

Graph Databases



图数据库

Ian Robinson

[美] *Jim Webber* 著

Emil Eifrem

刘璐 梁越 译

O'REILLY®



人民邮电出版社
POSTS & TELECOM PRESS

图数据库

Ian Robinson

[美] *Jim Webber* 著

Emil Eifrem

刘璐 梁越 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc. 授权人民邮电出版社出版

人 民 邮 电 出 版 社
北 京

图书在版编目(CIP)数据

图数据库 / (美) 罗宾逊 (Robinson, I.), (美) 韦伯 (Webber, J.), (美) 艾弗雷姆 (Eifrem, E.) 著; 刘璐, 梁越译. — 北京: 人民邮电出版社, 2015. 2

书名原文: Graph databases

ISBN 978-7-115-37604-6

I. ①图… II. ①罗… ②韦… ③艾… ④刘… ⑤梁… III. ①图象数据库 IV. ①TP311.132

中国版本图书馆CIP数据核字(2015)第002673号

版权声明

Copyright ©2013 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2015. Authorized translation of the English edition, 2015 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书中文简体版由 O'Reilly Media, Inc. 授权人民邮电出版社出版。未经出版者书面许可, 对本书的任何部分不得以任何方式复制或抄袭。

版权所有, 侵权必究。

◆ 著 [美] Ian Robinson Jim Webber Emil Eifrem

译 刘璐 梁越

责任编辑 杨海玲

责任印制 张佳莹 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

三河市海波印务有限公司印刷

◆ 开本: 787×1000 1/16

印张: 11.75

字数: 234千字

2015年2月第1版

印数: 1-3000册

2015年2月河北第1次印刷

著作权合同登记号 图字: 01-2013-8994号

定价: 49.00元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315

内容提要

世界上的大部分领域实际上都可以建模成图,而软件开发人员或是数据库管理人员却在辛辛苦苦地将这些图转化成关系型数据库中的表。想象一下,假如你再也不用去做这样的转化,假如数据库的迁移可以变得轻松简单,你愿意去接受一个全新的数据库吗?图数据库就是这样一个完全不同于关系型数据库的新型数据库,它处理的是大规模的数据和不断变化的需求。

本书系统地介绍了图数据库的历史由来、建模方法、工作原理和一些真实的用户用例,详细地说明了图数据解决的是什么样的问题,并以 Neo4j 数据库和 Cypher 查询语言为例,阐述了图数据库的建模方法和领域用例,最后还介绍了图数据库的工作原理以及一些实用的图论算法。本书的三位作者均为 Neo4j Technology 公司的技术高手,他们对图数据库及其解决方案有丰富的经验,其中一位甚至还是 Neo4j 图数据库的联合创始人。

本书适合开发人员和数据库管理人员了解和学习图数据库时阅读,作为一门新的知识和独特的数据库领域来拓宽视野,也适合提供解决方案的负责人了解行业动向和新的解决问题的方式。通过阅读本书,读者可以对图数据库这一领域有一个透彻的了解。

序

图无处不在，即我们所知的图数据库的诞生

1999 年，我们的开发团队每天都工作 23 小时，至少感觉上是这样的。似乎每天都有一个疯狂的想法付诸实施，而这些想法刚刚获得数百万美元的融资。我们的竞争对手都有数百个工程师，而我们只是一个 20 人左右的开发团队，其中还有 10 个工程师要将他们的大部分时间花在关系型数据库上。

我们花了一段时间去寻找原因。当深挖企业内容管理系统的数据持久层时，我们意识到我们的软件不但需要管理大量独立的离散数据项，还要管理数据项之间的关联。尽管离散数据可以很容易地使用关系表来处理，但对于关联数据，其存储更具挑战性，而且其查询速度也异常缓慢。

出于对关系型数据库完全的绝望，我和我的两个 Neo 联合创始人 Johan 和 Peter 开始尝试使用其他数据模型处理数据，尤其是那些基于图的模型。我们欣喜地发现以图为中心的模型似乎可以替代扁平的 SQL 语法，使开发者更容易地操作关联数据。我们观察到，有了图数据模型，我们的开发团队在数据库上节约了一半的时间。

当然，我们要说，我们不是独自在战斗。图论已经存在了近 300 年，且被广泛应用于各种不同的数学问题，一定可以有能拥抱图的数据库！

因此我们在早期互联网上使用 Altavista¹ 搜索类似的数据库却一无所获。经过几个月的调查，我们（天真地）着手从无到有地创建一个天然支持图的数据库。我们的愿景是：在保留所有关系型数据库的已被验证的特性（事务、ACID、触发器等）的基础上使用 21 世纪的数据模型。于是 Neo 项目诞生了，伴随它诞生的还有我们现在所知的图数据库。

