

分布式图数据处理系统、方法、装置、设备及存储介质

相关申请的交叉引用

本公开是以中国申请号为 202211010841.1，申请日为 2022 年 08 月 23 日的申请为基础，并主张其优先权，该中国申请的公开内容在此作为整体引入本公开中。

技术领域

本公开涉及数据处理领域，尤其涉及一种分布式图数据处理系统、方法、装置、设备及存储介质。

背景技术

大数据时代下，社交网络、物联网、电子商务等应用产生大量的数据，这些数据组织成图的格式，形成图数据。而为了高效的处理规模较大的图数据，分布式图数据处理系统被提出。

由于图数据的数据量较大，对存储资源要求较高，因此，在基于分布式图数据处理系统执行图计算、图挖掘等图处理任务的过程中，需要对图数据进行存储和管理。

发明内容

第一方面，本公开提供了一种分布式图数据处理系统，所述方法包括：采用分布式架构的计算节点，所述计算节点包括具有对应关系的计算单元和持久化存储单元，所述计算单元中部署有存储管理器；所述计算单元，用于接收被分配的图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；所述存储管理器，用于将所述图数据处理任务对应的图数据邻接表划分为图数据块，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系；其中，所述图数据块包括所述图数据邻接表中的节点数据和边数据，所述图数据块的存储位置信息用于获取所述图数据块；所述持久化存储单元，用于基于所述存储位置信息对所述图数据块进行持久化存储，以支持与所述持久化存储单元对应的计算单元执行所述图数据处理任务。

第二方面，本公开提供了一种分布式图数据处理方法，所述方法包括：通过分布式架构中的计算节点接收图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；控制所述计算节点中的计算单元调用存储管理器，将所述图数据处理任务对应的图数据邻接表划分为图数据块；所述图数据块包括所

述图数据邻接表中的节点数据和边数据，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系；其中，所述图数据块的存储位置信息用于获取所述图数据块；基于所述存储位置信息将所述图数据块存储至持久化存储单元中；控制所述计算节点中的计算单元，基于所述图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

第三方面，本公开提供了一种分布式图数据处理装置，所述装置包括：第一获取模块，用于通过分布式架构中的计算节点接收图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；第一存储模块，用于控制所述计算节点中的计算单元调用存储管理器，将所述图数据处理任务对应的图数据邻接表划分为图数据块；所述图数据块包括所述图数据邻接表中的节点数据和边数据，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系；其中，所述图数据块的存储位置信息用于获取所述图数据块；第二存储模块，用于基于所述存储位置信息将所述图数据块存储至持久化存储单元中；图数据处理模块，用于控制所述计算节点中的计算单元，基于所述图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

第四方面，本公开提供了一种计算机可读存储介质，所述计算机可读存储介质中存储有指令，当所述指令在终端设备上运行时，使得所述终端设备实现上述的方法。

第五方面，本公开提供了一种数据处理设备，包括：存储器，处理器，及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时，实现上述的方法。

第六方面，本公开提供了一种计算机程序产品，所述计算机程序产品包括计算机程序/指令，所述计算机程序/指令被处理器执行时实现上述的方法。

附图说明

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。

为了更清楚地说明本公开实施例或相关技术中的技术方案，下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，显而易见地，对于本领域普通技术人员而言，

在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本公开实施例提供的一种分布式图数据处理系统的结构示意图。

图 2 为本公开实施例提供的一种节点邻接表和边邻接表的示意图。

图 3 为本公开实施例提供的一种基于图数据挖掘三角形的搜索树的示意图。

5 图 4 为本公开实施例提供的一种分布式图数据处理系统的局部架构图。

图 5 为本公开实施例提供的一种分布式图数据处理方法流程图。

图 6 为本公开实施例提供的一种分布式图数据处理装置的结构示意图。

图 7 为本公开实施例提供的一种分布式图数据处理设备的结构示意图。

具体实施方式

10 为了能够更清楚地理解本公开的上述目的、特征和优点，下面将对本公开的方案进行进一步描述。需要说明的是，在不冲突的情况下，本公开的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本公开，但本公开还可以采用其他不同于在此描述的方式来实施；显然，说明书中的实施例只是本公开的一部分实施例，
15 而不是全部的实施例。

本公开实施例提供了一种分布式图数据处理系统，参考图 1，为本公开实施例提供的一种分布式图数据处理系统的结构示意图，其中，该分布式图数据处理系统 10 包括采用分布式架构的计算节点，如图 1 所示的计算节点 101、计算节点 102 和计算节点 103，分别属于采用分布式架构的计算节点。每个计算节点包括具有对应关系的计算单元和持久化存储单元，每个计算单元中部署有存储管理器。如图 1 所示，以计算节点 101 为例，包括计算
20 单元 1011 和持久化存储单元 1012，计算单元 1011 中部署有存储管理器 1013。

