图数据库系统及其查询方法

本申请要求于2024年03月07日提交的申请号为202410263169.X、发明名称为“一种图数据查询方法、装置及计算设备集群”的中国专利申请的优先权，以及于2024年04月29日提交的申请号为202410532715.5、发明名称为“图数据库系统及其查询方法”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

本申请涉及数据库技术领域，特别涉及一种图数据库系统及其查询方法。

背景技术

随着大数据时代的到来，物与物、人与人、知识与知识的关联越来越紧密。通过把人、物和知识表示为图（graph）中的点，他们之间的关系表示为图中的边，使得多种多样的数据均可采用图数据表示。例如，人与人之间的社交网络，机器与机器之间的通信网络，知识图谱等均能够采用图数据表示。图数据库作为专门研发的用于存储、管理和查询图相关数据的专用数据库，因此受到了越来越多的关注。

目前，大部分的图存在信息缺失的问题。例如，图缺少部分点的信息、缺少边或者缺少边的信息。在此基础上，用户在数据库中基于图进行查询时，通常直接在数据库中已有的图中进行查询。

这样一来，当被查询的图缺失信息时，会导致对图数据库进行查询的准确性受到影响。

发明内容

本申请提供了一种图数据库系统及其查询方法。本申请有助于提高图数据库系统查询结果的准确性。本申请提供的技术方案如下：

第一方面，本申请提供了图数据库系统。该图数据库系统包括：交互模块，用于接收第一用户发送的数据查询请求，数据查询请求指示在图数据库系统中查询指定事件；处理模块，用于基于数据查询请求，获取指定事件相关的第一图数据，第一图数据为第二图数据补全缺失的信息后的图数据，第二图数据基于数据查询请求在图数据库系统存储的图数据中查找得到，图数据用于指示多个实体及不同实体之间的关系；交互模块，还用于基于第一图数据向第一用户反馈针对数据查询请求的响应，响应携带有符合指定事件的数据。

由于数据查询请求的响应基于补全缺失的信息后的图数据得到，能够避免因图数据缺失信息导致无法查询得到所有查询结果，有助于提高查询结果的准确性。并且，本申请的图数据库系统通过自动化地补全图数据缺失的信息，能够在用户无感的情况下对图数据缺失的信息进行补全，不会使用户感知图数据缺失信息，能够提高用户对图数据库的信任程度，有助于提高图数据库的易用性和用户粘性。

在一种可能的实现方式中，响应还携带有数据符合指定事件的概率。这样一来，图数据库系统为第一用户提供了更详尽的查询结果，能够为第一用户提供更多的参考信息，有助于第一用户基于概率指示实体符合指定事件的程度作出更恰当的决策。

在一种可能的实现方式中，处理模块用于基于数据查询请求，获取指定事件相关的第二图数据及第二图数据缺失的目标信息，基于第二图数据、目标信息与指定事件的关联关系，获取目标信息的值，并向第二图数据添加目标信息的值，得到第一图数据；其中，目标信息用于指示以下一种或多种：实体的标识，实体的属性信息，不同实体之间的关系，或者，关系的属性信息。

在一种可能的实现方式中，图数据库系统还包括：存储模块和预测模块，存储模块用于存储图数据；处理模块用于基于数据查询请求，从存储模块获取指定事件相关的第二图数据，确定第二图数据缺失的目标信息，并向预测模块提供第二图数据和目标信息；预测模块用于基于第二图数据和目标信息，预测目标信息的值，并向第二图数据添加目标信息的值，得到第一图数据，向处理模块提供第一图数据；处理模块用于基于第一图数据，得到针对数据查询请求的响应。

由于该预测模块为图数据库系统的内置模块，预测模块与目前图数据库系统已有的功能能够深度整合，使得用户在使用图数据库系统的过程中能够无感使用预测模块进行查询，实现对具有预测功能的图数据库系统的无感访问，能够提高用户的使用体验。且相较于图数据库系统仅能通过外部调用的方式使用预测功能的技术，实现该预测功能的过程中图数据库系统与外部的预测模块之间无需进行反复的数据

迁移，降低了图数据库系统的查询开销，且降低了图数据库系统的使用门槛，有助于进一步提高图数据库系统的易用性。

在一种可能的实现方式中，预测模块基于经过预训练的机器学习模型实现。

在一种可能的实现方式中，预测模块用于获取经过预训练的机器学习模型预先计算得到的第一计算数据，基于第一计算数据、第二图数据和目标信息，预测目标信息的值。当基于预先计算得到的第一计算数据得到第一图数据时，使得在获取第一图数据的过程中无需基于第二图数据执行计算得到第一计算数据的计算过程，减少得到第一图数据过程中的计算量，能够加速得到第一图数据的过程，进而加速对图数据库系统60的查询过程。并且，由于该机器学习模型经过预训练得到，使得第一计算数据反映的关系在较大概率上能够满足第二图数据中各种数据与目标信息的关系，因此，即使该第一计算数据不是基于第二图数据和目标信息计算得到，也能够将第一计算数据对得到的第一图数据的准确度的影响控制在可接受范围内，从而保证得到的第一图数据的准确度。

在一种可能的实现方式中，机器学习模型包括深度学习模型，第一计算数据为经过预训练的深度学习模型的嵌入层基于训练使用的图数据的计算结果，且第一计算数据作为训练使用的图数据的属性信息存储在存储模块中。这样一来，在获取图数据时，能够直接得到基于其得到的第一计算数据。

在一种可能的实现方式中，交互模块还用于接收第二用户对第二图数据设置的配置信息，配置信息用于指示第二图数据需要记载的信息；处理模块用于基于配置信息，确定第二图数据中缺失的目标信息。这样一来，能够实现对第二图数据的定制化补全。

在一种可能的实现方式中，配置信息用于指示第二图数据需要记载的信息的类型，类型指示以下一种或多种：实体的标识，实体的属性信息，不同实体之间的关系，或者，关系的属性信息。

第二方面，本申请提供了一种图数据库系统的查询方法。该图数据库系统的查询方法应用于图数据库系统。图数据库方法包括：接收第一用户发送的数据查询请求，数据查询请求指示在图数据库系统中查询指定事件；基于数据查询请求，获取指定事件相关的第一图数据，第一图数据为第二图数据补全缺失的信息后的图数据，第二图数据基于数据查询请求在图数据库系统存储的图数据中查找得到，图数据用于指示多个实体及不同实体之间的关系；基于第一图数据向第一用户反馈针对数据查询请求的响应，响应携带有符合指定事件的数据。

在一种可能的实现方式中，响应还携带有数据符合指定事件的概率。

在一种可能的实现方式中，基于数据查询请求，获取指定事件相关的第一图数据，包括：基于数据查询请求，获取指定事件相关的第二图数据及第二图数据缺失的目标信息，基于第二图数据、目标信息与指定事件的关联关系，获取目标信息的值，并向第二图数据添加目标信息的值，得到第一图数据；其中，目标信息用于指示以下一种或多种：实体的标识，实体的属性信息，不同实体之间的关系，或者，关系的属性信息。

在一种可能的实现方式中，基于第二图数据、目标信息与指定事件的关联关系，获取目标信息的值，包括：基于第二图数据和目标信息，采用经过预训练的机器学习模型预测目标信息的值。

在一种可能的实现方式中，基于第二图数据和目标信息，采用经过预训练的机器学习模型预测目标信息的值，包括：获取经过预训练的机器学习模型预先计算得到的第一计算数据，采用经过预训练的机器学习模型基于第一计算数据、第二图数据和目标信息，预测目标信息的值。

在一种可能的实现方式中，机器学习模型包括深度学习模型，第一计算数据为经过预训练的深度学习模型的嵌入层基于训练使用的图数据的计算结果，且第一计算数据作为训练使用的图数据的属性信息进行存储。

在一种可能的实现方式中，该图数据库系统的查询方法还包括：接收第二用户对第二图数据设置的配置信息，配置信息用于指示第二图数据需要记载的信息。相应的，获取第二图数据缺失的目标信息，包括：基于配置信息，确定第二图数据中缺失的目标信息。

在一种可能的实现方式中，配置信息用于指示第二图数据需要记载的信息的类型，类型指示以下一种或多种：实体的标识，实体的属性信息，不同实体之间的关系，或者，关系的属性信息。

第三方面，本申请提供了一种计算设备，包括存储器和处理器，存储器存储有程序指令，处理器运行程序指令以实现本申请第一方面以及其任一种可能的实现方式中提供的图数据库系统。

第四方面，本申请提供了一种计算设备集群，包括多个计算设备，多个计算设备包括多个处理器和多个存储器，多个存储器中存储有程序指令，多个处理器运行程序指令，使得计算设备集群实现本申请

第一方面以及其任一种可能的实现方式中提供的图数据库系统。

第五方面，本申请提供了一种计算机可读存储介质，该计算机可读存储介质为非易失性计算机可读存储介质，该计算机可读存储介质包括程序指令，当程序指令在计算设备上运行时，使得计算设备实现本申请第一方面以及其任一种可能的实现方式中提供的图数据库系统。

第六方面，本申请提供了一种包含指令的计算机程序产品，当计算机程序产品在计算机上运行时，使得计算机实现本申请第一方面以及其任一种可能的实现方式中提供的图数据库系统。

附图说明

图1是本申请实施例提供的一种图数据的示意图；

图2是本申请实施例提供的另一种图数据的示意图；

图3是本申请实施例提供的一种图数据库系统及其查询方法涉及的实施场景的结构示意图；

图4是本申请实施例提供的另一种图数据库系统及其查询方法涉及的实施场景的结构示意图；

图5是本申请实施例提供的一种数据中心中基础资源部署方式的示意图；

图6是本申请实施例提供的一种图数据库系统的结构示意图；

图7是本申请实施例提供的另一种图数据库系统的结构示意图；

图8是本申请实施例提供的又一种图数据库系统的结构示意图；

图9是本申请实施例提供的再一种图数据库系统的结构示意图；

图10是本申请实施例提供的一种数据查询的过程示意图；

图11是本申请实施例提供的一种查询计划树的示意图；

图12是本申请实施例提供的一种训练过程的示意图；

图13是本申请实施例提供的一种用于训练的查询计划树的示意图；

图14是本申请实施例提供的一种图数据库系统的查询方法的流程图；

图15是本申请实施例提供的另一种图数据库系统的查询方法的流程图；

图16是本申请实施例提供的一种计算设备的结构示意图；

图17是本申请实施例提供的一种计算设备集群的结构示意图；

图18是本申请实施例提供的另一种计算设备集群的结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

随着大数据时代的到来，物与物、人与人、知识与知识的关联越来越紧密。通过把人、物和知识等信息表示为图（graph）中的点，将他们之间的关系表示为图中的边，使得多种多样的数据均可采用图表示。例如，人与人之间的社交网络，机器与机器之间的通信网络，知识图谱等均能够采用图表示。采用图表示的数据称为图数据。目前，图数据库作为专门研发的用于存储、管理和查询图有关数据的专用数据库，受到了越来越多的关注。

