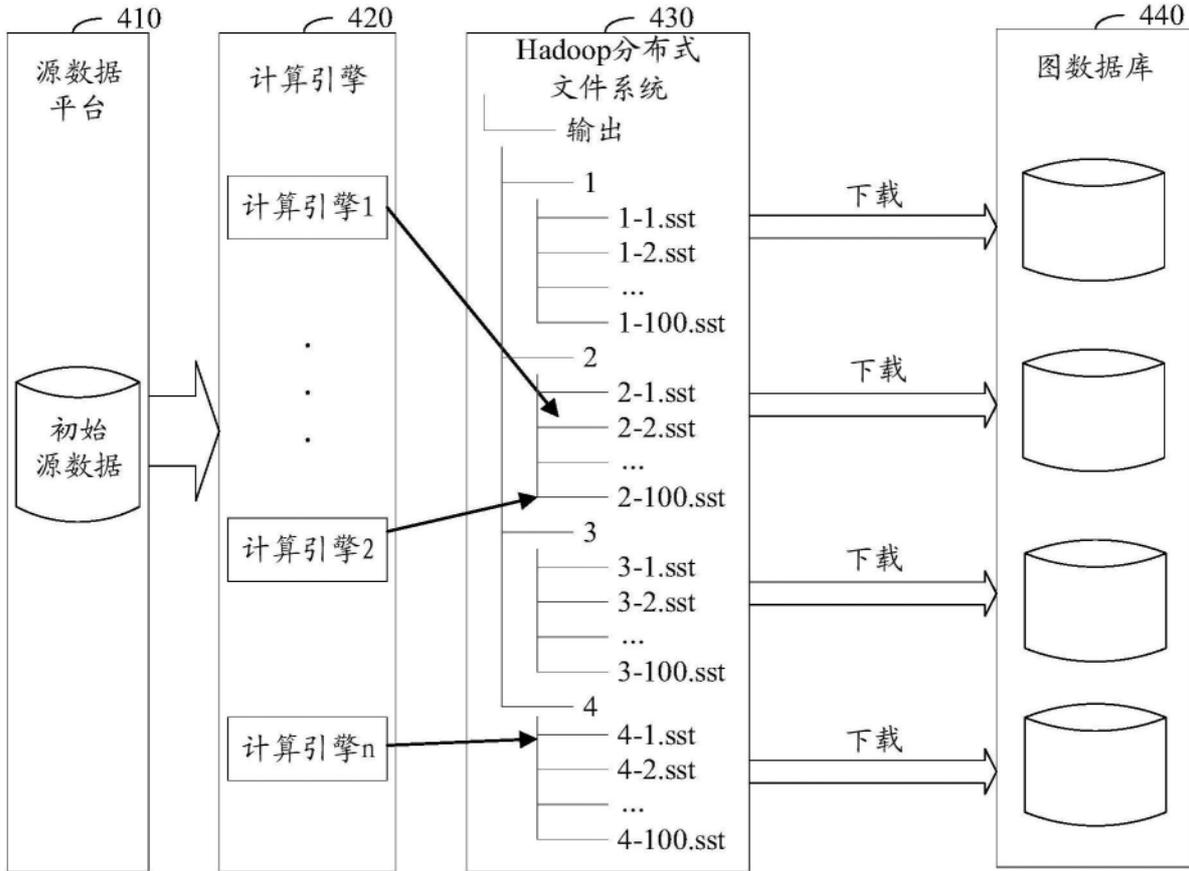


图4

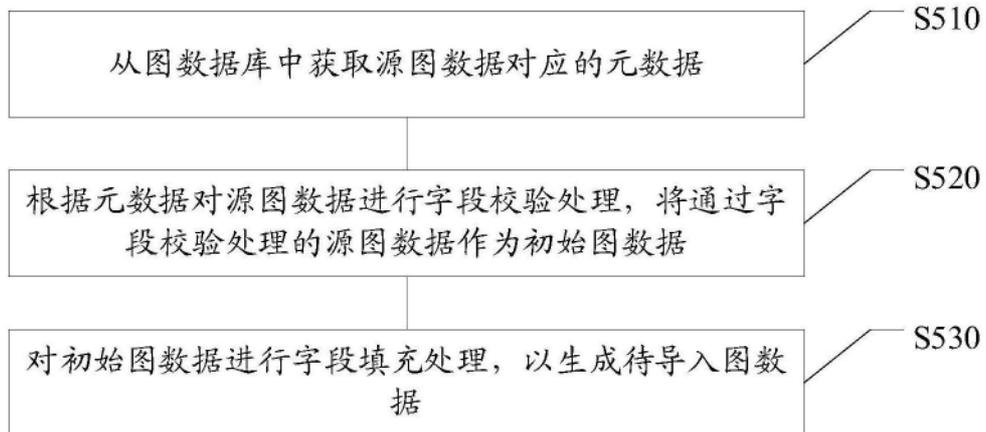


图5



图6



图7

