

数据存储方法及装置、数据读取方法及装置

5 本申请要求于 2022 年 03 月 02 日提交中国专利局、申请号为 202210197317.3、发明名称为“数据存储方法及装置、数据读取方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本说明书实施例涉及计算机技术领域，特别涉及一种数据存储方法、一种数据读取方法。

背景技术

10 知识图谱(Knowledge Graph)，在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。

15 目前，知识图谱的存储结构设计没有统一的标准，对于数据量不是很大且结构固定的图谱，通常使用传统数据库和关系表进行存储。但是，在数据量大的情况下，一个实体通常会包含众多属性，如果同时计算这些属性并存储到存储介质中，会大大降低数据计算、存储、检索的效率。

发明内容

有鉴于此，本说明书实施例提供了一种数据存储方法、一种数据读取方法。本说明书一个或者多个实施例同时涉及一种数据存储装置、一种数据读取装置，一种计算设备，一种计算机可读存储介质，一种计算机程序，以解决现有技术中存在的技术缺陷。

20 根据本说明书实施例的第一方面，提供了一种数据存储方法，应用于知识图谱平台，包括：

拆分目标知识图谱数据，确定至少两个待存储目标子图数据；

其中，所述目标知识图谱数据包括目标实体节点以及至少一个与所述目标实体节点相关联的边，各个所述待存储目标子图数据包括所述目标实体节点以及至少一个目标属性的边；

25 将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块，其中，前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目标子图数据的初始实体标识相同。

根据本说明书实施例的第二方面，提供了一种数据存储装置，应用于知识图谱平台，包括：

30 图谱拆分模块，被配置为拆分目标知识图谱数据，确定至少两个待存储目标子图数据，其中，所述目标知识图谱数据包括目标实体节点以及至少一个与所述目标实体节点相关联的边，各个所述待存储目标子图数据包括所述目标实体节点以及至少一个目标属性的边；

数据存储模块，被配置为将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块，其中，前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目标子图数据的初始实体标识相同。

根据本说明书实施例的第三方面，提供了一种数据读取方法，应用于知识图谱平台，包括：

35 接收针对目标数据的数据读取请求，基于所述数据读取请求在索引数组信息中确定目标存储位置；

其中，所述索引数组信息基于每个数据块中的索引信息而确定，并包括每个数据块中存储的图谱数

据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件；

基于所述目标存储位置从目标数据块中读取所述目标数据。

根据本说明书实施例的第四方面，提供了一种数据读取装置，应用于知识图谱平台，包括：

5 存储位置确定模块，被配置为接收针对目标数据的数据读取请求，基于所述数据读取请求在索引数组信息中确定目标存储位置；

其中，所述索引数组信息基于每个数据块中的索引信息而确定，并包括每个数据块中存储的图谱数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件；

数据读取模块，被配置为基于所述目标存储位置从目标数据块中读取所述目标数据。

根据本说明书实施例的第五方面，提供了一种计算设备，包括：

10 存储器和处理器；

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令，该计算机可执行指令被处理器执行时实现上述方法的步骤。

根据本说明书实施例的第六方面，提供了一种计算机可读存储介质，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现上述方法的步骤。

15 根据本说明书实施例的第七方面，提供了一种计算机程序，其中，当所述计算机程序在计算机中执行时，令计算机执行上述方法的步骤。

本说明书一个实施例，通过拆分目标知识图谱数据，确定至少两个待存储目标子图数据；其中，所述目标知识图谱数据包括目标实体节点以及至少一个与目标实体节点相关联的边，各个所述待存储目标子图数据包括所述目标实体节点以及至少一个目标属性的边；将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块，其中，前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目标子图数据的初始实体标识相同。

20 具体的，通过对构建的目标知识图谱进行拆分，确定至少两个待存储目标子图数据，并对至少两个待存储目标子图数据按边的属性存储至至少两个连续数据块，以解决数据量大的知识图谱一次性存储效率低的问题，通过拆分再进行存储也提高了图谱数据的存储性能，也便于后续对数据量大的图谱数据提供快速检索与访问的能力。

附图说明

图1是本说明书一个实施例提供的一种数据存储方法的流程图；

图2是本说明书一个实施例提供的一种数据存储方法中目标知识图谱数据拆分为至少两个待存储目标子图数据的结构示意图；

30 图3是本说明书一个实施例提供的一种数据存储方法中将待存储目标子图数据存储至至少两个连续数据块的结构示意图；

图4是本说明书一个实施例提供的一种数据存储方法中将待存储目标子图数据存储至至少两个连续数据块中缓冲区的结构示意图；

图5是本说明书一个实施例提供的一种数据存储方法中索引数组信息的生成方式示意图；

35 图6是本说明书一个实施例提供的一种数据读取方法的流程图；

图 7 是本说明书一个实施例提供的一种数据读取方法随机读取数据的过程示意图；

图 8 是本说明书一个实施例提供的一种数据读取方法分区读取数据的过程示意图；

图 9 是本说明书一个实施例提供的一种数据存储装置的结构示意图；

图 10 是本说明书一个实施例提供的一种数据读取装置的结构示意图；

5 图 11 是本说明书一个实施例提供的一种计算设备的结构框图。

具体实施方式

在下面的描述中阐述了很多具体细节以便于充分理解本说明书。但是本说明书能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本说明书内涵的情况下做类似推广，因此本说明书不受下面公开的具体实施的限制。

10 在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本说明书一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

15 应当理解，尽管在本说明书一个或多个实施例中可能采用术语第一、第二等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一也可以被称为第二，类似地，第二也可以被称为第一。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

首先，对本说明书一个或多个实施例涉及的名词术语进行解释。

20 知识图谱(Knowledge Graph)，在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。

一跳子图(OneHop Graph)：从图的某个顶点出发，传播一跳到相邻节点，形成的子图。

热点数据：与实体节点相连接的边的数据量大于预设阈值的子图数据。

25 热点子图(Hotspot Graph)：为热点数据的一跳子图，即一个顶点有大规模的出边或入边，该子图为热点子图。

B-Tree(多路搜索树，并不是二叉的)：是一种常见的数据结构。使用 B-tree 结构可以显著减少定位记录时所经历的中间过程，从而加快存取速度。

30 在图模型中，数据热点是一个常见问题，热点问题的解决关系到图谱数据的存储和访问性能。通常在交易类、商家、安全等业务场景中，存在较多的数据热点情况，比如，用户的交易信息、用户和特征标签等，这些热点场景会导致某个顶点存在大量的出边或入边。本说明书实施例提供的数据存储方法，对于数据量大的图谱数据可以有效地组织和存储，并提供对数据量大的数据的快速访问能力，不仅能够将数据量大的图片数据完整地存储下来，还可便于后续对数据量大的数据提供快速随机访问、顺序扫描、热点边截断的能力。