图是将信息中的实体和实体之间的关系，分别抽象表达为顶点（vertex）以及顶点之间的关系的一种数据结构。图包括至少两个顶点和至少一条边。顶点用于表示实体。客观存在并能够相互区分的事物叫实体。在一些情况中顶点又称为点、结点或端点等。顶点之间的边（edge）用于表示实体和实体之间的关系。当两个顶点表示的实体之间具有关联关系时，两个顶点之间通过边连接。当两个顶点表示的实体之间不具有关联关系时，两个顶点之间不连接。在一种实现方式中，一个实体与另一实体关联包括以下一种或多种情况：一个实体主动关联另一实体，另一实体主动关联一个实体。其中，顶点和边都携带有信息。顶点的信息可以包括：顶点所代表的实体的名称（又称标识或标签（label））及其属性（property）信息等。边的信息可以包括：表示该边连接的两个顶点对应表示的两个实体之间的关系的属性信息。可选地，图可以是有向图或无向图。当图是有向图时，一个顶点指向另一个顶点的边称为该一个顶点的出边，相应的，该边称为该另一个顶点的入边。

在不同的应用场景中，图中的顶点和边具有不同的实际意义。例如，应用于好友推荐的场景中，图中的每个顶点可以表示一个用户。顶点的标识指示用户的标识，如姓名。顶点的属性信息指示对应用户的特点，如用户的年龄、职业或爱好等。图中的边表示边连接的顶点表示的两个用户之间的社交关系，

如两个用户之间的熟悉程度等。又例如，应用于商品推荐的场景中，图中的部分顶点表示用户，部分顶点表示商品。图中表示用户的顶点的标识指示用户的标识，如姓名。该顶点的属性信息指示对应用户的特点，如用户的年龄、职业或商品偏好等。图中表示商品的顶点的标识指示商品的标识，如商品名称或货号。该顶点的属性信息指示对应商品的特点，如商品分类和适用范围等。表示用户的顶点与表示商品的顶点之间的边表示用户与商品之间的交互关系，如购买关系或收藏关系等。又例如，应用于员工信息统计的场景中，图中的部分顶点表示员工，部分顶点表示地点，部分顶点表示公司。图中表示员工的顶点的标识指示员工的标识，如姓名。图中表示地点的顶点的标识指示地点的标识，如地点的名称。图中表示公司的顶点的标识指示公司的标识，如公司名称。表示员工的顶点与表示地点的顶点之间的边表示顶点指示的员工居住在顶点指示的地点。表示员工的顶点与表示公司的顶点之间的边表示顶点指示的员工属于顶点指示的公司，表示公司的顶点与表示地点的顶点之间的边表示顶点指示的公司位于顶点指示的地点。

示例地，图1是本申请实施例提供的一种指示员工信息的图的示意图。图1所示的图包括4个顶点和5条边。4个顶点分别指示员工A、员工B、公司C和地点D。表示员工A的顶点与表示地点D的顶点之间的边表示员工A居住在地点D。表示员工B的顶点与表示地点D的顶点之间的边表示员工B居住在地点D。表示员工A的顶点与表示公司C的顶点之间的边表示员工A在公司C工作。表示员工B的顶点与表示公司C的顶点之间的边表示员工B在公司C工作。表示公司C的顶点与表示地点D的顶点之间的边表示公司C位于地点D。图2是本申请实施例提供的一种指示社交关系的图的示意图。图2所示的图包括5个顶点和4条边，且每个顶点具有标识和属性信息。5个顶点的标识分别指示小张、小李、小王、小N和小O。表示小张的顶点的属性信息表示：小张为指定商品的潜在客户的概率和是高净值客户的概率不详，即图2中{buyer:null,rich:null}。表示小李的顶点的属性信息表示：小李大概率是指定商品的潜在客户，但小李是高净值客户的概率不详，即图2中{buyer:true,rich:null}。表示小王的顶点的属性信息表示：小王为指定商品的潜在客户的概率和是高净值客户的概率不详，即图2中{buyer:null,rich:null}。表示小O的顶点的属性信息表示：小O为指定商品的潜在客户的概率不详，小O大概率是高净值客户，即图2中{buyer:null,rich:true}。表示小N的顶点的属性信息表示：小N为指定商品的潜在客户的概率和是高净值客户的概率不详，即图2中{buyer:null,rich:null}。两个顶点之间的边表示两个顶点表示的人员之间的社交关系，该社交关系通过边的属性信息呈现。表示小张的顶点与表示小李的顶点之间的边表示小张和小李是好朋友。表示小李的顶点与表示小O的顶点之间的边表示顶点指示的小李和小O是好朋友。表示小李的顶点与表示小王的顶点之间的边表示小李和小王是好朋友。表示小王的顶点与表示小N的顶点之间的边表示小王和小N是好朋友。

图数据库（graph database）是基于图论实现的一种新型非关系型的数据库。它的数据存储结构和数据的管理方式都是以图论为基础的。即图数据库中数据与数据之间的关系均通过图表示，且图数据库的所有特性均通过图实现。例如，对图执行创建（creat）、读取（read，也称查找）、更新（update）和删除（delete）等操作，还有处理事务的能力和高可用性等，都基于图执行。

目前，图数据库的使用普遍存在以下问题：大部分图均存在缺失信息的问题。图缺失信息是指图缺少部分点的标识和属性（property）信息、缺少边的属性信息、或者缺少点与点之间的边。例如，在有些图数据库中，图中93.8%的人没有出生地点信息，78.5%的人没有国籍，68%的人没有任何职业信息，而在另一些图数据库中，图中50%的艺术家没有出生日期。在此基础上，用户在数据库中基于图进行查询时，通常直接在数据库中已有的图中进行查询。这样一来，当图缺失一些信息时，图数据库就难以返回用户期望的所有结果，会导致对图数据库进行查询的准确性受到影响。而这个问题又进一步因为现实中大部分图都缺失信息而变得更为严重。

对应于图1所示的图数据，图1中的实线的边是图数据中存在的，虚线的边是图数据应该记载但实际缺失的。假定用户想要基于该图数据查找居住在地点D的人，其查询语句为：

Match (p: people) – [:居住] -> (c: country {name: 地点D})

Return p.name

用户将该查询语句输入图数据库后，该图数据库基于该图数据输出的结果为：员工A。但是，实际居住在地点D的人还有员工B。由此可以看出，由于图数据缺失虚线指示的边，导致图数据库无法输出居住在地点D的人还有员工B的信息，可见对图数据库进行查询的准确性较差。

对应于图2所示的图数据，图2中表示不详的属性信息即为该图数据缺失的信息。假定用户想要基于该图数据中查找可能的高净值客户和可能购买指定商品的潜在客户，以便于对可能的高净值客户和潜在

客户进行提前关注，其查询语句为：

Match (f:Person {rich:true})

Match (v:Person{buyer:true})

Return f.name,v.name

用户将该查询语句输入图数据库后，该图数据库基于该图数据输出的结果为：f.name：小O，v.name：小李。该结果表示小O可能是高净值客户，小李可能购买该指定商品。但是，实际小张、小王和小O也可能购买该指定商品，小李也可能是高净值客户。由此可以看出，由于图数据中顶点缺失属性信息，导致图数据库无法输出这些查询结果，可见对图数据库进行查询的准确性较差。

鉴于此，本申请实施例提供了一种图数据库系统及查询方法。该图数据库系统包括：交互模块和处理模块。交互模块用于接收第一用户发送的数据查询请求，数据查询请求指示在图数据库系统中查询指定事件。处理模块用于基于数据查询请求，获取指定事件相关的第一图数据。第一图数据为第二图数据补全缺失的信息后的图数据。第二图数据基于数据查询请求在图数据库系统存储的图数据中查找得到。图数据用于指示多个实体及不同实体之间的关系。交互模块还用于基于第一图数据向第一用户反馈针对数据查询请求的响应，响应携带有符合指定事件的数据。

由此可知，本申请的图数据库系统在接收到用户发送的数据查询请求后，能够自动化地对该数据查询请求查询的第二图数据缺失的信息进行补全，得到第一图数据，然后根据该第一图数据向用户反馈针对该数据查询请求的响应。由于数据查询请求的响应基于补全缺失的信息后的图数据得到，能够避免因图数据缺失信息导致无法查询得到所有查询结果，有助于提高查询结果的准确性。并且，本申请的图数据库系统通过自动化地补全图数据缺失的信息，能够在用户无感的情况下对图数据缺失的信息进行补全，不会使用户感知图数据缺失信息，能够提高用户对图数据库的信任程度，有助于提高图数据库的易用性和用户粘性。

本文从实施场景、方法流程、硬件装置、软件装置等多个角度，对本申请的技术方案进行详细介绍。

下面先对本申请实施例的实施场景举例说明。

图3是本申请实施例提供的一种图数据库系统及其查询方法涉及的实施场景的结构示意图。如图3所示，该实施场景包括：一个或多个计算设备10和客户端20。该一个或多个计算设备10用于实现本申请实施例提供的图数据库。客户端20能够与一个或多个计算设备10建立通信连接。例如，客户端20与一个或多个计算设备10之间可以通过网络建立通信连接。可选的，该网络可以为局域网，也可以为互联网，还可以为其它网络，本申请实施例不作限定。

客户端20用于供用户与一个或多个计算设备20进行交互。在一种实现方式中，客户端20用于按照用户的指示向计算设备10发送数据查询指令。一个或多个计算设备20用于基于该数据查询指令在图数据库中进行查询，并向客户端20反馈查询结果。