所述计算单元，用于接收被分配的图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表。

如图 1 所示，分布式图数据处理系统 10 中部署有弹性引擎 100，用于接收用户下发的图数据处理任务，并基于图数据处理任务的负载，分配计算资源和存储资源等。在为图
25 数据处理任务分配具有计算单元和持久化存储单元的计算节点之后，基于被分配的计算节点执行该图数据处理任务。

针对每个计算节点中的计算单元，在接收到被分配的图数据处理任务之后，首先从远程文件系统中获取该图数据处理任务对应的图数据分片。如图 1 所示的远程文件系统 110，

用于以文件格式存储图数据处理任务对应的原始图数据。

本公开实施例中，计算单元从远程文件系统中获取到图数据处理任务对应的图数据分片之后，将图数据分片存储于对应的图数据邻接表。

本公开实施例中的图数据处理任务可以为图计算任务、图挖掘任务等。

- 5 图计算任务可以包括基于像顶点一样思考（Think Like A Vertex，TLAV）编程范式的迭代图计算任务，如 PageRank 算法、社区检测算法等相关图计算任务。

图挖掘任务是指从社交网络、深网等图数据中发现用户感兴趣的信息的过程，图挖掘结果（如三角形、频繁子图、图模式等）可以直接用于图数据分析，也可以间接用于其他与图数据相关的任务，因此，图挖掘在科学和商业中的应用场景非常广泛和实用。

- 10 图计算任务和图挖掘任务均是基于图数据邻接表中存储的图数据执行的，图数据邻接表中的图数据包括节点数据和边数据。其中，图数据邻接表包括节点邻接表 Vertex table 和边邻接表 Edge table，节点数据是基于节点邻接表 Vertex table 存储，而边数据是利用边邻接表 Edge table 以连续的方式存储每个节点的边，保留节点邻接表中各节点的顺序。本公开实施例采用简单的无共享设计，在分布式环境中，对节点邻接表在分布式集群中进行分区，每个分区都与其对应的边相关联。为了实现负载均衡的最大并发，本公开实施例允许系统用户重写图分区器，根据图数据的节点或边进行分区实现图数据块的划分。如图 2 所示，为节点邻接表和边邻接表的示意图。
- 15

- 所述存储管理器，用于将所述图数据处理任务对应的图数据邻接表划分为图数据块，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系；其中，所述图数据块包括所述图数据邻接表中的节点数据和边数据，所述图数据块的存储位置信息用于获取所述图数据块。
- 20

存储管理器基于预设规则计算各图数据块对应的键值 key，并由计算单元将各图数据块存入持久化存储单元中，实现基于 key 的图数据存储。其中，预设规则可以是基于哈希 hash 函数的键值生成规则等。

- 25 存储管理器在为各图数据块计算对应的 key 之后，记录并存储图数据块的 key 与图数据块的存储位置信息之间的对应关系，以便后续可以基于 key 从持久化存储单元中读取图数据块。

另外，所述持久化存储单元，用于基于所述存储位置信息对所述图数据块进行持久化存储，以支持与所述持久化单元对应的所述计算单元执行所述图数据处理任务。

其中，持久化存储单元可以采用内存和持久化存储介质（如持久内存 Optane Persistent Memory, PMEM）的分层结构，或者也可以采用内存、持久化存储介质和硬盘的分层结构实现，相应的，持久化存储单元具体用于按照内存、持久化存储介质和硬盘三级存储层的优先级降序，将图数据块存储至对应的存储层。其中，硬盘可以为固态硬盘（Solid State Drive, SSD）等。

本公开实施例针对不同量级大小的图数据，可以利用不同的存储介质进行存储。具体的，当图数据的量级较小时，可以优先利用内存对其进行存储，以实现数据读取的较佳效率；对于中等量级的图数据，可以将大部分数据存储于持久内存 PMEM 中，PMEM 具有容量大且性能接近内存的优势，因此能够保证稍大量级的图数据的存储要求；对于较大量级的图数据，可以利用容量较大的硬盘（如 SSD）进行存储，通过持续但不太频繁的数据溢出，实现将较大量级的图数据存储至硬盘中。