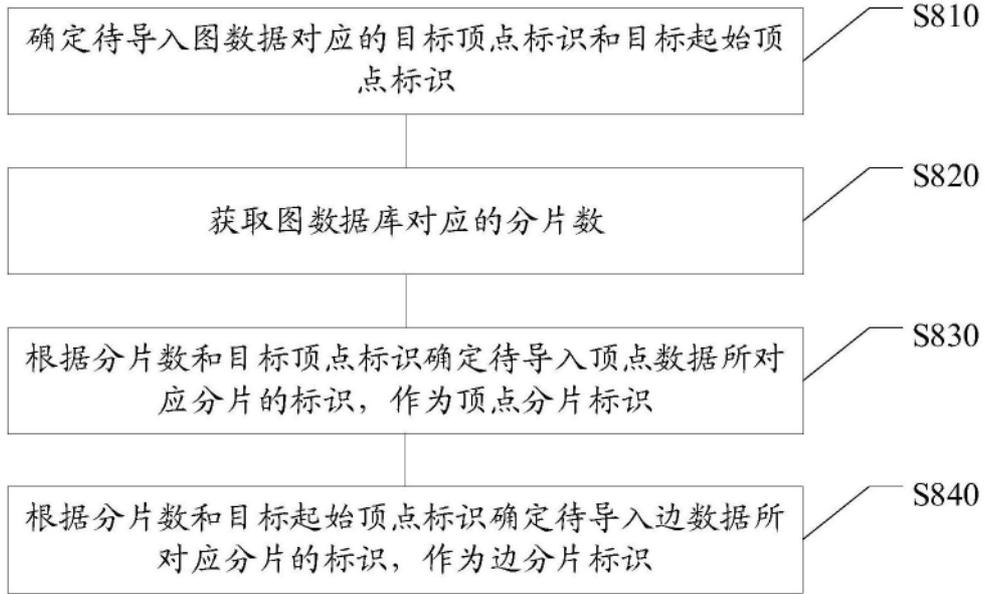


图8

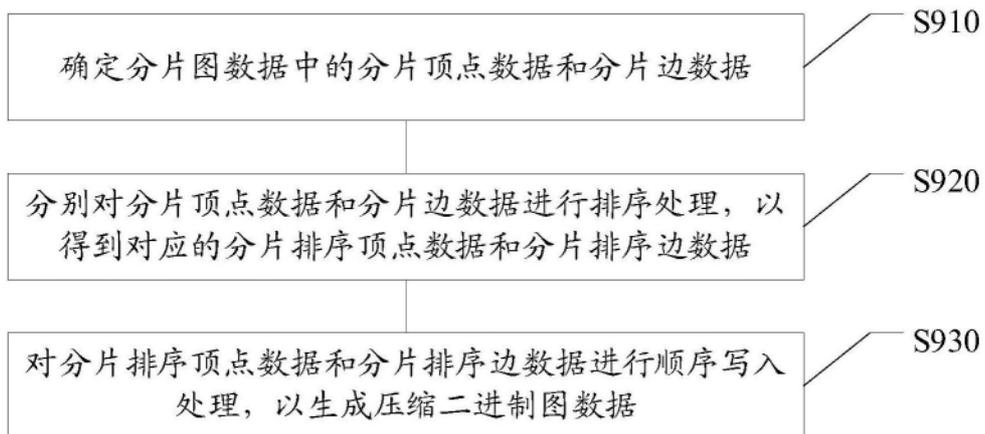


图9

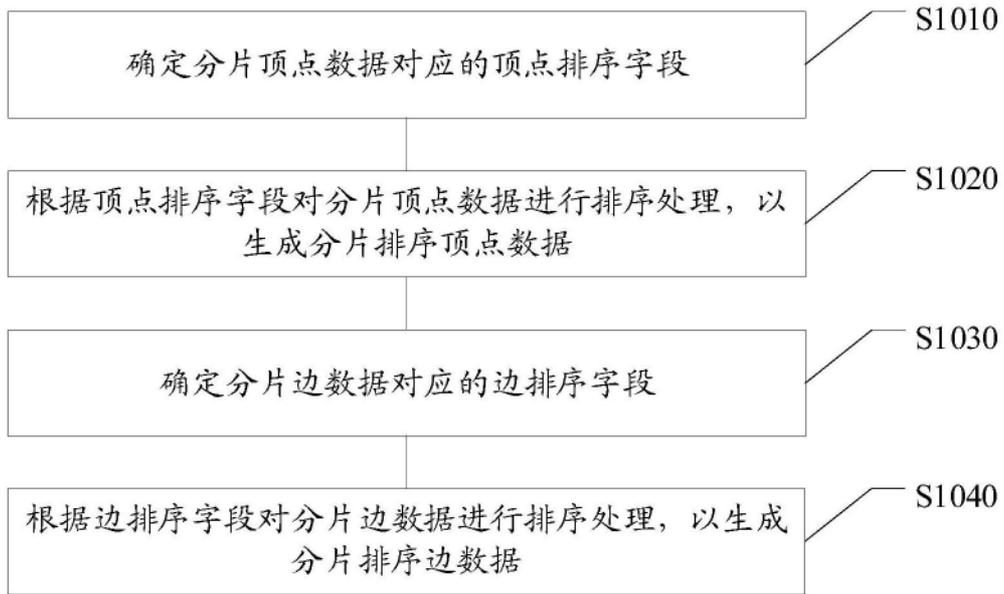