在一种可实现方式中，客户端20可以为台式计算机、膝上型计算机、移动电话、智能手机、平板电脑、多媒体播放器、智能家电、人工智能设备、智能可穿戴设备、电子阅读器、智能车载设备或物联网设备等。计算设备10可以为服务器（如云服务器）。当该实施场景包括多个计算设备10时，该多个计算设备10可称为计算设备集群。此时，计算设备集群为由若干台服务器组成的服务器集群，或者是一个云计算服务中心实现。其中，云计算服务中心中部署有云服务提供商拥有的大量基础资源。例如云计算服务中心中部署有计算资源、存储资源和网络资源等。云计算服务中心可以利用该大量基础资源，实现本申请实施例提供的图数据库。

当计算设备集群通过云计算服务中心实现时，计算设备集群实现的图数据库的功能可以由云服务提供商在云平台抽象成一种图数据库云服务。此时，用户能够通过客户端20访问云管理平台，在云管理平台购买该图数据库云服务后，通过云管理平台使用计算设备集群提供的该图数据库云服务。可选地，云管理平台可以是中心云的云管理平台、边缘云的云管理平台或包括中心云和边缘云的云管理平台，本申请实施例对其不做具体限定。并且，该图数据库云服务可选由云管理平台提供为一项单独的云服务，或者，将该图数据库云服务作为其他云服务的附加云服务提供，本申请实施例对其不做具体限定。

当计算设备集群通过云计算服务中心实现时，如图4所示，本申请实施例提供的一种图数据库系统及其查询方法涉及的实施场景包括：数据中心1和客户端20。数据中心1和客户端20之间能够通过网络建立

通信连接。可选的，该网络可以为互联网，还可以为其它网络，本申请实施例不作限定。租户能够通过客户端20与数据中心1进行交互。例如，租户可以通过客户端20向数据中心1发送数据查询请求等信息。数据中心1用于基于客户端20发送的信息进行响应。

数据中心1中部署有云服务提供商拥有的大量基础设施，如计算资源、存储资源和网络资源等。例如，计算资源可以是能够提供计算能力的计算设备（如服务器等）。此时，计算设备10部署在数据中心1中。如图4所示，数据中心1包括云管理平台和基础设施（图4中未示出）。云管理平台和基础设施之间通过数据中心内部网络连接。云管理平台用于管理基础设施。基础设施用于提供公有云服务，如提供上述图数据库云服务。基础设施包括多个服务器。云服务可选部署在服务器中。云服务通过运行虚拟实例实现，因此也称为用于实现租户业务的虚拟实例部署在服务器中。租户能够通过其使用的客户端20向服务器发送云服务请求和相关信息，服务器能够对该云服务请求和相关信息进行处理，并基于经过处理的云服务请求和相关信息向租户提供云服务。例如，服务器能够通过本申请实施例提供的图数据库系统的查询方法，接收第一用户发送的数据查询请求，基于数据查询请求查找到第二图数据，补全该第二图数据缺失的信息得到第一图数据，基于第一图数据向第一用户反馈针对数据查询请求的响应。

云管理平台从逻辑功能上能够划分为：租户控制台、计算管理服务、网络管理服务、存储管理服务、鉴权服务和镜像管理服务。租户控制台提供界面或应用程序接口（application program interface，API）与租户交互。计算管理服务用于管理运行虚拟实例的服务器以及裸金属服务器。网络管理服务用于管理网络服务（如网关、防火墙等）。存储管理服务用于管理存储服务（如数据桶服务）。鉴权服务用于管理租户的账号和密码。镜像管理服务用于管理虚拟实例的镜像。

在图4所示的实施场景中，一个数据中心中设置有多个服务器。服务器包括硬件层和软件层。硬件层为服务器的常规配置。硬件层部署有处理器、内存、网卡、磁盘和总线等硬件设备。软件层包括安装并运行在服务器上的操作系统。相对虚拟机的操作系统可称为宿主机操作系统。宿主机操作系统中运行有虚拟机管理器（又可称为Hypervisor）。虚拟机管理器的作用是实现虚拟机的计算虚拟化、网络虚拟化以及存储虚拟化，并负责管理虚拟机。

虚拟机管理器中运行有云管理平台客户端。云管理平台客户端能够接收云管理平台发送的控制面命令，根据控制面控制命令在服务器上创建虚拟实例，并对虚拟实例进行全生命周期管理。例如，云管理平台客户端能够实时检测所在服务器的硬件资源的使用情况并上报至云管理平台。云管理平台确认要在某个服务器上创建虚拟实例时，会向该服务器上的云管理平台客户端发送虚拟实例创建命令，云管理平台客户端接收该命令后，在该服务器上创建虚拟实例。这样一来，租户可通过云管理平台在数据中心中创建、管理、登录和操作虚拟实例。

服务器可用于运行不同规格的虚拟机。虚拟机规格分为：通用计算型、内存优化型、超大内存型等等，每个类型下面还有具体规格，租户选择好虚拟机规格后，云管理平台在数据中心中选择支持该规格的服务器，并确定该服务器中上具有足够的空闲硬件资源，然后在该服务器上创建设置有该规格的虚拟机。通过云管理平台对服务器进行配置，能够实现对服务器硬件资源的分析与规划，根据服务器的硬件性能规划物理硬件的对应计算产品，如规划不同规格的虚拟机，以满足不同租户的差异化诉求。并且，根据不同规格的虚拟机的性能差异，能够差异化定价策略。例如，将高性能规格的虚拟实例以更高的价格对外售卖，将普通性能规格的虚拟实例以较低价格对外售卖，以便于租户按需购买虚拟实例。

在一种实现方式中，如图5所示，数据中心中基础资源的位置可以通过云资源部署区域（region）和可用区（availability zone，AZ）进行描述。租户可选基于特定的region、AZ中的资源部署云服务。其中，region是从地理位置和网络时延维度进行划分的。同一个region内使用同一个资源池，可以理解为共享弹性计算、块存储、对象存储、虚拟私有云（virtual private cloud，VPC）网络、弹性网络互连协议（elastic internet protocol，EIP）地址和镜像等公共服务。region分为通用region和专属region。通用region指面向公共租户提供通用云服务的region。专属region指承载同一类业务或面向特定租户提供业务服务的专用region。一个region通常包括多个AZ。一个region中的多个AZ间通过高速光纤相连，以满足租户跨AZ构建高可用性系统的需求。一个AZ是一个或多个图5所示的数据中心的集合。AZ内的计算、网络和存储等资源在逻辑上被划分成多个集群。

租户可通过其使用的客户端20向云管理平台发送指令，以在服务器中创建、管理、登录和操作虚拟实例，并使用该虚拟实例提供的云服务。示例地，云管理平台能够提供访问接口。访问接口可选以界面或API的形式提供。租户可操作客户端远程接入访问接口，以在云管理平台注册云账号和密码，并使用云

账号和密码登录云管理平台。云管理平台还能够对云账号和密码进行鉴权。鉴权成功后，租户可进一步在云管理平台选择并付费购买特定规格（处理器、内存、磁盘）的虚拟实例。租户付费购买虚拟实例成功后，云管理平台向租户提供所购买的虚拟实例的远程登录账号和密码。租户可使用远程登录账号和密码，在客户端远程登录该虚拟实例，在该虚拟实例中安装并运行租户的应用，以通过该应用实现租户的业务。

在一种实现方式中，本申请实施例提供的图数据库系统及其查询方法，可以通过一个或多个计算设备10运行可执行程序实现。例如，该图数据库系统及其查询方法的可执行程序可选以应用程序安装包的形式呈现，一个或多个计算设备10中安装该应用程序安装包后，能够通过运行该可执行程序实现该图数据库系统及其查询方法。当计算设备集群通过云计算服务中心实现时，本申请实施例提供的图数据库系统及其查询方法，可以通过数据中心1中的计算设备运行可执行程序实现。

应当理解的是，以上内容是对本申请实施例提供的图数据库系统及其查询方法的实施场景的示例性说明，并不构成对于该图数据库系统及其查询方法的实施场景的限定，本领域普通技术人员可知，随着业务需求的改变，其实施场景可以根据应用需求进行调整，本申请实施例对其不做具体限定。并且，当本申请实施例提供的图数据库系统及其查询方法应用于其他场景时，该方法的可执行程序也可以应用程序安装包的形式呈现或者采用其他方式呈现，本申请实施例对其不做一一列举。

下面先对本申请实施例提供的图数据库系统的实现方式进行介绍。图6是本申请实施例提供的一种图数据库系统的结构示意图。如图6所示，该图数据库系统60包括：交互模块601和处理模块602。

交互模块601用于接收第一用户发送的数据查询请求，并向处理模块602提供该数据查询请求。数据查询请求指示在图数据库系统60中查询指定事件。例如，数据查询请求指示查询可能的高净值客户和可能购买指定商品的潜在客户。第一用户在需要查询指定事件时，可在其使用的客户端上执行指定操作，以触发客户端向图数据库系统60发送该数据查询请求。客户端向图数据库系统60发送该数据查询请求后，由交互模块601接收该数据查询请求。在一种可能的实现方式中，图数据库系统60能够为客户端提供应用程序接口（application programming interface，API），客户端通过访问该API能够触发数据查询请求。

处理模块602用于基于数据查询请求，获取指定事件相关的第一图数据。第一图数据为第二图数据补全缺失的信息后的图数据。也即是，第一图数据携带有第二图数据包括的信息及第二图数据中缺失的信息。第二图数据基于数据查询请求在图数据库系统60存储的图数据中查找得到。数据查询请求指示在图数据库系统60中查询指定事件，则第二图数据为图数据库系统60中与该指定事件相关的图数据。其中，图数据与指定事件相关是指该图数据中记载有指定事件的查询结果。例如，假设图数据库系统60中记载有图1和图2所示的图数据，当数据查询请求指示查询的指定事件为可能的高净值客户和可能购买指定商品的潜在客户时，由于图2所示的图数据记载有该指定事件的查询结果，图1所示的图数据未记载有该指定事件的查询结果，因此，处理模块602可以确定图2所示的图数据为第二图数据。需要说明的是，图数据库系统60通常会记载很多图数据，且记载有指定事件的查询结果的图数据可能有多个，则记载有指定事件的查询结果的多个图数据均为第二图数据。相应的，处理模块602能够获取与多个第二图数据一一对应的多个第一图数据，任一个第一图数据为对应的第二图数据补全缺失的信息后的图数据。