一些实施例中，持久化存储单元可以基于内存（也可以称为主存储器，例如 DRAM）、持久化存储介质（例如 PMEM）和硬盘（例如 SSD）的混合存储进行设计。具体的，由于主存储器具有访问速度快的优势，因此持久化存储单元优先提供主存储器，其次，持久化存储介质 PMEM 的访问速度快于硬盘，因此，持久化存储单元在主存储器被占用的情况下，优先提供持久化存储介质如 PMEM，在 PMEM 被占用的情况下，提供硬盘（例如 SSD），从而实现分层存储的持久化存储单元，这样既能够保证存储空间访问速度，又能够保证存储空间足够大。另外，由于本公开实施例采用了分层持久存储的设计，使得每个计算节点能够处理的数据量增加，从而整体上降低了所需计算设备的数量，降低了系统成本。

实际应用中，为了性能优势，DRAM 用于存储元数据，例如 PMEM 层和 SSD 层中分别存储的图数据块的存储索引。如果未指定，图数据块可以首先写入 DRAM 层，然后异步写入 PMEM 层，如果 PMEM 层的空闲资源不足，则可以将图数据块从 PMEM 层刷新到 SSD 层，利用 SSD 层对其进行存储。

另一些实施例中，持久化存储单元 102 可以基于内存和持久化存储介质的混合存储进行设计，由于持久化存储介质如 PMEM 既能够保证访问速度，也能够提供远远大于内存的存储空间，因此，上述实施方式也能够同时保证存储空间的访问速度以及存储空间足够大。

另外，所述计算单元，还用于基于所述存储管理器中存储的图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

实际应用中，计算单元在执行图数据处理任务的过程中，通过存储管理器从持久化存储单元中读取所需的图数据，用于图数据处理任务的执行。具体的，存储管理器基于计算单元的图数据读取请求生成 key，确定 key 对应的存储位置信息，然后基于该计算单元从持久化存储单元中读取对应的图数据块，从而获取所需的图数据。

- 5 其中，计算单元可以包括第一计算单元和/或第二计算单元，第一计算单元用于执行图计算任务，第二计算单元用于执行图挖掘任务。

本公开实施例提供的分布式图数据处理系统，将图计算和图挖掘两种编程范式统一到同一个图数据处理系统，也就是说，基于分布式图数据处理系统 10 能够处理图计算任务和图挖掘任务。尽管图计算和图挖掘引擎在底层计算模式上有所不同，但本公开实施例的分布式图数据处理系统提供一个统一的接口，使它们对外保持一致，例如使它们与面向开发人员的高级 SDK 和低级组件（如分层存储和图数据索引）保持一致。

本公开实施例提供的分布式图数据处理系统，包括采用分布式架构的计算节点，每个计算节点中包括具有对应关系的计算单元和持久化存储单元，每个计算单元中部署有存储管理器。当计算单元接收图数据处理任务，并获取到该图数据处理任务对应的图数据分片之后，将图数据分片存储于对应的图数据邻接表。该计算单元中部署的存储管理器将图数据邻接表划分为图数据块，确定每个图数据块对应的键值之后，记录并存储图数据块的键值与存储位置信息之间的对应关系。该计算单元对应的持久化存储单元基于存储位置信息对图数据块进行持久化存储之后，该计算单元可以基于存储管理器中存储的图数据块的键值与存储位置信息之间的对应关系，从持久化存储单元中读取图数据块执行图数据处理任务。

20 可见，本公开实施例提供的分布式图数据处理系统，利用存储管理器将图数据存储于持久化存储单元中，由于持久化存储单元作为外存设备，能够满足数据量较大的图数据的存储需求，且对于图数据处理的效率影响较小，因此，本公开实施例提供的分布式图数据处理系统能够满足图数据处理对存储资源的需求。

25

在上述实施例的基础上，计算单元在执行图数据处理任务的过程中会产生中间状态数据，为了支持分布式图数据处理系统的容错功能等，还可以将图数据处理任务对应的中间状态数据进行持久化存储。

在一些实施例中，存储管理器，还用于接收所述图数据处理任务执行过程中产生的中

间状态数据，将中间状态数据划分为中间状态数据块，并确定各中间状态数据块对应的键值，以及对所述键值与所述中间状态数据块的存储位置信息之间的对应关系进行存储；其中，所述中间状态数据块的存储位置信息用于获取所述中间状态数据块。

持久化存储单元，还用于基于所述存储位置信息对所述中间状态数据块进行存储，用于支持分布式图数据处理系统的容错功能。

本公开实施例中，可以基于分层存储方式，将图数据处理任务执行过程中产生的中间状态数据存入持久化存储单元，具体可参照上述实施例。

假设图数据处理任务为图计算任务，则图数据处理任务对应的中间状态数据包括消息数据和中间状态快照数据，中间状态快照数据包括节点快照数据和消息快照数据。

实际应用中，由于图计算任务通常为迭代图计算任务，为了支持容错功能，本公开实施例可以按照预设频率获取并存储图数据处理任务对应的中间状态数据，例如每执行 100 轮图计算任务获取一次中间状态数据，后续在检测到分布式图数据处理系统发生异常时，可以基于最新一轮的中间状态数据实现图数据处理任务的回滚止损。在一些实施例中，对于基于中间状态数据进行容错的实现方式，本公开实施例不做限制。