新千年的第一个 10 年，Google、Facebook、Twitter 这些改变世界的新业务融入大众生活。它们有一个共同点：将关联的数据——图——作为业务的中心。15 年后的今天，图无处不在。

① 年轻读者可能会震惊曾经存在一个没有 Google 的时代。是的，恐龙曾经统治地球，而搜索引擎只有 Altavista、Lycos 和 Excite 这些，我们主要用它们在互联网上寻找可以购买宠物食品的电子商务网站。

前言

图数据库应对的是当今一个宏观商业世界的大趋势：凭借高度关联的数据中复杂而动态的联系获得洞察力并赢得竞争优势。无论我们想了解的是客户之间的联系，电话或数据中心网络元素之间的联系，娱乐产品制作者和消费者之间的联系，还是基因和蛋白质之间的联系，都会涉及大量的高度关联的数据。这些数据又会构成庞大的图，而理解和分析这些图的能力将成为公司在未来十年的核心竞争力。

对于任何达到一定规模或价值的数据，图数据库都是呈现和查询这些关联数据的最好方式。关联数据是这样的一种数据：它需要我们首先理解它的组成元素之间的关联方式。为了理解这个，很多时候我们需要去给这些事物之间的关联加以命名和限定。

尽管在一段时间以前一些大公司就已经意识到这个问题，并着手开发各自专有的图处理技术，但我们正处在一个技术全民化的时代。现如今，通用的图数据库已经成为现实，主流用户不必去投资建设自己的图架构，就可以享受关联数据带来的好处。

这次图数据和图思考复兴的伟大之处正在于图论本身并不是一个新事物。自 18 世纪欧拉创建了图论以来，数学家、社会学家、人类学家和其他领域工作者一直在研究和完善图论。然而，图论和图思考在信息管理中的应用却是最近几年的事情。那个时候，图数据库已经在社交网络、主数据管理（master data management）、地理空间、推荐系统以及其他领域帮我们解决了许多重要问题。有两股力量驱动我们对图数据库的日益关注：一股力量是那些获得巨大商业成功的公司，如 Facebook、Google 和 Twitter，他们都将自己的商业模式紧紧地围绕在他们专有的图技术上；另一股力量就是通用的图数据库开始进入到技术领域里。

关于本书

本书的目的是为技术实践者介绍图和图数据库，包括开发人员、数据库专业人士和技术决策者。阅读本书将会让你对图数据库有一个贴近实际的理解。我们将演示图模型如何“塑造”数据，以及如何用图数据库查询、推断、理解和处理数据。我们讨论了多种适合用图数据库处理的问题，配以从实际应用中提取的用户案例，还将展示如何规划和实施一个图数据库解决方案。

美国:

O'Reilly Media Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

中国:

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室 (100035)

奥莱利技术咨询(北京)有限公司

我们为本书提供了专门的网页,上面有勘误表、示例,以及其他额外的信息,可以通过 <http://oreil.ly/graph-databases> 访问该网页。

为本书提供建议和咨询技术问题,请发送邮件到 bookquestions@oreilly.com。

想了解更多关于我们的书籍、课程、会议,以及新闻等信息,请登录我们的网站:
<http://www.oreilly.com>。

我们的其他联系方式如下。

Facebook: <http://facebook.com/oreilly>

Twitter: <http://twitter.com/oreillymedia>

YouTube: <http://www.youtube.com/oreillymedia>

致谢

我们要感谢本书的技术审稿人 Michael Hunger、Colin Jack、Mark Needham 和 Pramod Sadalage。

我们赞赏并感谢本书的编辑 Nathan Jepson。

在本书成书的过程中,Neo Technology 的同事极大地贡献了他们的时间、经验和努力。特别感谢 Anders Nawroth 对本书的工具组提供了宝贵的帮助;感谢 Andrés Taylor 对 Cypher 部分提供的热情帮助;感谢 Philip Rathle 对行文提供的建议和所做的贡献。

感谢 Neo4j 社区的所有人,感谢你们这些年对图数据库做出的许多贡献。

还要特别感谢我们的家人 Lottie、Tiger、Elliot、Kath、Billy、Madelene 和 Noomi,谢谢他们付出的爱和支持。

O'Reilly Media, Inc. 介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

以 Facebook 为例，它的创建理念认为：尽管诸如使用者的名字、行为等离散信息是有价值的，但是使用者的联系更具有价值。Facebook 的创始人 Mark Zuckerberg 正是基于在社交图（social graph）中捕获社交联系这样的洞见建立了一个宏大的帝国。

同样，Google 的 Larry Page 和 Sergey Brin 提出了存储和处理离散的网页文件以及网页文件之间的关联的方法。Google 建立了网络图（Web graph），这使其成为过去 10 年中最具影响力的公司。