另外，第一图数据和第二图数据可以均为一个完整的图，或者均为一个完整的图中的子图（subgraph）。子图为完整的图的一部分，包括完整的图中的部分顶点以及部分边。子图又称为图中的分区（partition）。一个完整的图可以有多个子图，不同子图包括的顶点中的至少部分不同，不同子图包括的边中的至少部分不同。当第二图数据为子图时，该子图可选在完整的图中基于数据查询请求划分得到。示例地，一个完整的图用于表示具有同一属性的多个实体之间的关系，该多个实体可划分为多个实体集合。每个实体集合包括多个实体中的部分实体，且每个实体集合具有一个子属性，不同实体集合具有的子属性不同。一个实体集合中的多个实体组成的图即为完整的图中的子图。数据查询请求可指示查询具有子属性的实体满足指定事件的情况，则基于该数据查询请求得到的第二图数据为指示具有该子属性的多个实体及其之间关系的子图。例如，假设某完整的图记载浏览过某品牌所有商品的客户的信息。即该完整的图中的多个实体具有的同一属性为：多个实体均浏览过该品牌的商品。当数据查询请求指示查询可能购买该品牌的某个商品的潜在客户，基于该数据查询请求可在该完整的图中确定指示浏览过该商品的客户的信息，得到图2所示的子图。此时，子图中的多个实体具有的子属性为：多个实体均浏览过该品牌的某一指定商品。需要说明的是，以上仅为对完整的图和子图对应的属性的示例性说明，完整的图和

子图还可以具有其他属性，此处不对其进行一一举例。

交互模块601还用于在处理模块602获取到数据查询请求查询的指定事件相关的第一图数据后，基于第一图数据向第一用户反馈针对数据查询请求的响应。响应携带有符合指定事件的数据。其中，处理模块602对第二图数据补全得到第一图数据后，基于该第一图数据得到数据查询请求的查询结果，然后向交互模块601提供该查询结果。交互模块601得到该查询结果后，向第一用户反馈指示该查询结果的响应。或者，处理模块602获取第一图数据后，向交互模块601提供该第一图数据。交互模块601得到该第一图数据后，基于该第一图数据得到数据查询请求的查询结果，然后向第一用户反馈指示该查询结果的响应。本申请实施例以处理模块602向交互模块601提供数据查询请求的查询结果，由交互模块601向第一用户反馈指示该查询结果的响应为例进行说明。符合指定事件的数据指示第一图数据中符合指定事件的顶点指示的实体。例如，数据查询请求指示查询可能的高净值客户和可能购买指定商品的潜在客户时，图2所示的图数据针对该数据查询请求的查询结果指示可能购买该指定商品的潜在客户为小李、小张、小王和小O，可能是高净值客户的是小O和小李，即针对该数据查询请求的响应携带的数据指示“可能购买该指定商品的潜在客户为小李、小张、小王和小O，可能是高净值客户的是小O和小李”。

可选地，响应可选不仅携带有符合指定事件的数据，还携带有该数据指示的实体符合指定事件的概率。也即是，响应携带有符合指定事件的实体及该实体符合该指定事件的概率。这样一来，图数据库系统为第一用户提供了更详尽的查询结果，能够为第一用户提供更多的参考信息，有助于第一用户基于概率指示实体符合指定事件的程度作出更恰当的决策。进一步地，响应携带的实体可选为符合指定事件的概率大于指定概率阈值的实体。也即是，响应携带的实体符合指定事件的概率均大于指定概率阈值，且该响应还携带有该实体符合指定事件的概率的具体数值。该指定阈值的取值可根据应用需求确定，本申请实施例对其不做具体限定。示例地，表1和表2展示了上面示例的数据查询请求的查询结果。根据表1可以看出，查询结果除了指示可能购买指定商品的潜在客户的姓名，还指示这些潜在客户购买指定商品的概率。根据表2可以看出，查询结果除了指示可能是高净值客户的姓名，还指示这些客户是高净值客户的概率。

表1

f:Person {rich:true}	概率
小李	1
小张	0.7656757
小王	0.202511
小N	0.31452
小O	1

表2

v:Person{buyer:true}	概率
小李	1
小张	0.8656757
小王	0.1231242
小N	0.4251211
小O	0.7546752

在一种可能的实现方式中，处理模块602在得到数据查询请求后，可先基于该数据查询请求获取第二图数据，并获取第二图数据缺失的目标信息，然后基于第二图数据、目标信息与指定事件的关联关系，获取该目标信息的值，并向第二图数据添加目标信息的值，得到第一图数据。示例地，处理模块602获取数据查询请求后，可获取该数据查询请求指示查询的指定事件，然后在图数据库系统60记载的图数据中选择记载有该指定事件的查询结果的图数据，然后将选择的图数据确定为第二图数据。例如，如图7所示，图数据库系统60还包括：存储模块603。存储模块603用于存储图数据。则处理模块602在得到数据查询请求后，可先基于该数据查询请求指示查询的指定事件，从存储模块603获取与指定事件相关的第二图数据。

目标信息用于指示需要为第二图数据补齐的信息所属的类型。在一种可能的实现方式中，目标信息用于指示以下一种或多种：图数据中顶点指示的实体的标识，图数据中顶点指示的实体的属性信息，图

数据中不同顶点指示的不同实体之间的关系，或者，该关系的属性信息。处理模块602获取数据查询请求后，可获取该数据查询请求指示查询的图数据应当记载的信息。数据查询请求指示查询的图数据应当记载的信息可通过该图数据的配置信息指示。处理模块602获取第二图数据的配置信息后，能够根据该配置信息确定第二图数据缺失的目标信息。其中，配置信息用于指示数据查询请求指示查询的图数据需要记载的信息所属的类型。例如，配置信息指示以下一种或多种：图数据中顶点指示的实体的标识，图数据中顶点指示的实体的属性信息，图数据中不同顶点指示的不同实体之间的关系，或者，该关系的属性信息。

在一种可能的实现方式中，处理模块602获取配置信息后，可基于该配置信息，检查第二图数据是否记载了该配置信息指示的所有信息，并将配置信息指示但该第二图数据未记载的信息确定为第二图数据缺失的目标信息。示例地，假设第二图数据的配置信息用于指示：第二图数据中每个顶点均需要记载顶点指示的实体的标识和分别指示多个指定属性的属性信息，每两个顶点之间是否具有边基于这两个顶点表示的实体之间的关系确定，且每条边需要记载其属性信息。则当第二图数据中存在顶点没有记载其指示的实体的标识时，该第二图数据缺失的目标信息指示的内容包括：需要为第二图数据补齐该顶点指示的实体的标识。当第二图数据中存在顶点没有记载多个指定属性中第一属性的属性信息时，该第二图数据缺失的目标信息指示的内容包括：需要为第二图数据补齐该顶点携带该第一属性的属性信息。该第一属性为多个指定属性中的任一个。当第二图数据中存在两个顶点之间未连接有边，但这两个顶点指示的实体之间的关系表示这两个实体之间有关联时，该第二图数据缺失的目标信息指示的内容包括：需要为第二图数据补齐这两个顶点之间的边。当第二图数据中存在两个顶点之间的边没有记载属性信息时，该第二图数据缺失的目标信息指示的内容包括：需要为第二图数据补齐这两个顶点之间的边的属性信息。例如，假设处理模块602基于数据查询请求确定的第二图数据为图2所示的图数据。该图数据的配置信息指示该图数据需要记载的信息包括：每个查询对象及其是否为高净值人群和购买指定商品的潜在客户的属性信息，以及浏览过该指定商品的查询对象之间的社交关系。处理模块602基于该配置信息能够确定图2所示的图缺少以下信息：指示小张的顶点缺少是否为高净值人群和购买指定商品的潜在客户的属性信息，指示小李的顶点缺少是否为高净值人群的属性信息，指示小王的顶点缺少是否为高净值人群和购买指定商品的潜在客户的属性信息，指示小N的顶点缺少是否为高净值人群和购买指定商品的潜在客户的属性信息，指示小O的顶点缺少是否为购买指定商品的潜在客户的属性信息。这些缺少的信息即为图2所示的图数据缺失的目标信息。需要说明的是，配置信息可能为图数据实际缺失的信息中的部分信息，此时配置信息未指示的信息可认为是在实现业务时无需关注或参考价值小的信息，即认为不是图数据缺失的信息。比如，图2所示的图数据实际还缺少各个顶点指示的客户的年龄和性别，但是客户的年龄和性别对查询提供的参考价值较小，因此可视为客户的年龄和性别不是图2所示的图数据缺失的信息。

可选地，图数据的配置信息可由图数据库的用户设置得到。则交互模块601还用于接收第二用户对第二图数据设置的配置信息。这样一来，能够实现对第二图数据的定制化补全。其中，用户可基于需求多次为图数据设置配置信息。例如，用户在初次为图数据设置配置信息后，还可对图数据的配置信息进行修改、增加和删除。该第二用户与第一用户可以相同或不同。例如，第二用户为图数据库系统60的管理员，第一用户为图数据库系统60的使用者，此时第一用户与第二用户不同。又例如，第一用户和第二用户均为使用图数据库系统60进行查询的用户。

第二用户对图数据设置配置信息可通过数据库定义语言（data definition language，DDL）实现。DDL用于指示图数据需要记载的信息，如图中的顶点和/或边的属性信息、顶点与顶点之间的边等。下面对DDL表示的配置信息进行示例性地说明：

[Create|Drop] ML Index on ([node selection query | edge selection query] {property selection query})，用于指示在图数据中构建或删除一些顶点或者是边的某些属性信息。当该DDL指示在图数据中构建一些顶点或者是边的某些属性信息时，是指该图数据需要记载有这些属性信息，此时该DDL即指示该图数据的配置信息。当图数据缺少这些属性信息时，处理模块602能够基于该DDL为图数据补全这些属性信息。其中，node selection query指示需要构建或删除的顶点，edge selection query指示需要构建或删除的边，property selection query指示需要构建或删除的属性信息。例如，对于图2，Create ML Index on (p: Person {buyer, rich})指示在所有指示Person的顶点上构建Person是否为指定商品的buyer和是否为rich的属性信息。