假设图数据处理任务为图挖掘任务，则图数据处理任务对应的中间状态数据包括图数据处理任务对应的搜索树中的当前增长层级以及所述当前增长层级中的当前增长位置。

实际应用中，图挖掘任务通常是一个搜索树增长的过程，以图 3 为例，图 3 为本公开实施例提供的一种基于图数据挖掘三角形的搜索树的示意图。在基于图数据挖掘三角形的过程中，首先确定第一层级的节点，如图 3 所示的 Level-1 对应的节点，然后在第一层级的节点的基础上进行节点增长，得到第二层级的节点，如图 3 所示的 Level-2 对应的节点，进而在第二层级的节点的基础上进行节点增长，得到第三层级的节点，如图 3 所示的 Level-3 对应的节点。通过判断第三层级的节点是否满足三角形条件，确定图挖掘任务的结果。

在上述图挖掘任务执行的过程中，中间状态数据包括图数据处理任务对应的搜索树中的当前增长层级以及当前增长层级中的当前增长位置。假设在搜索树增长到第二层级的增长位置 301 时记录中间状态数据，此时的中间状态数据包括的搜索树中的当前增长层级是指第二层级，当前增长层级中的当前增长位置是指增长位置 301，由于增长位置 301 处于第二层级的第三个增长位置，因此可以将中间状态数据记录为 (Level-2; 3)。

本公开实施例提供的分布式图数据处理系统，不仅能够利用存储管理器将图数据存储于持久化存储单元中，从存储空间和处理效率两个维度，对图数据处理所需的存储资源进

行保障，还可以利用存储管理器将图数据处理任务执行过程中产生的中间状态数据进行持久化存储，支持分布式图数据处理系统的容错功能。因此，本公开实施例提供的分布式图数据处理系统能够满足图数据处理对存储资源的需求，以及支持图数据处理任务的回滚止损。

5

为了便于对本公开实施例提供的分布式图数据处理系统中的计算单元、持久化存储单元以及存储管理器进行进一步的理解，本公开实施例提供的一种分布式图数据处理系统的局部架构示意图，参考图 4 所示。

其中，分布式图数据处理系统中可以同时包括用于执行图计算任务的计算单元 401，以及用于执行图挖掘任务的计算单元 402。本公开实施例针对分布式图数据处理系统中的每个计算单元分别设置一个存储管理器，实现计算单元与持久化存储单元之间的解耦，能够灵活扩展应用层（即计算单元）和存储层（即持久化存储单元）之间的可用的存储介质。如图 4 所示，计算单元 401 部署有存储管理器 403，计算单元 402 部署有存储管理器 404。

存储管理器中有两个重要部分，即基于规则的索引管理器和索引表，基于规则的索引管理器用于定义索引类型，然后将索引类型作为前缀和索引值进行组合得到整个键值，用于唯一标识图数据块或者中间状态数据块。索引表可以是采用并发哈希表实现，用于存储键值和图数据块/中间状态数据块的存储地址之间的对应关系。

基于索引管理器和索引表，可以将图数据块和中间状态数据块存储在持久化存储单元中，如图 4 所示，计算单元 401 可以将图数据块和中间状态数据块存储于持久化存储单元 405，计算单元 402 可以将图数据块和中间状态数据块存储于持久化存储单元 406。各个持久化存储单元可以基于分层结构对图数据块和中间状态数据块进行存储。在一些实施例中，具有分层结构的持久化存储单元可参照上述实施例进行理解，在此不再赘述。

基于对上述分布式图数据处理系统的理解，本公开实施例还提供了一种分布式图数据处理方法，参考图 5，为本公开实施例提供的一种分布式图数据处理方法流程图。

在一些实施例中，本公开实施例提供的分布式图数据处理方法可以应用于分布式图数据处理系统中的计算单元，该计算单元挂载有持久化存储单元，该方法包括：S501：通过分布式架构中的计算节点接收图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表。

其中，所述图数据邻接表中包括所述图数据处理任务对应的图数据分片，所述图数据块包括所述图数据邻接表中的节点数据和边数据。

本公开实施例中，计算单元在接收到被分配的图数据处理任务之后，从远程文件系统获取图数据处理任务对应的图数据分片，然后利用图数据邻接表对该图数据分片进行存储，
5 并将存储有图数据分片的图数据邻接表划分为图数据块，具体的划分方式本公开实施例不做限制。