35 本说明书一实施例还提供可以一种 K/V 存储的图存储解决方案，通常基于多级存储方式来解决热点数据问题，其中，热点数据问题可以理解为数据量较大的知识图谱数据，本说明书实施例还可提供一种基于 B-Tree 将热点边数据存储为 N 和 part，part 作为 B-Tree 的节点，数据访问时可通过 B-Tree 的查找机制读取对应的 part 数据块，从而获取要查找的边信息，需要说明的是，该热点解决方案适合作为图数据库的热点方案，能够实现热点边的快速随机访问，但缺陷就是 B-Tree 模型有一定的维护成本，需要充分地缓冲层来减少底层文件的随机访问，而在知识图谱数据平台中对数据的分析、规则推理批量加载图数据场景并无明显优势，而且还会增加文件的随机访问，导致 IO 性能降低，因此，本说明书实施例提供的数据存储方法，并未采用 B-Tree 来维护热点数据拆分后的多个 part 的存储，而是通过保证热点子图的边有序存储，然后将热点子图拆分为多个数据块连续存储，并在 meta(索引记录)中记录热点边信

40

息,进而能够保证存储热点数据的数据块的连续存储,在图的批量顺序读取场景避免文件查找操作,具有较高的性能,在图的随机查询场景中,通过对存储热点数据的数据块的二分查找进行快速定位和捞取部分变数据。与图数据库服务不同,本说明书实施例提供的数据存储方法是基于图数据的 layout 文件,提供了数据分析场景的随机查询能力,并未增加数据缓冲层,也不提供低时延的在线查询响应能力。

5 在本说明书中,提供了一种数据存储方法、一种数据读取方法,本说明书同时涉及一种数据存储装置,一种数据读取装置,一种计算设备,一种计算机可读存储介质以及一种计算机程序,在下面的实施例中逐一进行详细说明。

需要说明的是,本说明书实施例提供的数据存储方法应用于智能知识图谱平台,该智能知识图谱平台可提供图谱构建、图数据存储、图数据的规则推理等功能,本说明书实施例对具体的智能知识图谱平台不作任何限定,此外,在智能知识图谱平台存储子图数据片段时,可基于前置构建的链路的执行,比如,在针对全量的事件或交易类数据(可以理解为 K/V 数据)的情况下,将这些 K/V 数据转换成一跳子图(可以理解为分组操作,将一个实体以及实体的一跳邻居以顶点和边的形式组成一跳子图,且该一跳子图是一个顶点以及该顶点为中心的所有的出/入边);再对一跳子图数据进行切分,形成多个一跳子图片段,然后再依次写入连续的数据块中(若达不到切分阈值,则不会被切分)。

15 图 1 示出了根据本说明书一个实施例提供的一种数据存储方法的流程图,具体包括以下步骤。

步骤 102: 拆分目标知识图谱数据,确定至少两个待存储目标子图数据。

其中,所述目标知识图谱数据包括目标实体节点以及至少一个与所述目标实体节点相关联的边,各个所述待存储目标子图数据包括所述目标实体节点以及至少一个目标属性的边。

20 实际应用中,为了实现对大量的知识图谱数据快速地进行存储、以及为存储后的知识图谱数据提供快速的访问能力,智能知识图谱平台对于数据量较大的目标知识图谱数据,先进行拆分处理,拆分成至少两个待存储目标子图数据,目标知识图谱数据不仅包括待处理数据的目标实体节点,还包括至少一个与该目标实体节点相关联的边,其中,目标实体节点为待处理数据中多个实体节点之一,比如,待处理数据为店铺 A 的交易数据,那么,目标知识图谱数据的目标实体节点可以为店铺 A,与该店铺 A 相关联的边则为与店铺 A 相关的交易数据。需要说明的是,目标知识图谱数据可以理解为一跳子图,一跳子图的结构包括一个实体节点以及至少一个与该节点相关联的边,且一跳子图具体包含的信息有实体节点、实体类型、实体属性、边的目标节点、边类型、边属性,本说明书实施例对一跳子图中包含的信息不作具体限定,还可基于实际应用不同有所改变。

进一步地,智能知识图谱平台还可对待处理数据进行清洗、处理,进而构建目标知识图谱数据;具体的所述基于接收到的待处理数据构建目标知识图谱数据,包括:

30 对接收到的待处理数据进行处理,确定所述待处理数据的目标实体数据、以及与所述目标实体数据相关联的关系数据;

基于所述目标实体数据以及所述关系数据确定所述待处理数据的数据结构,并基于所述数据结构构建目标知识图谱数据。

35 实际应用中,在智能知识图谱平台接收到待处理数据后,可根据待处理数据构建目标知识图谱数据,确定待处理数据中的目标实体数据、以及与该目标实体数据相关联的关系数据,其中,与目标实体数据相关联的关系数据可以理解为与目标实体数据具有边的属性关系的数据;在对目标实体数据与关系数据确定待处理数据的整体图谱的数据结构,再基于该数据结构构建待处理数据的目标知识图谱数据。沿用上例,目标实体数据为店铺 A 的数据,且具有店铺 A 的实体标识 A,关系数据为与店铺 A 之间的交易数据,在根据店铺 A 的数据以及与店铺 A 之间的交易数据构建目标知识图谱数据,其中,交易数据的数量越大,构建的目标知识图谱数据中边的数量越大。

需要说明的是,根据不同的应用场景,对待处理数据处理后确定的目标实体数据不同,相应的确定与该目标实体数据相关联的关系数据也不同,本说明书实施例对此不作具体限定。

本说明书实施例提供的数据存储方法,通过对待处理数据进行处理,确定目标实体数据以及关系数

据,进而构建待处理数据的目标知识图谱数据,以便于后续将目标知识图谱数据按照图谱数据进行存储。

具体的,智能知识图谱平台可根据预先设定的拆分参数实现对目标知识图谱数据的拆分;具体的,所述拆分所述目标知识图谱数据,确定至少两个待存储目标子图数据,包括:

5 确定拆分参数,并基于所述拆分参数拆分所述目标知识图谱数据,确定至少两个待存储目标子图数据。

其中,拆分参数可以理解为智能知识图谱平台预先配置的拆分图谱数据中边的数量阈值参数,比如,拆分参数可以为100条、500条、1000条等;需要说明的是,本说明书实施例提供的数据存储方法可根据不同的应用场景确定的不同数量的目标知识图谱数据的边确定不同的拆分参数。

10 实际应用中,智能知识图谱平台可根据当前目标知识图谱数据的数据量大小,确定拆分参数,并根据该拆分参数拆分目标知识图谱数据,进而确定出至少两个待存储目标子图数据;比如,智能知识图谱平台确定的拆分参数为100,那么,按照拆分参数100拆分目标知识图谱数据的边,若目标知识图谱数据的边有450条,则拆分后每组待存储目标子图数据的边分别为100条、100条、100条、100条、50条。