[Create|Drop] ML Edge on ((node selection query 1) - [`edgeName`] -> (node selection query 2))，用于指示在图数据中构建或删除node selection query 1选中的node和node selection query 2选中的node之间的名为

“edgeName”的边。当该DDL指示在图数据中构建node selection query 1选中的node和node selection query 2选中的node之间的名为“edgeName”的边时，是指该图数据需要记载这些边，此时该DDL即指示该图数据的配置信息。当图数据缺少这些边时，处理模块602能够基于该DDL为图数据补全这些边。例如，对于图2，Create ML Edge on (p: Person - [:Know] -> p2: Person)指示在所有指示Person的顶点之间构建顶点指示的Person是否互相认识（Know）的边。

需要说明的是，以上只是对DDL表示的配置信息的示意性举例，其还可能有更多的实现方式。例如，DDL还能够指示配置信息的生效时限、生效范围和针对配置信息对图数据进行补全的方式等。其中，配置信息的生效时限是指为图数据设置配置信息后，该配置信息生效的时间。例如，配置信息的生效时限指示在设置配置信息后，立即生效。即在设置配置信息后，就按照新配置的配置信息确定图数据缺失的信息。或者，配置信息的生效时限指示在设置配置信息且到达配置信息的生效周期后，才生效该配置信息。即图数据系统中设置有配置信息的生效时长，在设置配置信息后，若旧的配置信息的生效时长未到达该生效时长，则依旧按照旧的配置信息确定图数据缺失的信息。在旧的配置信息的生效时长到达该生效时长后，再按照新配置的配置信息确定图数据缺失的信息。生效范围指示配置信息适用的范围。例如，生效范围指示配置信息对完整的图适用，或者，生效范围指示配置信息对完整的图的某些子图适用。

在一种可能的实现方式中，对第二图数据进行补全的操作可通过预测的方式得到。该预测的过程可由处理模块602执行，或者由处理模块602调用其他模块执行。示例地，如图7所示，图数据库系统60还包括：预测模块604。处理模块602可调用该预测模块604对第二图数据进行补全，以得到第一图数据。也即是，处理模块602用于基于数据查询请求获取第二图数据及其缺失的目标信息后，向预测模块604提供第二图数据和目标信息。预测模块604用于在得到第二图数据和目标信息后，基于第二图数据、目标信息与数据查询请指示查询的指定事件的关联关系，预测目标信息的值，然后向第二图数据添加预测得到的目标信息的值，以得到第一图数据，并向处理模块602提供该第一图数据。其实现过程为：预测模块604在得到第二图数据和目标信息后，对第二图数据中已有信息、目标信息和指定事件进行分析，得到第二图数据中各种数据、目标信息与指定事件的关联关系，并根据该关联关系对目标信息的值进行预测，然后基于预测得到的目标信息的值补全第二图数据，得到第一图数据。

在一种可能的实现方式中，预测模块604通过机器学习（machine learning，ML）技术实现。例如，预测模块604基于经过预训练的机器学习模型实现。示例地，机器学习模型可选为深度学习模型。例如，机器学习模型为图神经网络（graph neural network，GNN）、图卷积网络（graph convolutional network，GCN）、图注意力网络（graph attention network，GAT）、图采样与聚合（graph sample and aggregate，GraphSAGE）或其他能实现对应功能的机器学习模型。此时，处理模块602获取第二图数据及其缺失的目标信息后，可向机器学习模型输入第二图数据以及其缺失的目标信息，并接收机器学习模型输出的第一图数据。相应的，本申请的图数据库系统60又称为图神经数据库（neural graph database）系统。

其中，当预测模块604通过该深度学习模型实现时，深度学习模型的输出还包括目标信息的值的置信度。任一目标信息的值的置信度指示该目标信息的值的可信程度。则根据预测模块604提供的第一图数据得到数据查询请求的查询结果时，还能够基于该置信度得到查询结果涉及的实体符合指定事件的概率。例如，继续以前面图2所示的图数据为例，假设深度学习模型输出的第一图数据中补全的缺失信息指示以下内容：指示小李的顶点的属性信息指示小李为高净值人群的置信度为1；指示小张的顶点的属性信息指示小张为高净值人群的置信度为0.7656757，小张为购买指定商品的潜在客户的置信度为0.8656757；指示小王的顶点的属性信息指示小王为高净值人群的置信度为0.202511，小王为购买指定商品的潜在客户的置信度为0.1231242；指示小N的顶点的属性信息指示小N为高净值人群的置信度为0.31452，小N为购买指定商品的潜在客户的置信度为0.4251211；指示小O的顶点的属性信息指示小O为购买指定商品的潜在客户的置信度为0.7546752。在基于该第一图数据得到查询结果时，可将该置信度直接确定为查询结果涉及的实体符合指定事件的概率，即得到表1和表2所示的结果。

当预测模块604基于经过预训练的机器学习模型实现时，机器学习模型还用于获取经过预训练的机器学习模型预先计算得到的第一计算数据，然后基于第一计算数据、第二图数据和目标信息，预测目标信息的值，进而得到第一图数据。该第一计算数据可选基于训练数据计算得到。示例地，第一计算数据为经过预训练的深度学习模型的嵌入层（embedding）基于训练数据的计算结果。在一种可能的实现方式中，机器学习模型在获取第一计算数据、第二图数据和目标信息后，可根据第二图数据和目标信息在第一计算数据中筛选符合该第二图数据和该目标信息的计算数据，然后根据筛选得到的计算数据预测目标信息

的值，然后根据目标信息的值补全第二图数据，得到第一图数据。例如，用于对深度学习模型进行训练的图数据包括图数据库系统60中用于供用户查询的图数据。这些图数据通常采用文本表示，在深度学习模型的训练过程中，深度学习模型的嵌入层会将文本表示的图数据转换为向量（vector）表示，然后基于向量执行后续训练过程。在完成深度学习模型的训练过程后，预测模块604可获取并存储嵌入层在训练过程中转换得到的向量，并在深度学习模型预测目标信息的值时，根据第二图数据和目标信息在嵌入层在训练过程中转换得到的向量中筛选用于表示第二图数据的向量，然后基于筛选得到的向量预测目标信息的值。其中，嵌入层将文本表示的图数据转换为向量表示，是将文本表示的图数据转换为机器语言的一种实现方式，深度学习模型还可基于需求将文本标识的图数据转换为其他类型的机器语言，本申请实施例对其不做一一举例。

可选地，该第一计算数据可存储在存储模块603中。在一种可能的实现方式中，基于任一图数据得到的第一计算数据可选作为该图数据的属性信息存储在该图数据中。这样一来，在获取图数据时，能够直接得到基于其得到的第一计算数据。或者，图数据库系统60包括至少两个存储模块603，其中一个存储模块603用于存储图数据，另一个存储模块603用于存储该第一计算数据。例如，如图8所示，图数据库系统60包括图数据存储模块6031和计算数据存储模块6032，图数据存储模块6031用于存储图数据，计算数据存储模块6032用于存储第一计算数据。其中，计算数据存储模块6032在存储第一计算数据后，还可将该第一计算数据的元数据等相关数据存储在图数据存储模块6031中，以便于处理模块602在图数据存储模块6031中读取第二图数据时能够获取该相关数据，并基于该相关数据从计算数据存储模块6032中获取该相关数据指示的第一计算数据。

当基于预先计算得到的第一计算数据得到第一图数据时，使得在获取第一图数据的过程中无需基于第二图数据执行计算得到第一计算数据的计算过程，减少得到第一图数据过程中的计算量，能够加速得到第一图数据的过程，进而加速对图数据库系统60的查询过程。并且，由于该机器学习模型经过预训练得到，使得第一计算数据反映的关系在较大概率上能够满足第二图数据中各种数据与目标信息的关系，因此，即使该第一计算数据不是基于第二图数据和目标信息计算得到，也能够将第一计算数据对得到的第一图数据的准确度的影响控制在可接受范围内，从而保证得到的第一图数据的准确度。

当采用机器学习模型对第二图数据进行补全得到第一图数据时，图数据库系统60能够自动化地对第二图数据进行补全，能够在用户无感知的情况下实现缺失信息的补全，不会使用户感知图数据缺失信息，能够提高用户对图数据库系统60的信任程度，有助于提高图数据库系统60的易用性和用户粘性。并且，当采用机器学习模型对第二图数据进行补全得到第一图数据时，可降低图数据库系统60对图数据的信息完备性的要求，使得仅标记有少量点和边的信息的图数据也能够加入图数据系统中。当图数据中的点和边的信息由用户（如图数据系统的管理员）标注得到时，能够减少用户标注图数据的工作量。另外，由于该预测模块604为图数据库系统60的内置模块，预测模块604与目前图数据库系统60已有的功能能够深度整合，使得用户在使用图数据库系统60的过程中能够无感使用预测模块604进行查询，实现对具有预测功能的图数据库系统60的无感访问，能够提高用户的使用体验。且相较于图数据库系统60仅能通过外部调用的方式使用预测功能的技术，实现该预测功能的过程中图数据库系统60与外部的预测模块604之间无需进行反复的数据迁移，降低了图数据库系统60的查询开销，且降低了图数据库系统60的使用门槛，有助于进一步提高图数据库系统60的易用性。其中，用户无感使用预测模块604进行查询，包括但不限于用户对查询语言无感，用户对图数据库原有功能与预测模块604之间的交互无感等。

另外，处理模块602在得到第一图数据后，可选将该第一图数据保存。在处理模块602再接收到数据查询请求时，可选先基于数据查询请求确定该第一图数据是否为该数据查询请求指示查询的图数据，当该第一图数据是该数据查询请求指示查询的图数据时，处理模块602可直接基于该第一图数据反馈针对该数据查询请求的响应。此时无需再对第二图数据执行补全操作，能够简化处理模块602的运算过程，减少计算开销。当该第一图数据不是该数据查询请求指示查询的图数据时，处理模块602再基于数据查询请求获取第二图数据，然后对第二图数据补全得到第一图数据，再基于补全得到的第一图数据反馈响应。