本公开实施例中的计算单元部署有存储管理器，由存储管理器对图数据处理任务对应的图数据邻接表进行划分，得到图数据块。

S502：控制所述计算节点中的计算单元调用存储管理器，将所述图数据处理任务对应的图数据邻接表划分为图数据块；所述图数据块包括所述图数据邻接表中的节点数据和边数据，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系。
10

其中，所述图数据块的存储位置信息用于获取所述图数据块。

本公开实施例中，在划分得到各个图数据块之后，确定各个图数据块分别对应的键值
15 key。其中，各图数据块对应的 key 由前缀和图数据块 ID 组合而成，前缀可以为索引类型，索引类型可以基于图数据块中存储的数据类型确定，例如，节点数据对应的索引类型为 kVertex，边数据对应的索引类型可以为 kEdge。在前缀后增加图数据块 ID，以确定同一索引类型的不同图数据块。另外，基于索引类型的前缀，能够对同一索引类型的图数据块进行读取等操作，提升图数据读取的效率。

20 另外，为了便于后续对图数据块的读取，可以将图数据块的键值与存储位置信息之间的对应关系进行持久化存储。基于该对应关系，可以通过图数据块的键值确定该图数据块对应的存储位置信息，进而基于该存储位置信息读取到该图数据块中的图数据。

一些实施例中，可以基于哈希函数为各图数据块生成键值，用于作为图数据块的存储索引。另外，对于其他生成键值的方式本公开实施例不做限定。

25 S503：基于所述存储位置信息将所述图数据块存储至持久化存储单元中。

S504：控制所述计算节点中的计算单元，基于所述图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

其中，用于持久化存储图数据块的存储单元可以基于分层结构实现，例如分层结构可

以采用内存、持久化存储介质如 PMEM 和硬盘，或者也可以采用内存和持久化存储介质如 PMEM，其中，内存的存储优先级高于持久化存储介质，持久化存储介质的存储优先级高于硬盘。

一些实施例中，针对分层结构采用内存和持久化存储介质 PMEM 实现的情况，本公开
5 开实施例在接收到图数据块之后，优先利用内存基于所述存储位置信息对所述图数据块进行存储；如果确定所述内存中的空闲资源不足，则利用持久化存储介质对所述图数据块进行存储。

另外，针对分层结构采用内存、持久化存储介质如 PMEM 和硬盘实现的情况，本公开
10 开实施例在接收到图数据块之后，优先利用内存基于所述存储位置信息对所述图数据块进行存储；如果确定所述内存中的空闲资源不足，则再利用持久化存储介质对所述图数据块进行存储；如果确定所述持久化存储介质中的空闲资源不足，则再利用硬盘对所述图数据块进行存储。

本公开实施例提供的分布式图数据处理方法，将图数据存储于能够满足数据量较大的图数据的存储需求的持久化存储单元，且对于图数据处理的效率影响较小，因此，本公开
15 实施例提供的分布式图数据处理系统能够满足图数据处理对存储资源的需求。

在上述实施例的基础上，为了支持分布式图数据处理方法的容错功能，本公开实施例还可以对图数据处理任务执行过程中产生的中间状态数据进行持久化存储，以便后续用于分布式图数据处理方法的容错处理中。

20 实际应用中，控制计算节点中的计算单元调用存储管理器，将图数据处理任务执行过程中产生的中间状态数据划分为中间状态数据块；然后，确定每个中间状态数据块对应的键值，并对所述键值与所述中间状态数据块的存储位置信息之间的对应关系进行存储；其中，所述中间状态数据块的存储位置信息用于获取所述中间状态数据块；进而基于所述存储位置信息对所述中间状态数据块进行持久化存储。例如，中间状态数据块可以以键值形式存储
25 在持久内存 PMEM 上。

另外，本公开实施例可以基于持久化存储单元中的中间状态数据块，对分布式图数据处理系统进行容错处理。

在一些实施例中，持久化存储单元可以采用分层结构实现，分层结构可以包括内存、持久化存储介质如 PMEM 和硬盘，或者包括内存和持久化存储介质如 PMEM，且存储优

优先级为内存最高且硬盘最低。

一些实施例中，由于 PMEM 的访问速度快于硬盘，因此，在 PMEM 中存在空闲存储资源的情况下，优先将中间状态数据块存储于 PMEM 中，在 PMEM 中无空闲存储资源的情况下，将中间状态数据块存储于硬盘（如 SSD）中，实现分层存储。