15 参见图2,图2示出了本说明书一实施例提供的一种数据存储方法中目标知识图谱数据拆分为至少两个待存储目标子图数据的结构示意图。

图2中的(a)为目标知识图谱数据的结构示意图,图2中的(b)为待存储目标子图数据的结构示意图,其中,该目标知识图谱数据包括一个目标实体数据以及10条与该目标实体数据连接的边数据;对目标知识图谱数据按照预设拆分参数3进行拆分后,可获得4个待存储目标子图数据,分别为图2中的(b)的(1)、(2)、(3)、(4)所示,其中,每个待存储目标子图数据具有相同的目标实体数据,且具有拆分参数的数据边,若拆分到最后的待存储目标子图数据无法满足拆分参数的情况下,可将剩余的数据边作为最后一个待存储目标子图数据,比如图2中(b)的(4);需要说明的是,图2仅为对目标知识图谱数据进行拆分再存储的示例图,实际应用中,目标知识图谱数据的边由于应用场景的不同,与目标实体数据相关联的数据边带有的属性数据并不相同,每条边可带有目标实体数据对应的实体数据的实体标识、以及与对应的实体数据相关联的属性数据等。

25 本说明书实施例提供的数据存储方法,通过对目标知识图谱数据进行拆分后,获得至少两个待存储目标子图数据,便于后续将拆分后的待存储目标子图数据存储至数据块中。

步骤104:将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块。

其中,前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目标子图数据的初始实体标识相同。

30 边的属性可以理解为存储至少两个待存储目标子图数据的边的存储逻辑信息,与一跳子图中的边属性有所区别,该边的属性可以理解为按照边的顺序存储逻辑,比如先存储出边数据、再存储入边数据,且在出边数据中按照边类型进行分类存储,在入边数据中按照边类型分类存储等。

35 实际应用中,为了实现对大量的知识图谱数据快速地进行存储、以及为存储后的知识图谱数据提供快速的访问能力,智能知识图谱平台对于数据量较大的目标知识图谱数据,先进行拆分处理,拆分成至少两个待存储目标子图数据,其中,每个待存储目标子图数据均包含目标知识图谱数据的目标实体节点,每个待存储目标子图数据中的边均为目标知识图谱数据中至少一个目标属性的边;再将至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块。

40 进一步地,智能知识图谱平台在拆分目标知识图谱数据之后,可将拆分后的待存储目标子图数据按预设顺序存储至数据块中;具体的,所述将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块,包括:

确定所述待存储目标子图数据中边的方向,并基于所述边的方向对所述待存储目标子图数据中的边进行分类,确定至少一个边的类型,其中,所述边的方向包括出边方向和入边方向,所述出边方向为由目标实体节点指向其他实体节点的方向,所述入边方向为由其他实体节点指向目标实体节点的方向;

基于所述至少一个边的类型将所述至少两个待存储目标子图数据存储至至少两个连续数据块。

其中，边的类型可以理解为对待存储目标子图数据中边的数据属性所区分的类型，比如数据边可有数值类型、关系类型等；需要说明的是，边的类型在本说明书实施例中不作任何限定。

数据块可以理解为存储知识图谱数据的数据存储块，每个数据块具有固定的存储空间，比如，一个数据块的存储空间为 8M，本说明书实施例对数据块的存储空间不作任何限定。

实际应用中，智能知识图谱平台先确定每个待存储目标子图数据中边的方向，其中，边方向包括出边方向和入边方向，然后对待存储目标子图数据中出边方向的边进行分类，确定至少一个边的类型，同时，再对待存储目标子图数据中入边方向的边进行分类，确定至少一个边的类型，再根据每个边的类型将拆分后的所有待存储目标子图数据存储至至少两个连续数据块。沿用图 2 的示例，标号为 (1) 的待存储目标子图数据中出边方向的数据边有 3 条，入边方向有 0 条，标号为 (2) 的待存储目标子图数据中出边方向的数据边有 2 条，入边方向有 1 条，标号为 (3) 的待存储目标子图数据中出边方向的数据边有 1 条，入边方向有 2 条，标号为 (4) 的待存储目标子图数据中出边方向的数据边有 0 条，入边方向有 1 条，那么，每个待存储目标子图数据都按照出边方向顺序存储，再按照入边方向顺序存储，需要说明的是，本说明书实施例提供的存储顺序为先存储出边方向的数据，再存储入边方向的数据，但不对此做具体限定。

参见 3，图 3 示出了本说明书实施例提供的一种数据存储方法中将待存储目标子图数据存储至至少两个连续数据块的结构示意图。

图 3 中，数据块可分别为数据块 1、数据块 2...数据块 n，其中，数据块的数量在本实施例中不作具体限定，图 3 中标号 (1) 至 (4) 的数据为待存储目标子图数据，实际应用中，智能知识图谱平台将拆分后的 4 个待存储目标子图数据存储至至少两个连续数据块中，由图 3 中可以看出，将待存储目标子图数据 (1) 存储至数据块 1，待存储目标子图数据 (2) 存储至数据块 1，待存储目标子图数据 (3) 存储至数据块 2，待存储目标子图数据 (4) 存储至数据块 2，需要说明的是，在将待存储目标子图数据进行存储时，要依次存储至连续的数据块中，在数据块 1 的剩余存储空间无法存储下一个待存储目标子图数据的情况下，再将下一个待存储目标子图数据存储至数据块 2，同时，每个待存储目标子图数据要按照边的方向以及每个方向边的属性分类后，依次进行顺序存储至数据块中。

此外，图 3 中所示，在将至少两个待存储目标子图数据写入同一个数据块中的过程（待存储目标子图数据 (1) 和待存储目标子图数据 (2) 写入数据块 1 的过程）中，需要将具有相同目标实体数据的至少两个待存储目标子图数据进行合并处理，即可以理解为将每个待存储目标子图数据中相同的目标实体数据合并为一个目标实体数据写入至数据块中，进而保证了单个数据块中的每个实体数据标识有一个，避免了数据写入的重复性，增强数据存储效率，节省数据块的存储空间。

本说明书实施例提供的数据存储方法，通过确定待存储目标子图数据中边的方向以及边的类型，对待存储目标子图数据按顺序存储至数据块中，进而保证了对数据量大的目标知识图谱数据存储的有序性。

此外，本说明书实施例提供的数据存储方法中的数据块还包括缓冲区，用于存储待存储目标子图数据中的剩余数据；具体的，所述数据块还包括缓冲区，

相应地，所述将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块，包括：在所述待存储目标子图数据中确定无法存储至至少两个连续数据块的剩余数据，将所述剩余数据存储至所述至少两个连续数据块的末尾数据块的缓冲区内。

其中，剩余数据可以理解为待存储目标子图数据中无法全部存储至同一数据块中所剩余的数据，比如，待存储目标子图数据的数据量为 5M，存储至数据块中的待存储目标子图数据已经存储了 4M，还差 1M 数据未存储，那么剩余的 1M 数据则可以理解为剩余数据。

实际应用中，同一数据块中可以存储多个待存储目标子图数据，在多个待存储目标子图数据按照存储顺序写入至数据块的时候，若写入该数据块中的最后一个待存储目标子图数据无法全部存储在该数据块中，还剩余一部分剩余数据的情况下，可将该剩余数据存储至该数据块的缓冲区内，以保证最后一个

待存储目标子图数据能够存储在同一数据块中；此外，在至少两个待存储目标子图数据可占用至少两个连续数据块进行存储时，每个数据块中均可设置缓冲区，以解决待存储目标子图数据中的大部分数据已经存储至当前数据块中，而还有一小部分剩余数据无法在该数据块中存储的问题，通过设置缓冲区以解决剩余数据能够与待存储目标子图数据中的大部分数据存在同一数据块的问题。