需要说明的是，处理模块602在基于数据查询请求获取第一图数据的过程中，可以先基于数据查询请求进行预处理，然后基于预处理的结果获取第一图数据。示例地，处理模块602的功能可通过解析单元（parser）、优化单元（optimizer）和处理单元（processor）实现。解析单元用于对文本表示的数据查询请求进行解析，并向优化单元提供解析结果。解析结果用于指示查询诉求和该数据查询请求指示查询的图数据应当记载的信息，即解析结果指示数据查询请求的查询诉求及其对应的配置信息。解析结果可采

用抽象语法树（abstract syntax tree，AST）表示。抽象语法树是源代码语法结构的一种抽象表示。它以树状的形式表现编程语言的语法结构，树上的每个节点都表示源代码中的一种结构。优化单元用于基于解析结果生成数据查询请求需要执行的操作流，并向处理单元提供该操作流。处理单元用于基于该操作流，获取数据查询请求指示查询的第二图数据。操作流指示具有先后顺序的多个操作。操作流可通过查询计划树（query plan tree）表示。查询计划树是图数据库系统60能够识别的呈树状的数据结构，其具有多个节点，每个节点用于指示一个操作命令。操作流从查询计划树的根节点开始执行，除根结点外的其余节点的执行依赖于其父节点的执行结果。由于操作流指示具有先后顺序的多个操作，该多个操作的执行需要参考的数据不同，则该操作流也指示需要获取的数据的先后顺序。相应的，处理单元从存储模块603获取第二图数据时，可根据该操作流指示的数据依次获取该第二图数据中的数据。

由此可以看出，本申请的图数据库系统60可由多个部件实现。图9为图8所示的图数据库系统60中通过多个部件实现数据查询的示意图。如图9所示，交互模块601也称图查询接口（graph query interface），用于负责和用户进行对接，向用户提供图查询的能力。例如，交互模块601用于接收第一用户发送的数据查询请求，并向第一用户反馈针对数据查询请求的响应。处理模块602包括解析单元6021、优化单元6022和处理单元6023。解析单元用于接收交互模块601提供的数据查询请求，如图10中的步骤①，然后将用户的数据查询请求解析为抽象语法树，并向优化单元提供该抽象语法树，如图10中的步骤②。优化单元用于将抽象语树转换为查询计划树，并按需对该查询计划树进行优化，并向处理单元提供优化后的查询计划树，如图10中的步骤③。处理单元用于根据查询计划树从图数据存储模块6031查询相关的点和边，如图10中的步骤④。图数据存储模块6031向处理单元提供该相关的点和边，如图10中的步骤⑤。处理单元在得到图数据存储模块6031提供的点和边后，向预测模块604提供从图数据存储模块6031得到的点和边，如图10中的步骤⑥。处理单元从图数据存储模块6031得到的点和边组成第二图数据。预测模块604得到处理单元提供的第二图数据后，基于该第二图数据，对该第二图数据进行补全，得到第一图数据，并向处理单元输出第一图数据，如图10中的步骤⑨。处理单元接收预测模块604输出的第一图数据，并基于该第一图数据向交互模块601提供数据查询请求的查询结果，如图10中的步骤⑩。其中，若计算数据存储模块6032中存储有第一计算数据，预测模块604对第二图数据进行补全时，可先从计算数据存储模块6032中获取该第一计算数据，如图10中的步骤⑦。计算数据存储模块6032向预测模块604提供该第一计算数据如图10中的步骤⑧。预测模块604得到第一计算数据后，再基于该第一计算数据对第二图数据进行补全。其中，图10中步骤顺序用于指示对应步骤的执行顺序，数字小的步骤先执行。

示例地，假设机器学习模型为GCN模型，依旧以前面对图2的查询为例进行说明，当数据查询请求指示查询的指定事件为小李是否为高净值客户时，图11为处理单元生成的查询计划树的示意图。如图11所示，该查询计划树包括两个分支，即图11中从左侧汇入Union算子的左分支和从右侧汇入Union算子的右分支。左分支的根节点指示：先从图数据存储模块6031中，获取所有指示小李的顶点及顶点之间的边。该功能通过图11中NodeByPropertyScan(x.name=小李)算子实现。该根节点的第一级子节点指示从获取的所有顶点中选取指示是否为高净值人群的属性信息的值为true的所有顶点。该功能通过图11中Filter(x.rich != null & x.rich = true)算子实现。右分支的根节点指示：先从图数据存储模块6031中，获取所有指示小李的顶点及顶点之间的边。该功能通过图11中NodeByPropertyScan(x.name=小李)算子实现。该根节点的第一级子节点指示从获取的所有顶点中选取指示是否为高净值人群的属性信息的值为不详的所有顶点。该功能通过图11中Filter(x.rich=null)算子实现。该根节点的第二级子节点（即第一级子节点的子节点）指示从图数据存储模块6031中获取选取的顶点在两跳之内的所有顶点和边，基于所有从图数据存储模块6031获取的顶点和边，得到第二图数据。该功能通过图11中ExpandAsSubgraph(x,2)算子实现。任一顶点在两跳之内的顶点是指以该任一顶点为起点，经过两条边能够到达的顶点。由于该第二图数据为图数据存储模块6031中存储的图中的一部分，因此该第二图数据为该图的子图。该根节点的第三级子节点（即第二级子节点的子节点）指示处理单元调用预测模块604的接口使用GCN模型。GCN模型基于第二图数据对所有值为不详的指示是否为高净值人群的属性信息进行预测，并将预测结果记录在第二图数据中，得到第一图数据，并向处理单元提供该第一图数据。该功能通过图11中GNNInference(GCN,Person.rich)算子实现。该根节点的第四级子节点（即第三级子节点的子节点）指示处理单元从第一图数据中筛选出所有属性信息指示顶点指示的客户为高净值人群的顶点，得到数据查询请求的查询结果。该功能通过图11中Filter(x.rich=ture)算子实现。Union算子用于汇总左分支和右分支的查询结果，并向ProduceResults算子提供汇总结果。ProduceResults算子用于基于汇总结果向交互模块601提供查询结果。

下面对机器学习模型的训练过程进行说明。该训练过程包括依次执行的多个子训练过程。子训练过程包括：处理模块602获取训练使用的数据查询请求，基于该数据查询请求获取该数据查询请求指示查询的训练用的图数据，并确定该训练用的图数据缺失的目标信息，并向机器学习模型输入该目标信息和训练用的图数据，使得机器学习模型基于该目标信息和训练用的图数据进行模型训练。在完成多个子训练过程后，处理模块602可获取经过训练的机器学习模型对每个训练用的图数据计算得到的第一计算数据，并存储该第一计算数据。

类似于图数据库系统60的查询过程，处理模块602也可先对训练使用的数据查询请求进行预处理，然后基于预处理的执行后续的训练操作。示例地，当处理模块602的功能通过解析单元、优化单元和处理单元实现时，解析单元用于获取交互模块601提供的训练使用的数据查询请求，如图12中的步骤①，然后对文本表示的训练使用的数据查询请求进行解析，并向优化单元提供解析结果，如图12中的步骤②。解析结果用于指示训练使用的数据查询请求指示查询的图数据应当记载的信息，即解析结果指示训练使用的数据查询请求对应的配置信息。优化单元用于基于解析结果生成训练使用的数据查询请求需要执行的操作流，并向处理单元提供该操作流，如图12中的步骤③。处理单元用于基于该操作流，从图数据存储模块6031获取训练使用的数据查询请求指示查询的第二图数据，如图12中的步骤④。图数据存储模块6031向处理单元提供该相关的点和边，如图12中的步骤⑤。其中，处理单元从存储模块603获取第二图数据时，可根据该操作流指示的数据依次获取该第二图数据中的数据。处理单元在得到图数据存储模块6031提供的点和边后，向预测模块604提供从图数据存储模块6031得到的点和边，如图12中的步骤⑥。处理单元从图数据存储模块6031得到的点和边组成第二图数据。预测模块604得到处理单元提供的第二图数据后，基于该第二图数据，对该第二图数据进行补全，得到第一图数据，并向处理单元输出第一图数据，如图12中的步骤⑦。在完成多个子训练过程后，处理模块602可从经过训练的机器学习模型对每个训练用的图数据计算得到的第一计算数据，将该第一计算数据存储在计算数据存储模块6032中，如图12中的步骤⑧。计算数据存储模块6032在存储该第一计算数据后，可将该第一计算数据的元数据等相关数据存储在图数据存储模块6031中，如图12中的步骤⑨。其中，图12中步骤顺序用于指示对应步骤的执行顺序，数字小的步骤先执行。

示例地，假设机器学习模型为GCN模型，依旧以前面对图2的查询为例进行说明，当训练使用的数据查询请求指示查询的指定事件为可能的高净值客户时，图13为处理单元生成的用于训练的查询计划树的示意图。如图13所示，该查询计划树的根节点指示：先从图数据存储模块6031中，获取所有指示人的顶点及顶点之间的边。该功能通过图13中NodeByPropertyScan(Person)算子实现。该根节点的第一级子节点指示从图数据存储模块6031中获取根节点获取的所有顶点在两跳之内的所有顶点和边，基于所有从图数据存储模块6031获取的顶点和边，得到第二图数据（即子图）。该功能通过图13中ExpandAsSubgraph(2)算子实现。该根节点的第二级子节点（即第一级子节点的子节点）指示处理单元调用预测模块604的接口对GCN模型进行训练。在训练过程中，GCN模型需要基于第二图数据对所有值为不详的指示是否为高净值人群的属性信息进行预测，并将预测结果记录在第二图数据中，得到第一图数据，并向处理单元提供该第一图数据。该功能通过图13中GNNTraining(GCN)算子实现。该根节点的第三级子节点（即第二级子节点的子节点）指示处理单元从预测模块604获取GCN模型的embedding层基于训练用的图数据的计算结果（即第一计算数据），然后将该第一计算数据存储在计算数据存储模块6032。计算数据存储模块6032在存储该第一计算数据后，可将该第一计算数据的元数据等相关数据存储在图数据存储模块6031中。获取第一计算数据的功能通过图13中Pre-ComputeEmbedding算子实现。该根节点的第四级子节点（即第三级子节点的子节点）指存储完成训练的GCN模型。该功能通过图13中StoreModel算子实现。