5 一些实施例中，图数据处理任务可以为图计算任务，也可以为图挖掘任务。

假设图数据处理任务为图计算任务，图数据处理任务对应的中间状态数据包括消息数据和中间状态快照数据，所述中间状态快照数据包括节点快照数据和消息快照数据。消息数据是指图计算任务执行过程中，当前计算节点与其他计算节点之间传输的消息，用于当前计算节点的图计算任务的执行。中间状态快照数据是指每执行 N 轮图计算任务时，针对
10 当前轮的节点数据和消息数据获取的快照数据，用于容错功能的回滚止损。其中，N 可以基于需求进行设置。

假设图数据处理任务为图挖掘任务，图数据处理任务对应的中间状态数据包括搜索树中的当前增长层级以及所述当前增长层级中的当前增长位置。执行图挖掘任务是一个搜索树增长的过程，为了执行图挖掘任务执行过程中的容错，计算单元在执行图挖掘任务的过程
15 中，按照预设频率获取搜索树的当前增长层级以及当前增长层级中的当前增长位置，用于容错功能的回滚止损。

基于上述系统和方法实施例，本公开还提供了一种分布式图数据处理装置，参考图 6，
为本公开实施例提供的一种分布式图数据处理装置的结构示意图，所述装置包括：第一获取
20 模块 601，用于通过分布式架构中的计算节点接收图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；第一存储模块 602，用于控制所述计算节点中的计算单元调用存储管理器，将所述图数据处理任务对应的图数据邻接表划分为图数据块；所述图数据块包括所述图数据邻接表中的节点数据和边数据，确定每个图数据块对应的键值，记录并存储图数据块的键值与
25 所述图数据块的存储位置信息之间的对应关系；其中，所述图数据块的存储位置信息用于获取所述图数据块；第二存储模块 603，用于基于所述存储位置信息将所述图数据块存储至持久化存储单元中；图数据处理模块 604，用于控制所述计算节点中的计算单元，基于所述图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

一些实施例中，所述装置还包括：划分模块，用于控制所述计算节点中的计算单元调用所述存储管理器，将所述图数据处理任务执行过程中产生的中间状态数据划分为中间状态数据块；第三存储模块，用于确定每个中间状态数据块对应的键值，记录并存储所述中间状态数据块的键值与所述中间状态数据块的存储位置信息之间的对应关系进行存储；其中，所述中间状态数据块的存储位置信息用于获取所述中间状态数据块；第四存储模块，用于基于所述存储位置信息将所述中间状态数据块存储至所述持久化存储单元中。

一些实施例中，所述图数据处理任务为图计算任务，所述图数据处理任务对应的中间状态数据包括消息数据和中间状态快照数据，所述中间状态快照数据包括节点快照数据和消息快照数据，所述消息数据为所述图数据处理任务执行过程中各计算节点之间传输的计算结果数据，所述中间状态快照数据用于所述图数据处理任务的恢复执行。

一些实施例中，所述图数据处理任务为图挖掘任务，所述图数据处理任务对应的中间状态数据包括所述图数据处理任务对应的搜索树中的当前增长层级以及所述当前增长层级中的当前增长位置。

一些实施例中，所述第二存储模块，包括：第一存储子模块，用于在接收到所述图数据块之后，优先利用内存基于所述存储位置信息对所述图数据块进行存储；第二存储子模块，用于如果确定所述内存中的空闲资源不足，则再利用持久化存储介质对所述图数据块进行存储；第三存储子模块，用于如果确定所述持久化存储介质中的空闲资源不足，则再利用硬盘对所述图数据块进行存储。

一些实施例中，所述第二存储模块，包括：第四存储子模块，用于在接收到所述图数据块之后，优先利用内存基于所述存储位置信息对所述图数据块进行存储；第五存储子模块，用于如果确定所述内存中的空闲资源不足，则利用持久化存储介质对所述图数据块进行存储。

一些实施例中，所述持久化存储介质包括持久内存。

本公开实施例提供的分布式图数据处理装置，不仅能够利用存储管理器将图数据存储于持久化存储单元中，从存储空间和处理效率两个维度对图数据处理所需的存储资源进行保障，还可以利用存储管理器将图数据处理任务执行过程中产生的中间状态数据进行持久化存储，支持分布式图数据处理系统的容错功能。因此，本公开实施例提供的分布式图数据处理装置能够满足图数据处理对存储资源的需求以及支持图数据处理任务的回滚止损。

除了上述方法和装置以外，本公开实施例还提供了一种计算机可读存储介质，计算机可读存储介质中存储有指令，当所述指令在终端设备上运行时，使得所述终端设备实现本公开实施例所述的分布式图数据处理方法。