5 参见图4，图4示出了本说明书实施例提供的一种数据存储方法中将待存储目标子图数据存储至至少两个连续数据块中缓冲区的结构示意图。

需要说明的是，图4中以每个数据块中写入的最后一个待存储目标子图数据无法全部存储至当前数据块时，产生剩余数据的处理方式；本说明书实施例中对于同一个目标知识图谱数据拆分后的至少两个待存储目标子图数据可连续存储至数据块中，该缓冲区的存储应用可以在每个数据块中应用也可以在所有的待存储目标子图数据按顺序存储的末尾应用，本说明书实施例对此不作任何限定。

10 图4中，以四个待存储目标子图数据（1）、（2）、（3）、（4）为例，存储至连续的两个数据块为例进行说明，其中，连续的两个数据块可参见图4中的数据块1和数据块2；在实际应用中，在待存储目标子图数据（1）存储至数据块1后，数据块1的剩余空间则可存储待存储目标子图数据（2），在待存储目标子图数据（2）中还有一小部分剩余数据无法存储至当前数据块中的剩余空间的情况下，可
15 占用数据块1中的缓冲区，将这一小部分剩余数据存储至数据块1的缓冲区内，需要说明的是，所确定的剩余数据的大小一定是小于缓冲区预留的存储空间，也可以理解为，若剩余数据量的大小能够全部存储至该数据块的缓冲区内，才可启动缓冲区进行数据存储，若剩余数据量的大小仅能在缓冲区中存储一部分，则不启动缓冲区进行存储，可直接将待存储目标子图数据存储至下一个数据块2即可。

20 相应的，图4中的待存储目标子图数据（4）与待存储目标子图数据（2）中的剩余数据的存储机制相一致，在此则不作过多赘述。

需要说明的是，图4中同一数据块中的待存储目标子图数据进行合并处理的过程可参考图3中的合并处理的描述，在此则不作过多赘述。

25 本说明书实施例提供的数据存储方法，通过在数据块中设置缓冲区，以满足待存储目标子图数据中剩余数据小于缓冲区预设数据量大小的情况下，保证同一个待存储目标子图数据可以存储于同一数据块中，实现数据存储的完整性。

此外，在将待存储目标子图数据存储至两个连续的数据块之后，还可在每个数据块中记录索引数据，以便于后续可根据索引数据的记录实现知识图谱数据快速访问的能力；具体的，所述将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块之后，还包括：

30 记录每个数据块的索引信息，并基于所述每个数据块中的索引信息，确定索引数组信息，其中，所述索引信息包括所述每个数据块中存储的待存储目标子图数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件。

其中，索引数组信息可以理解为将每个数据块中记录的索引信息相链接后的索引组的信息，比如，将每个数据块中的索引信息抽象为一个链接块的索引信息，抽象后的索引数据可理解为索引数组信息。

35 实际应用中，智能知识图谱平台对于每个数据块中的存储数据记录相应的索引信息，其中，索引信息包括当前数据块中存储第一个待存储目标子图数据的初始实体标识、存储在当前数据块末尾的最后一个待存储目标子图数据的末尾实体标识、以及与末尾实体标识相关联的边属性条件，其中，与末尾实体标识相关联的边属性条件可以理解为存储在当前数据块中最后一个待存储目标子图数据中的边的方向以及边的类型，比如，数据块1记录的索引信息为vid=1、vid=3、出边方向、交易数据类型，其中，vid表示待存储目标子图数据的实体标识。

40 需要说明的是，索引信息可以理解为meta数据，meta是一个类似csv的存储文件，一行数据描述一个数据块，共有7个字段：vertexType（实体数据类型）、startVertexId（初始实体数据标识）、endVertexId（末尾实体数据标识）、filename（文件名称）、blockIndex（数据块索引信息）、edgeMode（末尾实体数据中的最后一条边的方向）、edgeType（末尾实体数据中的最后一条边的类型）。且上述的索引信息

仅为存储示意,表明 meta 数据中记录的信息以及 meta 数据间的排列情况,但为了减少剩余存储,在存储文件时,还可进行一些存储优化,实际的落盘数据可能与上述格式有所区别,在本说明书实施例对此则不作具体限定。

5 本说明书实施例提供的数据存储方法,通过记录每个数据块中的索引信息,进而将每个数据块中的索引信息进行抽象处理,确定索引数组信息,便于后续通过该索引数组信息快速地查找访问的待存储目标子图数据。

进一步地,根据每个数据块中记录的索引信息进行格式转换构建索引数组信息;具体的,所述基于所述每个数据块中的索引信息,确定索引数组信息,包括:

10 确定每个数据块中待存储目标子图数据的初始实体标识、末尾实体标识;

在确定两个连续数据块中相邻的末尾实体标识与初始实体标识相同的情况下,对所述两个连续数据块中的索引信息进行处理,确定索引数组信息,其中,所述索引数组信息中的实体标识基于存储顺序排列。

实际应用中,知识图谱数据平台可先确定每个数据块存储所有待存储目标子图数据的初始实体标识和末尾实体标识,其中,初始实体标识可以理解为存储在当前数据块中的第一个待存储目标子图数据的目标实体数据的标识,末尾实体标识可以理解为存储在当前数据块中最后一个待存储目标子图数据的末尾实体数据的标识;然后在每个数据块中均确定了初始实体标识和末尾实体标识之后,需要判断连续的两个数据块的末尾实体标识与初始实体标识是否相同,若在相同的情况下,则说明当前两个连续数据块中存储的待存储目标子图数据的目标实体数据一致,即可以理解为当前两个连续数据块中所有存储的待存储目标子图数据都是同一个目标知识图谱数据拆分来的,本说明书实施例提供的数据存储方法,是将
20 同一个目标知识图谱数据所存储的数据块的索引数据抽象合并为一个索引数组数据,且抽象后的一个索引数组数据与数据块的存储机制相同,均为有序存储。