由上可知，本申请的图数据库系统在接收到用户发送的数据查询请求后，能够自动化地对该数据查询请求查询的第二图数据缺失的信息进行补全，得到第一图数据，然后根据该第一图数据向用户反馈针对该数据查询请求的响应。由于数据查询请求的响应基于补全缺失的信息后的图数据得到，能够避免因图数据缺失信息导致无法查询得到所有查询结果，有助于提高查询结果的准确性。并且，本申请的图数据库系统通过自动化地补全图数据缺失的信息，能够在用户无感的情况下对图数据缺失的信息进行补全，不会使用户感知图数据缺失信息，能够提高用户对图数据库的信任程度，有助于提高图数据库的易用性和用户粘性。并且，本申请提供的图数据库系统能够较大程度地复用现有图数据库系统的功能，在做图数据库系统的升级更新时，能够复用用于实现现有图数据库系统的一些可执行程序，能够减少重复的开发工作，降低开发维护成本，提高图数据库系统的升级更新效率。

下面对本申请实施例提供的图数据库系统的查询方法的实现过程进行说明。该图数据库系统的查询方法应用于本申请实施例提供的图数据库系统。图14是本申请实施例提供的一种图数据库系统的查询方法的流程图。如图14所示，该图数据库系统的查询方法包括以下步骤：

步骤1401、接收第一用户发送的数据查询请求，数据查询请求指示在图数据库系统中查询指定事件。

该步骤1401的实现过程相应参考前面图数据库系统中的相关描述，此处不再赘述。

步骤1402、基于数据查询请求，获取指定事件相关的第一图数据，第一图数据为第二图数据补全缺失的信息后的图数据，第二图数据基于数据查询请求在图数据库系统存储的图数据中查找得到，图数据用于指示多个实体及不同实体之间的关系。

在一种可能的实现方式中，基于数据查询请求，获取指定事件相关的第一图数据，包括：基于数据查询请求，获取指定事件相关的第二图数据及第二图数据缺失的目标信息，基于第二图数据、目标信息与指定事件的关联关系，获取目标信息的值，并向第二图数据添加目标信息的值，得到第一图数据。其中，目标信息用于指示以下一种或多种：实体的标识，实体的属性信息，不同实体之间的关系，或者，关系的属性信息。

可选地，基于第二图数据、目标信息与指定事件的关联关系，获取目标信息的值，包括：基于第二图数据和目标信息，采用经过预训练的机器学习模型预测目标信息的值。其中，基于第二图数据和目标信息，采用经过预训练的机器学习模型预测目标信息的值，包括：获取经过预训练的机器学习模型预先计算得到的第一计算数据，采用经过预训练的机器学习模型基于第一计算数据、第二图数据和目标信息，预测目标信息的值。

在一种可能的实现方式中，机器学习模型包括深度学习模型，第一计算数据为经过预训练的深度学习模型的嵌入层基于训练使用的图数据的计算结果，且第一计算数据作为训练使用的图数据的属性信息进行存储。

在一种可能的实现方式中，如图15所示，该方法还包括：步骤1404、接收第二用户对第二图数据设置的配置信息，配置信息用于指示第二图数据需要记载的信息。相应的，获取第二图数据缺失的目标信息，包括：基于配置信息，确定第二图数据中缺失的目标信息。也即是，步骤1402包括步骤14021、基于数据查询请求，获取指定事件相关的第二图数据，基于配置信息，确定第二图数据中缺失的目标信息，基于第二图数据、目标信息与指定事件的关联关系，获取目标信息的值，并向第二图数据添加目标信息的值，得到第一图数据。其中，配置信息用于指示第二图数据需要记载的信息的类型，类型指示以下一种或多种：实体的标识，实体的属性信息，不同实体之间的关系，或者，关系的属性信息。

该步骤1402的实现过程相应参考前面图数据库系统中的相关描述，此处不再赘述。

步骤1403、基于第一图数据向第一用户反馈针对数据查询请求的响应，响应携带有符合指定事件的数据。

可选地，响应还携带有数据符合指定事件的概率。该步骤1403的实现过程相应参考前面图数据库系统中的相关描述，此处不再赘述。

综上所述，在本申请实施例提供的图数据库的查询方法中，图数据库系统在接收到用户发送的数据查询请求后，能够自动化地对该数据查询请求查询的第二图数据缺失的信息进行补全，得到第一图数据，然后根据该第一图数据向用户反馈针对该数据查询请求的响应。由于数据查询请求的响应基于补全缺失的信息后的图数据得到，能够避免因图数据缺失信息导致无法查询得到所有查询结果，有助于提高查询结果的准确性。并且，本申请的图数据库系统通过自动化地补全图数据缺失的信息，能够在用户无感的情况下对图数据缺失的信息进行补全，不会使用户感知图数据缺失信息，能够提高用户对图数据库的信任程度，有助于提高图数据库的易用性和用户粘性。并且，本申请提供的图数据库系统能够较大程度地复用现有图数据库系统的功能，在做图数据库系统的升级更新时，能够复用用于实现现有图数据库系统的一些可执行程序，能够减少重复的开发工作，降低开发维护成本，提高图数据库系统的升级更新效率。

需要说明的是，本申请实施例提供的图数据库系统的查询方法的步骤先后顺序可以进行适当调整，步骤也可以根据情况进行相应增减。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，可轻易想到变化的方法，都应涵盖在本申请的保护范围之内，因此不再赘述。

在本申请中，交互模块601、处理模块602、存储模块603和预测模块604均可以通过软件实现，或者

可以通过硬件实现。示例性地，接下来以交互模块601为例，介绍交互模块601的实现方式。类似的，处理模块602、存储模块603和预测模块604的实现方式可以参考交互模块601的实现方式。

作为软件功能单元的一种举例，交互模块601可以包括运行在计算实例上的代码。其中，计算实例可以包括物理主机（计算设备）、虚拟机、容器中的至少一种。进一步地，上述计算实例可以是一台或者多台。例如，交互模块601可以包括运行在多个主机/虚拟机/容器上的代码。需要说明的是，当本申请的图数据库系统部署于云平台中时，用于运行该代码的多个主机/虚拟机/容器可以分布在相同的区域（region）中，也可以分布在不同的region中。进一步地，用于运行该代码的多个主机/虚拟机/容器可以分布在相同的可用区（availability zone，AZ）中，也可以分布在不同的AZ中，每个AZ包括一个云数据中心或多个地理位置相近的云数据中心。其中，通常一个region可以包括多个AZ。

同样，用于运行该代码的多个主机/虚拟机/容器可以分布在同一个虚拟私有云（virtual private cloud，VPC）中，也可以分布在多个VPC中。其中，通常一个VPC设置在一个region内，同一region内两个VPC之间，以及不同region的VPC之间跨区通信需在每个VPC内设置通信网关，经通信网关实现VPC之间的互连。

作为硬件功能单元的一种举例，交互模块601可以包括至少一个计算设备，如服务器等。或者，交互模块601也可以是利用专用集成电路（application-specific integrated circuit，ASIC）实现、或可编程逻辑器件（programmable logic device，PLD）实现的设备等。其中，上述PLD可以是复杂程序逻辑器件（complex programmable logical device，CPLD）、现场可编程门阵列（field-programmable gate array，FPGA）、通用阵列逻辑（generic array logic，GAL）或其任意组合实现。

交互模块601包括的多个计算设备可以分布在相同的region中，也可以分布在不同的region中。交互模块601包括的多个计算设备可以分布在相同的AZ中，也可以分布在不同的AZ中。同样，交互模块601包括的多个计算设备可以分布在同一个VPC中，也可以分布在多个VPC中。其中，多个计算设备可以是服务器、ASIC、PLD、CPLD、FPGA和GAL等计算设备的任意组合。

需要说明的是，在其他实施例中，交互模块601、处理模块602、存储模块603和预测模块604中任一个可以用于执行图数据库系统的查询方法中的任意步骤。交互模块601、处理模块602、存储模块603和预测模块604负责实现的步骤可根据需要指定，通过交互模块601、处理模块602、存储模块603和预测模块604分别实现图数据库系统的查询方法中不同的步骤来实现图数据库系统的查询装置的全部功能。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的各个组成部分的具体工作过程，可以参考前述方法实施例中的对应内容，在此不再赘述。

下面对本申请实施例涉及的基本硬件结构举例说明。

本申请还提供一种计算设备1600。如图16所示，计算设备1600包括：总线1602、处理器1604、存储器1606和通信接口1608。处理器1604、存储器1606和通信接口1608之间通过总线1602通信。计算设备1600可以是服务器或终端设备。应理解，本申请不限定计算设备1600中的处理器、存储器的个数。

总线1602可以是外设部件互连标准（peripheral component interconnect，PCI）总线或扩展工业标准结构（extended industry standard architecture，EISA）总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图16中仅用一条线表示，但并不表示仅有一根总线或一种类型的总线。总线1602可包括在计算设备1600各个部件（例如，存储器1606、处理器1604、通信接口1608）之间传送信息的通路。

处理器1604可以包括中央处理器(central processing unit，CPU)、图形处理器（graphics processing unit，GPU）、微处理器（micro processor，MP）或者数字信号处理器（digital signal processor，DSP）等处理器中的任意一种或多种。

存储器1606可以包括易失性存储器(volatile memory)，例如随机存取存储器(random access memory，RAM)。处理器1604还可以包括非易失性存储器(non-volatile memory)，例如只读存储器(read-only memory，ROM)，快闪存储器，机械硬盘（hard disk drive，HDD）或固态硬盘（solid state drive，SSD）。

存储器1606中存储有可执行的程序代码，处理器1604执行该可执行的程序代码以分别实现前述交互模块601、处理模块602、存储模块603和预测模块604的功能，从而实现图数据库系统的查询方法。也即，存储器1606上存有用于执行图数据库系统的查询方法的指令。

通信接口1608使用例如但不限于网络接口卡、收发器一类的收发模块，来实现计算设备1600与其他设备或通信网络之间的通信。

本申请实施例还提供了一种计算设备集群。该计算设备集群包括至少一台计算设备。该计算设备可以是服务器，例如是中心服务器、边缘服务器，或者是本地数据中心中的本地服务器。在一些实施例中，计算设备也可以是台式机、笔记本电脑或者智能手机等终端设备。