如今，图已成功地被网络巨头以外的公司接纳。全球最大的物流公司之一使用图数据库建立包裹的实时路由，一家大型航空公司使用图处理其多媒体内容元数据，一家顶尖金融公司基于 Neo4j 完全改写了其信息系统基础设施。图数据库在几年前还鲜为人知，而如今已广泛应用于医疗、零售、石油和天然气、媒体、游戏，以及更多的领域，并以爆炸式的速度繁荣发展。

这些想法发展出一种新的工具——通用数据库管理技术，即包含关联数据且支持图的思想。这就是我在 1999 年挣扎于关系型数据库时所希望找到的现成的可用工具。

我希望本书能很好地将图数据库技术这个新兴世界介绍给你，同时希望本书能鼓舞你在下一个项目中使用图数据库，这样你也将能解开图的非凡力量。祝你好运！

——Emil Eifrem

Neo4j 联合创始人，NeoTechnology 公司 CEO

本书的排版约定

本书使用下列排版约定。

- 中文楷体：用于新术语、邮件地址、文件名称以及文件扩展名。
- 等宽字体：用于程序代码，包括文字段落中引用的程序元素，如变量、函数名、数据库、数据类型、环境变量、代码语句和关键字。
- 加粗等宽字体：用于命令以及其他需要用户逐字输入的文字。
- 等宽斜体：应该由用户提供的值或根据上下文确定的值。



此图标表示一个提示、建议或一般性注释。



此图标表示警告或慎用。

代码示例的使用

这本书的目的是帮助你完成工作。一般来说，如果这本书包括代码示例，你可以在你的程序和文档中使用本书的代码。你不需要联系我们申请权限，除非你要直接复制相当大的一部分代码。例如，在编写程序的过程中使用了本书中的几段代码，这不需要授权。售卖或者分发 O'Reilly 的图书示例光盘显然是需要授权的。引用本书或引用示例代码来回答问题是不需要授权的，但将本书的大量示例代码纳入产品的文档是需要授权的。

我们对你在使用时声明引用信息表示赞赏，但并不做强制要求。引用信息通常包括书名、作者、出版社和 ISBN，如“*Graph Databases* by Ian Robinson, Jim Webber, and Emil Eifrem (O'Reilly). Copyright 2013 Neo Technology, Inc., 978-1-449-35626-2”。

如果你认为对示例代码的使用需要授权，请通过这个邮箱联系我们：permissions@oreilly.com。

联系我们

如果你想就本书发表评论或有任何疑问，敬请联系出版社。

目录

第 1 章 简介	1	3.4.3 测试模型	32
1.1 图是什么	1	3.5 跨域模型	34
1.2 图领域概览	3	3.5.1 创建莎士比亚图	37
1.2.1 图数据库	4	3.5.2 开始查询	38
1.2.2 图计算引擎	5	3.5.3 声明查找的信息模式	39
1.3 图数据库的威力	6	3.5.4 约束匹配	40
1.3.1 性能	6	3.5.5 处理结果	41
1.3.2 灵活性	7	3.5.6 查询链	42
1.3.3 敏捷性	7	3.6 建模时常见的陷阱	43
1.4 小结	7	3.6.1 电子邮件起源问题域	43
第 2 章 关联数据的存储选择	9	3.6.2 敏感的第一个迭代	43
2.1 关系型数据库缺少联系	9	3.6.3 第二次的魅力	46
2.2 NoSQL 数据库也缺少联系	12	3.6.4 发展中的领域	48
2.3 图数据库拥抱联系	15	3.7 避免反模式	52
2.4 小结	19	3.8 小结	53
第 3 章 使用图进行数据建模	21	第 4 章 构建基于图数据库的 应用	55
3.1 模型和目标	21	4.1 数据建模	55
3.2 属性图模型	22	4.1.1 用应用程序所需的 术语描述模型	55
3.3 查询图: Cypher 简介	22	4.1.2 用节点表示事物, 用联系 表示结构	56
3.3.1 Cypher 的理念	23	4.1.3 细粒度联系与通用联系	57
3.3.2 START	24	4.1.4 将事实建模为节点	58
3.3.3 MATCH	25	4.1.5 将复杂的值类型表示为 节点	60
3.3.4 RETURN	25	4.1.6 时间	60
3.3.5 其他 Cypher 语句	25	4.1.7 迭代开发和增量开发	63
3.4 关系建模和图建模对比	26	4.2 应用程序架构	64
3.4.1 系统管理领域中的 关系建模	28	4.2.1 嵌入式与服务器	64
3.4.2 系统管理领域中的 图建模	31		