参见图 5,图 5 示出了本说明书实施例提供的一种数据存储方法中索引数组信息的生成方式示意图。

图 5 中上半部分为 4 个数据块存储待存储目标子图数据,其中,在数据块 1 (8M) 中,存储的是目标实体数据标识为 1 的待存储目标子图数据 (vid=1)、目标实体数据标识为 2 的待存储目标子图数据
25 (vid=2); 数据块 2 (8M) 中,存储的是目标实体数据标识为 2 的待存储目标子图数据 (vid=2),其中,数据块 2 中的虚线分界线表示具有两个 vid=2 的待存储目标子图数据;数据块 3 (8M) 中,存储的是目标实体数据标识为 2 的待存储目标子图数据 (vid=2)、目标实体数据标识为 3 的待存储目标子图数据 (vid=3)、目标实体数据标识为 4 的待存储目标子图数据 (vid=4); 数据块 4 (8M) 中,存储的是目标实体数据标识为 4 的待存储目标子图数据 (vid=4); 下半部分为抽象出的索引数组信息 1 以及
30 索引数组信息 2 的示意图,其中,索引数组信息 1 中包括数据块 1、数据块 2 以及数据块 3,索引数组信息 2 包括数据块 3 和数据块 4;在索引数组信息 1 中,每个数据块中均存有目标实体数据标识为 2 的待存储目标子图数据,在实际应用中,对索引数组信息进行抽象链接时,可确定每个数据块中的第一个待存储目标子图数据的目标实体标识,以及存储在该数据块的最后一个待存储目标子图数据的目标实体标识,通过判断第一个待存储目标子图数据的目标实体标识与最后一个待存储目标子图数据的目标实体标识是否相同,进而实现对数据块的链接,需要说明的是,索引数组信息中数据块的链接不是物理层面上对存储介质的链接,而是抽象为在逻辑层面上,对每个数据块中的索引信息进行链接的过程,便于后续基于链接后的索引数组信息快速地实现数据访问的能力,也实现了在索引数组信息 1 中目标实体标识
35 相同的数据存储的数据块索引信息在同一个索引数组信息中存储。

进一步地,在索引数组信息 2 中包括数据块 3 和数据块 4,原因是数据块 3 的末尾实体标识为 vid=4,数据块 4 的初始实体标识为 vid=4,因此,将数据块 3 与数据块 4 的索引记录信息链接到索引数组信息
40 2 中。

需要说明的是,图 5 中数据块的斜线部分为表示两个索引数组信息中重复存储的索引数据,且在后续数据访问的过程中,在索引数组信息中可忽略访问,避免重复访问带来的计算资源的浪费。

本说明书实施例提供的数据存储方法，通过对拆分后的待存储目标子图数据从逻辑上聚合为一个完整的子图，将每个数据块中记录的索引信息进行处理，生成索引数组信息，便于后续可快速地根据索引数组信息查找相应的待存储目标子图数据。

5 此外，在目标知识图谱数据被拆分再进行存储之后，还可对拆分后的待存储目标子图数据进行读取，即可利用记录的索引数组信息对待存储目标子图数据进行查询，参见图6，图6示出了本说明书实施例提供了一种数据读取方法的流程图，具体包括以下步骤：

步骤602：接收针对目标数据的数据读取请求，基于所述数据读取请求在所述索引数组信息中确定目标存储位置；

10 其中，所述索引数组信息基于每个数据块中的索引信息而确定，并包括每个数据块中存储的图谱数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件；

步骤604：基于所述目标存储位置从目标数据块中读取所述目标数据。

其中，目标存储位置可以理解为目标数据存储和数据块中的逻辑存储位置，即可在索引数组信息中根据逻辑存储位置再确定物理存储地址，需要说明的是，本说明书实施例中对于目标存储位置的类型不作限定。

15 具体实施时，智能知识图谱平台在接收到用户发送的针对目标数据的数据读取请求的情况下，可根据数据读取请求在索引数组信息中找到目标数据对应的目标存储位置之后，再根据这个目标存储位置从相应的目标数据块中读取对应的数据。

本说明书实施例提供的数据读取方法，通过在索引数组信息中确定目标存储位置，以实现在索引数据层面上快速地查找到具体存储目标数据的存储位置。

20 进一步地，在对待存储目标子图数据进行随机读取的过程中，还可基于聚合后的数据块中的初始实体标识和末尾实体标识，按照目标实体数据标识进行查找并读取；具体的，所述接收针对目标数据的数据读取请求，基于所述数据读取请求在所述索引数组信息中确定目标存储位置，包括：

接收针对目标数据的数据读取请求，其中，所述数据读取请求中携带有目标数据的实体标识、数据边类型；

25 基于所述目标数据的实体标识、数据边类型在所述索引数组信息中进行查找，确定所述目标数据对应的目标存储位置。

实际应用中，在智能知识图谱平台接收到用户发送的针对目标数据的数据读取请求之后，可根据数据读取请求中携带的目标数据的实体标识、数据边类型，从索引数组信息中可采用二分法进行查找，来判定目标数据的实体标识所存储的存储地址，进而确定目标数据存储的目标存储位置。比如，用户对于30 店铺A在时间段B的商品交易额，则该数据读取请求为商品交易额读取请求，其中，该商品交易额读取请求携带有实体标识-店铺A，数据边类型为入边方向数据，然后，可在至少两个索引数组信息中通过二分法查找实体标识为店铺A所在的索引数组信息，进而确定一个目标索引数组信息，在该目标索引数组信息中继续查找店铺A在时间段B的商品交易额数据。

35 需要说明的是，由于上述实施例中描述的在数据块中对待存储目标子图数据进行存储的过程中，是按照目标实体标识的顺序进行存储，比如vid=1、vid=2、vid=3...，由于每个目标实体数据均有相应的实体标号，所以在后续数据访问的过程中，可采用二分法确定对应的目标存储位置。

本说明书实施例提供的数据读取方法，通过在索引数组信息中查找与目标数据中的实体标识以及数据边类型，进行确定目标存储位置，以减少遍历查找所有存储介质的查找流程，实现快速查找目标数据的目标存储位置。

40 更进一步地，所述基于所述目标存储位置从目标数据块中读取所述目标数据，包括：

基于所述目标存储位置在所述至少两个连续数据块中确定目标数据块；

基于所述数据读取请求中携带的边属性条件从所述目标数据块中读取所述目标数据。

实际应用中，智能知识图谱平台在索引数组信息中确定出目标数据对应的目标存储位置之后，可基

于该目标存储位置在至少两个连续数据块中确定存储目标数据的目标数据块，然后再根据数据读取请求中携带的边的属性条件，从目标数据块中读取目标数据，其中，边的属性条件可以理解为待存储目标子图数据中数据边的属性类型，比如数值类型等；具体的，可在确定的目标数据块中读取数值类型的数据作为目标数据。

5 参见图 7，图 7 示出了本说明书实施例提供的一种数据读取方法随机读取数据的过程示意图。

图 7 中具体的示意结构与图 5 中示意结构相同，在此则不作过多赘述，但图 7 中的索引数组信息 1 可看作为逻辑数据块 1，索引数组信息 2 可看作为逻辑数据块 2，在数据读取的过程中，在确定待读取的待存储目标子图数据的目标实体标识为 vid=2 的情况下，则可确定在索引数组信息 1 中进行数据查找，由于索引数据信息 1 中还存储了 vid=1、vid=3、vid=4 的索引信息，则在读取层可忽略掉除目标实体标识为 vid=2 的其他数据的索引信息（图 6 中的索引数组信息 1 斜线覆盖的问题），其中，vid=1 的数据在数据块 1 中存储，并未跨不同的数据块存储，可表示 vid=1 的数据的数据量不大，无需对其进行拆分存储（可以理解为非热点数据，但当前服务器对是否为热点数据并不感知），在索引数组信息 1 中需要访问 vid=1 的数据时，再进行读取；同样地，对于 vid=4 的数据也可参考上述读取 vid=2 数据的方式，在此不作过多赘述。此外，在同一数据块中具有数据量较小的知识图谱数据（非热点数据）和数据量较大的拆分后的子图数据时，如索引数组信息 2 中存在 vid=3 和 vid=4 的数据，即可在读取时 vid=4 的过程中需要跳过一个数据块中遗留的数据（即 vid=3 的数据）。