如图17所示，计算设备集群包括至少一个计算设备1600。计算设备集群中的一个或多个计算设备1600中的存储器1606中可以存有相同的用于执行图数据库系统的查询方法的指令。

在一些可能的实现方式中，该计算设备集群中的一个或多个计算设备1600的存储器1606中也可以分别存有用于执行图数据库系统的查询方法的部分指令。换言之，一个或多个计算设备1600的组合可以共同执行用于执行图数据库系统的查询方法的指令。

需要说明的是，计算设备集群中的不同的计算设备1600中的存储器1606可以存储不同的指令，分别用于执行图数据库系统的查询装置的部分功能。也即，不同的计算设备1600中的存储器1606存储的指令可以实现交互模块601、处理模块602、存储模块603和预测模块604中的一个或多个模块的功能。

在一些可能的实现方式中，计算设备集群中的一个或多个计算设备可以通过网络连接。其中，网络可以是广域网或局域网等等。图18示出了一种可能的实现方式。如图18所示，两个计算设备1600A和1600B之间通过网络进行连接。具体地，通过各个计算设备中的通信接口与网络进行连接。在这一类可能的实现方式中，计算设备1600A中的存储器1606中存有执行交互模块601和处理模块602的功能的指令。同时，计算设备1600B中的存储器1606中存有执行存储模块603和预测模块604的功能的指令。

图18所示的计算设备集群之间的连接方式可以是考虑到本申请提供的图数据库系统的查询方法需要大量地存储数据，因此考虑将存储模块603和预测模块604实现的功能交由计算设备1600B执行。

应理解，图18中示出的计算设备1600A的功能也可以由多个计算设备1600完成。同样，计算设备1600B的功能也可以由多个计算设备1600完成。

本申请实施例还提供了另一种计算设备集群。该计算设备集群中各计算设备之间的连接关系可以类似的参考图17和图18计算设备集群的连接方式。不同的是，该计算设备集群中的一个或多个计算设备1600中的存储器1606中可以存有相同的用于执行图数据库系统的查询方法的指令。

在一些可能的实现方式中，该计算设备集群中的一个或多个计算设备1600的存储器1606中也可以分别存有用于执行图数据库系统的查询方法的部分指令。换言之，一个或多个计算设备1600的组合可以共同执行用于执行图数据库系统的查询方法的指令。

本申请实施例还提供了一种包含指令的计算机程序产品。计算机程序产品可以是包含指令的，能够运行在计算设备上或被储存在任何可用介质中的软件或程序产品。当计算机程序产品在至少一个计算设备上运行时，使得至少一个计算设备执行图数据库系统的查询方法。

本申请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可以是计算设备能够存储的任何可用介质或者是包含一个或多个可用介质的数据中心等数据存储设备。可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘）等。该计算机可读存储介质包括指令，指令指示计算设备执行图数据库系统的查询方法，或指示计算设备执行图数据库系统的查询方法。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，该程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

需要说明的是，本申请所涉及的信息（包括但不限于用户设备信息、用户个人信息等）、数据（包括但不限于用于分析的数据、存储的数据、展示的数据等）以及信号，均为经用户授权或者经过各方充分授权的，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。例如，本申请中涉及到的原始数据和可执行代码等都是在充分授权的情况下获取的。

在本申请实施例中，术语“第一”、“第二”和“第三”仅用于描述目的，而不能理解为指示或暗示相对重要性。术语“至少一个”是指一个或多个，术语“多个”指两个或两个以上，除非另有明确的限定。

本申请中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/

或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

最后应说明的是：以上实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的保护范围。

权 利 要 求 书

1.一种图数据库系统，其特征在于，所述图数据库系统包括：

交互模块，用于接收第一用户发送的数据查询请求，所述数据查询请求指示在所述图数据库系统中查询指定事件；

处理模块，用于基于所述数据查询请求，获取所述指定事件相关的第一图数据，所述第一图数据为第二图数据补全缺失的信息后的图数据，所述第二图数据基于所述数据查询请求在所述图数据库系统存储的图数据中查找得到，所述图数据用于指示多个实体及不同实体之间的关系；

所述交互模块，还用于基于所述第一图数据向所述第一用户反馈针对所述数据查询请求的响应，所述响应携带有符合所述指定事件的数据。

2. 如权利要求1所述的系统，其特征在于，所述响应还携带有所述数据符合所述指定事件的概率。

3. 如权利要求1或2所述的系统，其特征在于，

所述处理模块用于基于所述数据查询请求，获取所述指定事件相关的所述第二图数据及所述第二图数据缺失的目标信息，基于所述第二图数据、所述目标信息与所述指定事件的关联关系，获取所述目标信息的值，并向所述第二图数据添加所述目标信息的值，得到所述第一图数据；

其中，所述目标信息用于指示以下一种或多种：所述实体的标识，所述实体的属性信息，所述不同实体之间的关系，或者，所述关系的属性信息。

4. 如权利要求3所述的系统，其特征在于，所述图数据库系统还包括：存储模块和预测模块，所述存储模块用于存储图数据；

所述处理模块用于基于所述数据查询请求，从所述存储模块获取所述指定事件相关的所述第二图数据，确定所述第二图数据缺失的目标信息，并向所述预测模块提供所述第二图数据和所述目标信息；

所述预测模块用于基于所述第二图数据和所述目标信息，预测所述目标信息的值，并向所述第二图数据添加所述目标信息的值，得到所述第一图数据，向所述处理模块提供所述第一图数据；

所述处理模块用于基于所述第一图数据，得到所述针对所述数据查询请求的响应。

5. 如权利要求4所述的系统，其特征在于，所述预测模块基于经过预训练的机器学习模型实现。

6. 如权利要求5所述的系统，其特征在于，

所述预测模块用于获取所述经过预训练的机器学习模型预先计算得到的第一计算数据，基于所述第一计算数据、所述第二图数据和所述目标信息，预测所述目标信息的值。

7. 如权利要求6所述的系统，其特征在于，所述机器学习模型包括深度学习模型，所述第一计算数据为经过预训练的深度学习模型的嵌入层基于训练使用的图数据的计算结果，且所述第一计算数据作为所述训练使用的图数据的属性信息存储在所述存储模块中。

8. 如权利要求3至7任一所述的系统，其特征在于，

所述交互模块还用于接收第二用户对所述第二图数据设置的配置信息，所述配置信息用于指示所述第二图数据需要记载的信息；

所述处理模块用于基于所述配置信息，确定所述第二图数据中缺失的所述目标信息。

9. 如权利要求8所述的系统，其特征在于，所述配置信息用于指示所述第二图数据需要记载的信息的类型，所述类型指示以下一种或多种：所述实体的标识，所述实体的属性信息，所述不同实体之间的关系，或者，所述关系的属性信息。

10.一种图数据库系统的查询方法，其特征在于，所述方法应用于图数据库系统，所述图数据库方法

包括：

接收第一用户发送的数据查询请求，所述数据查询请求指示在所述图数据库系统中查询指定事件；

基于所述数据查询请求，获取所述指定事件相关的第一图数据，所述第一图数据为第二图数据补全缺失的信息后的图数据，所述第二图数据基于所述数据查询请求在所述图数据库系统存储的图数据中查找得到，所述图数据用于指示多个实体及不同实体之间的关系；

基于所述第一图数据向所述第一用户反馈针对所述数据查询请求的响应，所述响应携带有符合所述指定事件的数据。

11. 如权利要求10所述的方法，其特征在于，所述响应还携带有所述数据符合所述指定事件的概率。

12. 如权利要求10或11所述的方法，其特征在于，所述基于所述数据查询请求，获取所述指定事件相关的第一图数据，包括：

基于所述数据查询请求，获取所述指定事件相关的所述第二图数据及所述第二图数据缺失的目标信息，基于所述第二图数据、所述目标信息与所述指定事件的关联关系，获取所述目标信息的值，并向所述第二图数据添加所述目标信息的值，得到所述第一图数据；

其中，所述目标信息用于指示以下一种或多种：所述实体的标识，所述实体的属性信息，所述不同实体之间的关系，或者，所述关系的属性信息。

13. 如权利要求12所述的方法，其特征在于，所述基于所述第二图数据、所述目标信息与所述指定事件的关联关系，获取所述目标信息的值，包括：

基于所述第二图数据和所述目标信息，采用经过预训练的机器学习模型预测所述目标信息的值。

14. 如权利要求13所述的方法，其特征在于，所述基于所述第二图数据和所述目标信息，采用经过预训练的机器学习模型预测所述目标信息的值，包括：

获取所述经过预训练的机器学习模型预先计算得到的第一计算数据，采用所述经过预训练的机器学习模型基于所述第一计算数据、所述第二图数据和所述目标信息，预测所述目标信息的值。

15. 如权利要求14所述的方法，其特征在于，所述机器学习模型包括深度学习模型，所述第一计算数据为经过预训练的深度学习模型的嵌入层基于训练使用的图数据的计算结果，且所述第一计算数据作为所述训练使用的图数据的属性信息进行存储。

16. 如权利要求12至15任一所述的方法，其特征在于，所述方法还包括：

接收第二用户对所述第二图数据设置的配置信息，所述配置信息用于指示所述第二图数据需要记载的信息；

所述获取所述第二图数据缺失的目标信息，包括：

基于所述配置信息，确定所述第二图数据中缺失的所述目标信息。

17. 如权利要求16所述的方法，其特征在于，所述配置信息用于指示所述第二图数据需要记载的信息的类型，所述类型指示以下一种或多种：所述实体的标识，所述实体的属性信息，所述不同实体之间的关系，或者，所述关系的属性信息。

18. 一种计算设备集群，其特征在于，包括多个计算设备，所述多个计算设备包括多个处理器和多个存储器，所述多个存储器中存储有程序指令，所述多个处理器运行程序指令，使得所述计算设备集群实现权利要求1至9任一所述的图数据库系统。

19. 一种计算机可读存储介质，其特征在于，包括程序指令，当所述程序指令在计算设备上运行时，使得所述计算设备实现权利要求1至9任一所述的图数据库系统。

20. 一种包含指令的计算机程序产品，其特征在于，当所述指令被计算设备集群运行时，使得所述计算设备集群实现权利要求1至9任一所述的图数据库系统。