本公开实施例还提供了一种计算机程序产品，所述计算机程序产品包括计算机程序/指令，所述计算机程序/指令被处理器执行时实现本公开实施例所述的分布式图数据处理方法。

另外，本公开实施例还提供了一种分布式图数据处理设备，参见图 7 所示，可以包括：处理器 701、存储器 702、输入装置 703 和输出装置 704。分布式图数据处理设备中的处理器 701 的数量可以一个或多个，图 7 中以一个处理器为例。在本公开的一些实施例中，处理器 701、存储器 702、输入装置 703 和输出装置 704 可通过总线或其它方式连接，其中，图 7 中以通过总线连接为例。

存储器 702 可用于存储软件程序以及模块，处理器 701 通过运行存储在存储器 702 的软件程序以及模块，从而执行分布式图数据处理设备的各种功能应用以及数据处理。存储器 702 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序等。此外，存储器 702 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。输入装置 703 可用于接收输入的数字或字符信息，以及产生与分布式图数据处理设备的用户设置以及功能控制有关的信号输入。

具体在本实施例中，处理器 701 会按照如下的指令，将一个或一个以上的应用程序的进程对应的可执行文件加载到存储器 702 中，并由处理器 701 来运行存储在存储器 702 中的应用程序，从而实现上述分布式图数据处理设备的各种功能。

需要说明的是，在本文中，诸如“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

以上所述仅是本公开的具体实施方式，使本领域技术人员能够理解或实现本公开。对这些实施例的多种修改对本领域的技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本公开的精神或范围的情况下，在其它实施例中实现。因此，本公开将不会被限制于本文所述的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

5

权 利 要 求

1、一种分布式图数据处理系统，包括：

采用分布式架构的计算节点，所述计算节点包括具有对应关系的计算单元和持久化存储单元，所述计算单元中部署有存储管理器；

所述计算单元，用于接收被分配的图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；

所述存储管理器，用于将所述图数据处理任务对应的图数据邻接表划分为图数据块，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系，其中，所述图数据块包括所述图数据邻接表中的节点数据和边数据，所述图数据块的存储位置信息用于获取所述图数据块；

所述持久化存储单元，用于基于所述存储位置信息对所述图数据块进行持久化存储，以支持与所述持久化存储单元对应的计算单元执行所述图数据处理任务。

2、根据权利要求 1 所述的分布式图数据处理系统，其中：

所述存储管理器，还用于接收所述图数据处理任务执行过程中产生的中间状态数据，将所述中间状态数据划分为中间状态数据块，并确定所述中间状态数据块对应的键值，以及对所述键值与所述中间状态数据块的存储位置信息之间的对应关系进行存储，其中，所述中间状态数据块的存储位置信息用于获取所述中间状态数据块；

所述持久化存储单元，还用于基于所述存储位置信息对所述中间状态数据块进行存储。

3、根据权利要求 2 所述的分布式图数据处理系统，其中，所述图数据处理任务为图计算任务，所述图数据处理任务对应的中间状态数据包括消息数据和中间状态快照数据，所述中间状态快照数据包括节点快照数据和消息快照数据，所述消息数据为所述图数据处理任务执行过程中各计算节点之间传输的计算结果数据，所述中间状态快照数据用于所述图数据处理任务的恢复执行。

4、根据权利要求 2 或 3 所述的分布式图数据处理系统，其中，所述图数据处理任务为图挖掘任务，所述图数据处理任务对应的中间状态数据包括所述图数据处理任务对应的搜索树中的当前增长层级以及所述当前增长层级中的当前增长位置。

5、根据权利要求 1-4 任一项所述的分布式图数据处理系统，其中，所述持久化存储单元采用内存、持久化存储介质和硬盘的分层结构；

所述持久化存储单元，用于按照内存、持久化存储介质和硬盘三级存储层的优先级降序，将所述图数据块存储至对应的存储层。

6、根据权利要求 1-5 任一项所述的分布式图数据处理系统，其中，所述持久化存储单元采用内存和持久化存储介质的分层结构；

所述持久化存储单元，用于在接收到所述图数据块之后，优先利用所述内存进行存储，并在所述内存中的空闲资源不足时，再利用所述持久化存储介质对所述图数据块进行存储。

7、根据权利要求 5 或 6 所述的分布式图数据处理系统，其中，所述持久化存储介质包括持久内存。

8、根据权利要求 1-7 任一项所述的分布式图数据处理系统，其中，所述计算单元包括第一计算单元和第二计算单元的至少一种，所述图数据处理任务包括图计算任务和图挖掘任务的至少一种；