需要说明的是，本说明书实施例提供的数据读取方法中的数据读取过程，是基于上述实施例中描述的数据存储过程的基础上实现的，即在前述数据存储的过程就会为后续的数据查找提供了快速地数据访问能力。

20 本说明书实施例提供的数据读取方法，通过目标存储位置确定具体存储目标数据的目标数据块，以实现通过索引数据快速地确定具体的目标数据存储位置，无需遍历查找更多的数据块，造成不必要计算资源开销。

此外，本说明书实施例提供的数据读取方法还可提供数据分区读取能力，该应用场景可为分布式数据读取场景；具体的，所述数据读取请求在索引数组信息中确定目标存储位置之后，还包括：

25 基于负载均衡条件对所述索引数组信息进行分区处理，确定至少两个数据分区；

相应地，所述确定至少两个数据分区之后，还包括：

接收用户发送的针对目标数据的数据分区读取请求，基于所述数据分区读取请求中携带的分区索引信息确定目标数据分区；

基于所述数据分区读取请求中携带的实体标识从所述目标数据分区中读取所述目标数据。

30 实际应用中，为了实现分布式场景下快速进行数据读取的过程，通常是分布式多任务的，读取后的图数据可进行分布式的图计算和规则推理。对聚合后的索引数组信息进行分区处理，每个分区由一个 worker 实例负责数据加载（热点子图的边截断也可在 worker 进行数据加载时实现），进而，在分区的过程中，将考虑各个分布式服务器的负载均衡条件，避免一部分服务器的处理量较大，一部分服务器的处理量较小，导致的负载均衡的问题出现；在将索引数组信息分区为至少两个数据分区之后，在接收到用户发送的针对目标数据的数据分区读取请求之后，还可基于数据分区读取请求中携带的分区索引信息进而确定目标数据分区，再根据目标数据分区读取与数据分区读取请求中携带的实体标识相应的目标数据。

需要说明的是，本说明书实施例还可提供分区计算公式，以计算某个分区的数据块索引范围 [int startIndex, int endIndex]，即 $\text{int startIndex} = ((\text{workerIndex} * \text{blockCount} + \text{workerCount} - 1) / \text{workerCount})$ ；
40 $\text{int endIndex} = ((\text{workerIndex} + 1) * \text{blockCount} - 1) / \text{workerCount}$ 。

本说明书实施例提供的数据读取方法，在分布式数据读取场景下，即可采用对索引数组信息进行分区处理，以满足不同实例处理过程的负载均衡，进而快速地实现数据访问能力。

参见图 8，图 8 示出了本说明书实施例提供的一种数据读取方法分区读取数据的过程示意图。

图 8 中的具体的示意结构与图 5 中示意结构相同,在此则不作过多赘述,但图 8 中的索引数组信息 1 可分区为分区 1,索引数组信息 2 可分区为分区 2,则每个分区可对应一个分布式服务器,以满足负载均衡的条件划分至少两个索引数组信息,具体的,可根据每个索引数组信息中存储的数据块索引信息的数量确定的,还可采用其他方式实现,在本说明书实施例中则不做具体限定。在数据读取的过程中,可先确定目标数据的存储分区,再根据分区确定目标数据块,进行读取目标数据,在此对具体的过程不展开描述。

此外,本说明书实施例提供的数据读取方法,以一跳子图为基础,将一跳子图有序存储在数据块中,当一跳子图较大时会在写入侧拆分为多个小的子图,分多次写入,存储层不感知热点的存在。数据块通常拥有近似固定的大小,由写入操作触发 block 的切分。一般情况下,只存储在单个数据块中的一跳子图,可以理解为非热点数据,数据量较大的数据在单个数据块中存在存储不下的问题,为了避免数据量大的数据分散到多个文件夹中存储,导致的某个单个文件数据过大,且不好实现数据访问,则可将数据量大的一跳子图进行拆分存储。

需要说明的是,数据量大的一跳子图拆分约束为:1、数据量大的一跳子图的边按出边和入边类型进行先后存储;2、数据量大的一跳子图中相同方向类型的边(出边或入边),按边的类型进行字典序排序后存储;3、数据量大的一跳子图按数据块的大小对拆分后的子图进行点切分,将切分后的子图有序存储在 N 个数据块中,且切分后的单个子图内部的边也是有序存储的。

此外,对于热点子图的构建方法,热点子图由于数据量较大,可跨越至少两个数据块存储,可将该类数据块成为热点数据块,热点数据块的数目与热点子图的规模成线性关系。在对拆分后的子图进行存储的过程中,子图需要按出边、入边、边类型顺序切分,子图拆分后可从左到右进行顺序写入,其中,子图的拆分过程在构建热点数据的链路中执行的,SDK 并未提供拆分机制。存储热点子图的数据块也是具有连续性的,依赖于实体标识的全局有序性来保证,其中,全局有序性可分为单个数据块中 vid 有序,单文件内数据块之间的 vid 有序,文件间的 vid 有序,相同 vid 的热点数据块也是连续的,进而才能保证后续可有序快速提供数据访问能力。

综上,本说明书实施例提供的数据读取方法,在存储时要弱化热点的存在,访问时根据索引信息判定热点,进而在随机读或批量读时决定是否跳过热点等操作,再对一跳子图进行拆分和重组,实现数据量大的一跳子图可存储到多个数据块的存储,并保证子图的有序存储特性。数据块的有序存储,并将热点信息记录到索引信息中,可通过二分查找从热点数据中快速定位所需数据。且在数据访问时,将 N 个热点数据块抽象为 chained block(索引数组信息),将热点和非热点数据存储模型进行统一,保证随机读、数据块分区等机制的有效执行。

与上述方法实施例相对应,本说明书还提供了数据存储装置实施例,图 9 示出了本说明书一个实施例提供的一种数据存储装置的结构示意图。如图 9 所示,该装置包括:

图谱拆分模块 902,被配置为拆分目标知识图谱数据,确定至少两个待存储目标子图数据,其中,所述目标知识图谱数据包括目标实体节点以及至少一个与所述目标实体节点相关联的边,各个所述待存储目标子图数据包括所述目标实体节点以及至少一个目标属性的边;

数据存储模块 904,被配置为将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块,其中,前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目标子图数据的初始实体标识相同。

可选地,所述装置,还包括:

索引数据信息确定模块,被配置为记录每个数据块的索引信息,并基于所述每个数据块中的索引信息,确定索引数组信息,其中,所述索引信息包括所述每个数据块中存储的待存储目标子图数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件。

可选地,所述索引数据信息确定模块,进一步被配置为:

确定每个数据块中待存储目标子图数据的初始实体标识、末尾实体标识;

在确定两个连续数据块中相邻的末尾实体标识与初始实体标识相同的情况下，对所述两个连续数据块中的索引信息进行处理，确定索引数组信息，其中，所述索引数组信息中的实体标识基于存储顺序排列。