图10

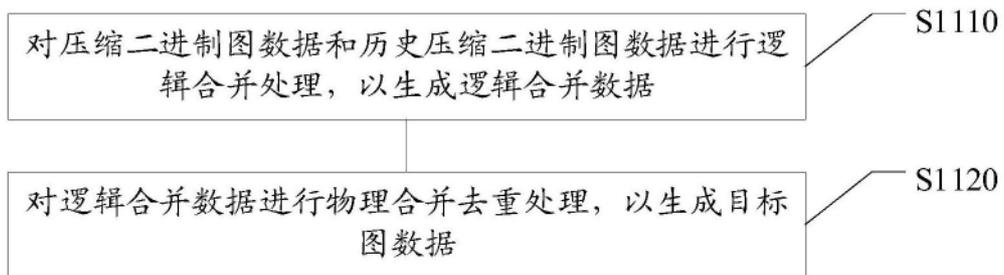


图11

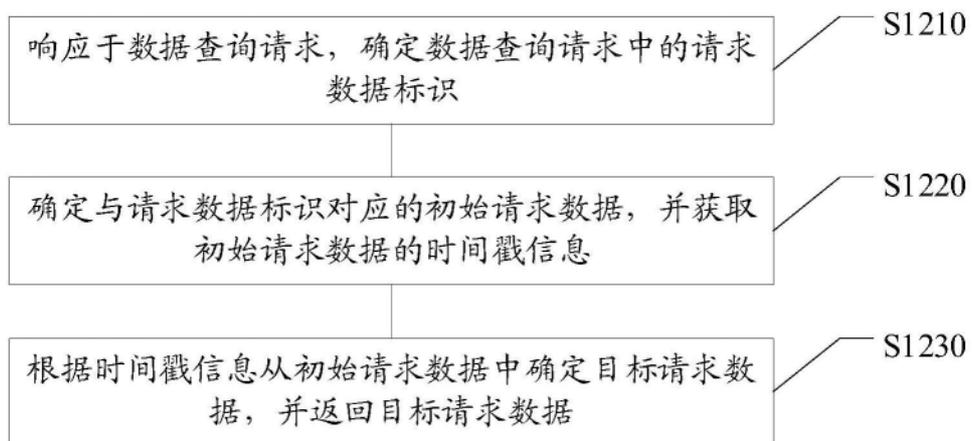


图12

1300



图13