所述第一计算单元，用于执行所述图计算任务；

所述第二计算单元，用于执行所述图挖掘任务。

9、一种分布式图数据处理方法，包括：

通过分布式架构中的计算节点接收图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；

控制所述计算节点中的计算单元调用存储管理器，将所述图数据处理任务对应的图数据邻接表划分为图数据块，其中，所述图数据块包括所述图数据邻接表中的节点数据和边数据，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系，其中，所述图数据块的存储位置信息用于获取所述图数据块；

基于所述存储位置信息，将所述图数据块存储至持久化存储单元中；

控制所述计算节点中的计算单元，基于所述图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

10、根据权利要求 9 所述的分布式图数据处理方法，还包括：

控制所述计算节点中的计算单元调用所述存储管理器，将所述图数据处理任务执行过程中产生的中间状态数据划分为中间状态数据块；

确定每个中间状态数据块对应的键值，记录并存储所述中间状态数据块的键值与所述中间状态数据块的存储位置信息之间的对应关系进行存储，其中，所述中间状态数据块的

存储位置信息用于获取所述中间状态数据块；

基于所述存储位置信息，将所述中间状态数据块存储至所述持久化存储单元中。

11、根据权利要求 10 所述的分布式图数据处理方法，其中，所述图数据处理任务为图计算任务，所述图数据处理任务对应的中间状态数据包括消息数据和中间状态快照数据，所述中间状态快照数据包括节点快照数据和消息快照数据，所述消息数据为所述图数据处理任务执行过程中各计算节点之间传输的计算结果数据，所述中间状态快照数据用于所述图数据处理任务的恢复执行。

12、根据权利要求 10 或 11 所述的分布式图数据处理方法，其中，所述图数据处理任务为图挖掘任务，所述图数据处理任务对应的中间状态数据包括所述图数据处理任务对应的搜索树中的当前增长层级以及所述当前增长层级中的当前增长位置。

13、根据权利要求 9-12 中的任一项所述的分布式图数据处理方法，其中，所述基于所述存储位置信息，将所述图数据块存储至持久化存储单元中，包括：

在接收到所述图数据块之后，优先利用内存基于所述存储位置信息对所述图数据块进行存储；

在确定所述内存中的空闲资源不足的情况下，再利用持久化存储介质对所述图数据块进行存储；

在确定所述持久化存储介质中的空闲资源不足的情况下，再利用硬盘对所述图数据块进行存储。

14、根据权利要求 9-13 中的任一项所述的分布式图数据处理方法，其中，所述基于所述存储位置信息，将所述图数据块存储至持久化存储单元中，包括：

在接收到所述图数据块之后，优先利用内存基于所述存储位置信息对所述图数据块进行存储；

在确定所述内存中的空闲资源不足的情况下，利用持久化存储介质对所述图数据块进行存储。

15、根据权利要求 13 或 14 所述的分布式图数据处理方法，其中，所述持久化存储介质包括持久内存。

16、一种分布式图数据处理装置，包括：

第一获取模块，用于通过分布式架构中的计算节点接收图数据处理任务，并获取所述图数据处理任务对应的图数据分片，将所述图数据分片存储于对应的图数据邻接表；

第一存储模块，用于控制所述计算节点中的计算单元调用存储管理器，将所述图数据处理任务对应的图数据邻接表划分为图数据块，其中，所述图数据块包括所述图数据邻接表中的节点数据和边数据，确定每个图数据块对应的键值，记录并存储图数据块的键值与所述图数据块的存储位置信息之间的对应关系，其中，所述图数据块的存储位置信息用于获取所述图数据块；

第二存储模块，用于基于所述存储位置信息将所述图数据块存储至持久化存储单元中；

图数据处理模块，用于控制所述计算节点中的计算单元，基于所述图数据块的键值与图数据块的存储位置信息之间的对应关系，从所述持久化存储单元中读取所述图数据处理任务对应的图数据块进行图数据处理。

17、一种计算机可读存储介质，所述计算机可读存储介质中存储有指令，当所述指令在终端设备上运行时，使得所述终端设备实现如权利要求 9-15 任一项所述的分布式图数据处理方法。

18、一种分布式图数据处理设备，包括：

存储器；以及

耦接至所述存储器的处理器，所述处理器被配置为基于存储在所述存储器的指令，执行如权利要求 9-15 任一项所述的分布式图数据处理方法。

19、一种计算机程序产品，所述计算机程序产品包括计算机程序/指令，所述计算机程序/指令被处理器执行时实现如权利要求 9-15 任一项所述的分布式图数据处理方法。

20、一种计算机程序，包括：

指令，所述指令当由处理器执行时使所述处理器执行根据权利要求 9-15 任一项所述的分布式图数据处理方法。