可选地，所述图谱拆分模块 902，进一步配置为：

- 5 确定拆分参数，并基于所述拆分参数拆分所述目标知识图谱数据，确定至少两个待存储目标子图数据；

可选地，所述装置，还包括：

图谱构建模块，被配置为对接收到的待处理数据进行处理，确定所述待处理数据的目标实体数据、以及与所述目标实体数据相关联的关系数据；

- 10 基于所述目标实体数据以及所述关系数据确定所述待处理数据的数据结构，并基于所述数据结构构建目标知识图谱数据。

可选地，所述数据存储模块 904，进一步被配置为：

确定所述待存储目标子图数据中边的方向，并基于所述边的方向对所述待存储目标子图数据中的边进行分类，确定至少一个边的类型，其中，所述边的方向包括出边方向和入边方向，所述出边方向为由目标实体节点指向其他实体节点的方向，所述入边方向为由其他实体节点指向目标实体节点的方向；

- 15 基于所述至少一个边的类型将所述至少两个待存储目标子图数据存储至至少两个连续数据块。

可选地，所述数据块还包括缓冲区，

可选地，所述数据存储模块 904，进一步被配置为：

- 20 在所述待存储目标子图数据中确定无法存储至至少两个连续数据块的剩余数据，将所述剩余数据存储至所述至少两个连续数据块的末尾数据块的缓冲区。

本说明书实施例提供的数据存储装置，通过对构建的目标知识图谱进行拆分，确定至少两个待存储目标子图数据，并对至少两个待存储目标子图数据按边的属性存储至至少两个连续数据块，以解决数据量大的知识图谱一次性存储效率低的问题，通过拆分再进行存储也提高了图谱数据的存储性能，也便于后续对数据量大的图谱数据提供快速访问的能力。

- 25 上述为本实施例的一种数据存储装置的示意性方案。需要说明的是，该数据存储装置的技术方案与上述的数据存储方法的技术方案属于同一构思，数据存储装置的技术方案未详细描述的细节内容，均可以参见上述数据存储方法的技术方案的描述。

与上述方法实施例相对应，本说明书还提供了数据读取装置实施例，图 10 示出了本说明书一个实施例提供的一种数据读取装置的结构示意图。如图 10 所示，该装置应用于知识图谱平台，包括：

- 30 存储位置确定模块 1002，被配置为接收针对目标数据的数据读取请求，基于所述数据读取请求在索引数组信息中确定目标存储位置；

其中，所述索引数组信息基于每个数据块中的索引信息而确定，并包括每个数据块中存储的图谱数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件；

数据读取模块 1004，被配置为基于所述目标存储位置从目标数据块中读取所述目标数据。

- 35 可选地，所述存储位置确定模块 1002，进一步被配置为：

接收针对目标数据的数据读取请求，其中，所述数据读取请求中携带有目标数据的实体标识、数据边类型；

基于所述目标数据的实体标识、数据边类型在所述索引数组信息中进行查找，确定所述目标数据对应的目标存储位置。

- 40 可选地，所述数据读取模块 1004，进一步配置为：

基于所述目标存储位置在至少两个连续数据块中确定目标数据块；

基于所述数据读取请求中携带的边属性条件从所述目标数据块中读取所述目标数据。

可选地，所述装置，还包括：

分区确定模块，被配置为基于负载均衡条件对所述索引数组信息进行分区处理，确定至少两个数据分区；

可选地，所述数据读取模块 1004，进一步配置为：

接收针对目标数据的数据分区读取请求，基于所述数据分区读取请求中携带的分区索引信息确定目标数据分区；

基于所述数据分区读取请求中携带的实体标识从所述目标数据分区中读取所述目标数据。

本说明书实施例提供的数据读取装置，在存储时要弱化热点的存在，访问时根据索引信息判定热点，进而在随机读或批量读时决定是否跳过热点等操作，再对一跳子图进行拆分和重组，实现数据量大的一跳子图可存储到多个数据块的存储，并保证子图的有序存储特性。数据块的有序存储，并将热点信息记录到索引信息中，可通过二分查找从热点数据中快速定位所需数据。且在数据访问时，将 N 个热点数据块抽象为 chained block（索引数组信息），将热点和非热点数据存储模型进行统一，保证随机读、数据块分区等机制的有效执行。

上述为本实施例的一种数据读取装置的示意性方案。需要说明的是，该数据读取装置的技术方案与上述的数据读取方法的技术方案属于同一构思，数据读取装置的技术方案未详细描述的细节内容，均可以参见上述数据读取方法的技术方案的描述。

图 11 示出了根据本说明书一个实施例提供的一种计算设备 1100 的结构框图。该计算设备 1100 的部件包括但不限于存储器 1110 和处理器 1120。处理器 1120 与存储器 1110 通过总线 1130 相连接，数据库 1150 用于保存数据。

计算设备 1100 还包括接入设备 1140，接入设备 1140 使得计算设备 1100 能够经由一个或多个网络 1160 通信。这些网络的示例包括公用交换电话网（PSTN）、局域网（LAN）、广域网（WAN）、个域网（PAN）或诸如因特网的通信网络的组合。接入设备 1140 可以包括有线或无线的任何类型的网络接口（例如，网络接口卡（NIC））中的一个或多个，诸如 IEEE802.11 无线局域网（WLAN）无线接口、全球微波互联接入（Wi-MAX）接口、以太网接口、通用串行总线（USB）接口、蜂窝网络接口、蓝牙接口、近场通信（NFC）接口，等等。

在本说明书的一个实施例中，计算设备 1100 的上述部件以及图 11 中未示出的其他部件也可以彼此相连接，例如通过总线。应当理解，图 11 所示的计算设备结构框图仅仅是出于示例的目的，而不是对本说明书范围的限制。本领域技术人员可以根据需要，增添或替换其他部件。

计算设备 1100 可以是任何类型的静止或移动计算设备，包括移动计算机或移动计算设备（例如，平板计算机、个人数字助理、膝上型计算机、笔记本计算机、上网本等）、移动电话（例如，智能手机）、可佩戴的计算设备（例如，智能手表、智能眼镜等）或其他类型的移动设备，或者诸如台式计算机或 PC 的静止计算设备。计算设备 1100 还可以是移动式或静止式的服务器。

其中，处理器 1120 用于执行如下计算机可执行指令，该计算机可执行指令被处理器执行时实现上述方法的步骤。

上述为本实施例的一种计算设备的示意性方案。需要说明的是，该计算设备的技术方案与上述的方法的技术方案属于同一构思，计算设备的技术方案未详细描述的细节内容，均可以参见上述方法的技术方案的描述。

本说明书一实施例还提供一种计算机可读存储介质，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现上述方法的步骤。

上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是，该存储介质的技术方案与上述的方法的技术方案属于同一构思，存储介质的技术方案未详细描述的细节内容，均可以参见上述方法的技术方案的描述。

本说明书一实施例还提供一种计算机程序，其中，当所述计算机程序在计算机中执行时，令计算机执行上述方法的步骤。

上述为本实施例的一种计算机程序的示意性方案。需要说明的是，该计算机程序的技术方案与上述的方法的技术方案属于同一构思，计算机程序的技术方案未详细描述的细节内容，均可以参见上述方法的技术方案的描述。

5 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。

10 所述计算机指令包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、电载波信号、电信号以及软件分发介质等。需要说明的是，所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减，例如在某些司法管辖区，根据立法和专利实践，计算机可读介质不包括电载波信号和电信号。

15 需要说明的是，对于前述的各方法实施例，为了简便描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本说明书实施例并不受所描述的动作顺序的限制，因为依据本说明书实施例，某些步骤可以采用其它顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本说明书实施例所必须的。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中未详述的部分，可以参见其它实施例的相关描述。

20 以上公开的本说明书优选实施例只是用于帮助阐述本说明书。可选实施例并没有详尽叙述所有的细节，也不限制该发明仅为所述的具体实施方式。显然，根据本说明书实施例的内容，可作很多的修改和变化。本说明书选取并具体描述这些实施例，是为了更好地解释本说明书实施例的原理和实际应用，从而使所属技术领域技术人员能很好地理解和利用本说明书。本说明书仅受权利要求书及其全部范围和等效物的限制。

25

权 利 要 求

- 1、一种数据存储方法，应用于知识图谱平台，包括：
- 拆分目标知识图谱数据，确定至少两个待存储目标子图数据；
- 其中，所述目标知识图谱数据包括目标实体节点以及至少一个与所述目标实体节点相关联的边，各
- 5 个所述待存储目标子图数据包括所述目标实体节点以及至少一个目标属性的边；
- 将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块，其中，前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目标子图数据的初始实体标识相同。
- 2、根据权利要求 1 所述的数据存储方法，所述将所述至少两个待存储目标子图数据按照边的属性
- 10 存储至至少两个连续数据块之后，还包括：
- 记录每个数据块的索引信息，并基于所述每个数据块中的索引信息，确定索引数组信息，其中，所述索引信息包括所述每个数据块中存储的待存储目标子图数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件。
- 3、根据权利要求 2 所述的数据存储方法，所述基于所述每个数据块中的索引信息，确定索引数组
- 15 信息，包括：
- 确定每个数据块中待存储目标子图数据的初始实体标识、末尾实体标识；
- 在确定两个连续数据块中相邻的末尾实体标识与初始实体标识相同的情况下，对所述两个连续数据块中的索引信息进行处理，确定索引数组信息，其中，所述索引数组信息中的实体标识基于存储顺序排列。
- 4、根据权利要求 1 所述的数据存储方法，所述拆分所述目标知识图谱数据，确定至少两个待存储
- 20 目标子图数据，包括：
- 确定拆分参数，并基于所述拆分参数拆分所述目标知识图谱数据，确定至少两个待存储目标子图数据；
- 相应地，所述拆分目标知识图谱数据之前，还包括：
- 25 对接收到的待处理数据进行处理，确定所述待处理数据的目标实体数据、以及与所述目标实体数据相关联的关系数据；
- 基于所述目标实体数据以及所述关系数据确定所述待处理数据的数据结构，并基于所述数据结构构建目标知识图谱数据。
- 5、根据权利要求 4 所述的数据存储方法，所述将所述至少两个待存储目标子图数据按照边的属性
- 30 存储至至少两个连续数据块，包括：
- 确定所述待存储目标子图数据中边的方向，并基于所述边的方向对所述待存储目标子图数据中的边进行分类，确定至少一个边的类型，其中，所述边的方向包括出边方向和入边方向，所述出边方向为由目标实体节点指向其他实体节点的方向，所述入边方向为由其他实体节点指向目标实体节点的方向；
- 基于所述至少一个边的类型将所述至少两个待存储目标子图数据存储至至少两个连续数据块。
- 6、根据权利要求 1 所述的数据存储方法，所述数据块还包括缓冲区，
- 35

— 16 —

相应地,所述将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数据块,包括:

在所述待存储目标子图数据中确定无法存储至至少两个连续数据块的剩余数据,将所述剩余数据存储至所述至少两个连续数据块的末尾数据块的缓冲区。

7、一种数据读取方法,应用于知识图谱平台,包括:

5 接收针对目标数据的数据读取请求,基于所述数据读取请求在索引数组信息中确定目标存储位置;

其中,所述索引数组信息基于每个数据块中的索引信息而确定,并包括每个数据块中存储的图谱数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件;

基于所述目标存储位置从目标数据块中读取所述目标数据。

8、根据权利要求 7 所述的数据读取方法,所述接收针对目标数据的数据读取请求,基于所述数据
10 读取请求在所述索引数组信息中确定目标存储位置,包括:

接收针对目标数据的数据读取请求,其中,所述数据读取请求中携带有目标数据的实体标识、数据边类型;

基于所述目标数据的实体标识、数据边类型在所述索引数组信息中进行查找,确定所述目标数据对应的目标存储位置。

9、根据权利要求 7 所述的数据读取方法,所述基于所述目标存储位置从目标数据块中读取所述目
15 标数据,包括:

基于所述目标存储位置在至少两个连续数据块中确定目标数据块;

基于所述数据读取请求中携带的边属性条件从所述目标数据块中读取所述目标数据。

10、根据权利要求 7 所述的数据读取方法,所述数据读取请求在索引数组信息中确定目标存储位置
20 之后,还包括:

基于负载均衡条件对所述索引数组信息进行分区处理,确定至少两个数据分区;

相应地,所述确定至少两个数据分区之后,还包括:

接收针对目标数据的数据分区读取请求,基于所述数据分区读取请求中携带的分区索引信息确定目
标数据分区;

25 基于所述数据分区读取请求中携带的实体标识从所述目标数据分区中读取所述目标数据。

11、一种数据存储装置,应用于知识图谱平台,包括:

图谱拆分模块,被配置为拆分目标知识图谱数据,确定至少两个待存储目标子图数据,其中,所述
目标知识图谱数据包括目标实体节点以及至少一个与所述目标实体节点相关联的边,各个所述待存储目
标子图数据包括所述目标实体节点以及至少一个目标属性的边;

30 数据存储模块,被配置为将所述至少两个待存储目标子图数据按照边的属性存储至至少两个连续数
据块,其中,前一数据块中存储的待存储目标子图数据的末尾实体标识与后一数据块中存储的待存储目
标子图数据的初始实体标识相同。

12、一种数据读取装置,应用于知识图谱平台,包括:

存储位置确定模块,被配置为接收针对目标数据的数据读取请求,基于所述数据读取请求在索引数

组信息中确定目标存储位置；

其中，所述索引数组信息基于每个数据块中的索引信息而确定，并包括每个数据块中存储的图谱数据的初始实体标识、末尾实体标识、与末尾实体标识相关联的边属性条件；

数据读取模块，被配置为基于所述目标存储位置从目标数据块中读取所述目标数据。

5

13、一种计算设备，包括：

存储器和处理器；

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1-6、7-10 任意一项所述方法的步骤。

10

14、一种计算机可读存储介质，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1-6、7-10 任意一项所述方法的步骤。