4.2.2 集群.....	69	5.4 小结.....	122
4.2.3 负载均衡.....	69	第 6 章 图数据库的内部结构.....	123
4.3 测试.....	72	6.1 原生图处理.....	123
4.3.1 数据模型的测试驱动 开发.....	72	6.2 原生图存储.....	125
4.3.2 性能测试.....	78	6.3 用于编程的 API.....	130
4.4 容量规划.....	82	6.3.1 内核 API.....	131
4.4.1 优化条件.....	82	6.3.2 核心 (或 Beans) API.....	131
4.4.2 性能.....	83	6.3.3 遍历 API.....	132
4.4.3 冗余.....	85	6.4 非功能型特性.....	134
4.4.4 负载.....	85	6.4.1 事务.....	134
4.5 小结.....	86	6.4.2 可恢复性.....	135
第 5 章 现实世界中的图.....	87	6.4.3 可用性.....	136
5.1 为什么选择图.....	87	6.4.4 可扩展性.....	137
5.2 常见用例.....	88	6.5 小结.....	140
5.2.1 社交.....	88	第 7 章 使用图论预分析.....	141
5.2.2 推荐.....	89	7.1 深度优先搜索和广度优先 搜索.....	141
5.2.3 地理空间.....	90	7.2 使用 Dijkstra 算法寻找路径.....	142
5.2.4 主数据管理.....	90	7.3 A*算法.....	150
5.2.5 网络和数据中心管理.....	90	7.4 图论和预测建模.....	150
5.2.6 授权和访问控制 (通信).....	91	7.4.1 三元闭包.....	151
5.3 实际示例.....	92	7.4.2 结构平衡.....	152
5.3.1 社交推荐 (专业社交网络).....	92	7.5 局部桥.....	156
5.3.2 授权和访问控制.....	102	7.6 小结.....	158
5.3.3 地理空间 (物流).....	109	附录 NoSQL 概览.....	159

虽然本书大部分内容是讨论图数据模型的，但这并不是一本关于图论的书。^① 使用图数据库并不需要太多的理论知识：只要知道什么是图就够了。记住这一点，下面来大体回顾一下我们对图的认识。

1.1 图是什么

说得正式一点儿，图仅仅是顶点和边的集合，或者说更简单一点儿，图就是一些节点和关联这些节点的联系（relationship）的集合。图将实体表现为节点，实体与其他实体连接的方式表现为联系。我们可以用这个通用的、富有表现力的结构来建模各种场景，从宇宙火箭的建造到道路系统，从食物的供应链及原产地追踪到人们的病历，甚至更多其他的场景。

无处不在的图

在我们了解科学、政府和商业领域的数据集广泛多样性的过程中，图起到了极大的作用。现实世界完全不同于关系型数据库背后的基于表的模型，它是丰富的且相互之间充满关联：有些部分是统一而规则的，而其他部分是特殊的、不规则的。一旦理解了图，你就会发现图无处不在。比如，Gartner 定义了商业世界的 5 个图——社交、意向、消费、兴趣和移动，并指出运用这些图的能力是一个“可持续的竞争优势”。

^① 关于图论的介绍，请参考 Richard J. Trudeau 的《Introduction to Graph Theory》（Dover，1993）和 Gary Chartrand 的《Introductory Graph Theory》（Dover，1985）。如果想要了解图是怎样给复杂的时间和行为提供洞察力的，请参考 David Easley 和 Jon Kleinberg 的《Networks, Crowds, and Markets: Reasoning about a Highly Connected World》（Cambridge University Press，2010）。

就拿 Twitter 来说，它的数据很容易表示为一张图。在图 1-1 中我们可以看到由互相关注的人组成的一个小的社交网络。联系是这里建立语义上下文的关键，也就是说，Billy 关注了（FOLLOWS）Harry，反过来，Harry 也关注了 Billy，Ruth 和 Harry 同样也是互相关注的，但是，尽管 Ruth 关注了 Billy，但 Billy 却对他无动于衷。

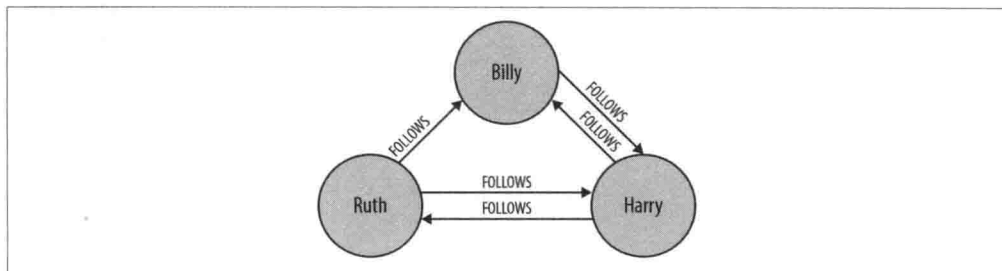


图 1-1 一个小型社交图

当然，实际的 Twitter 图比图 1-1 要大数亿倍，但它们的工作原理是一样的。在图 1-2 中，我们把 Ruth 发布的消息也包含到图里面来。

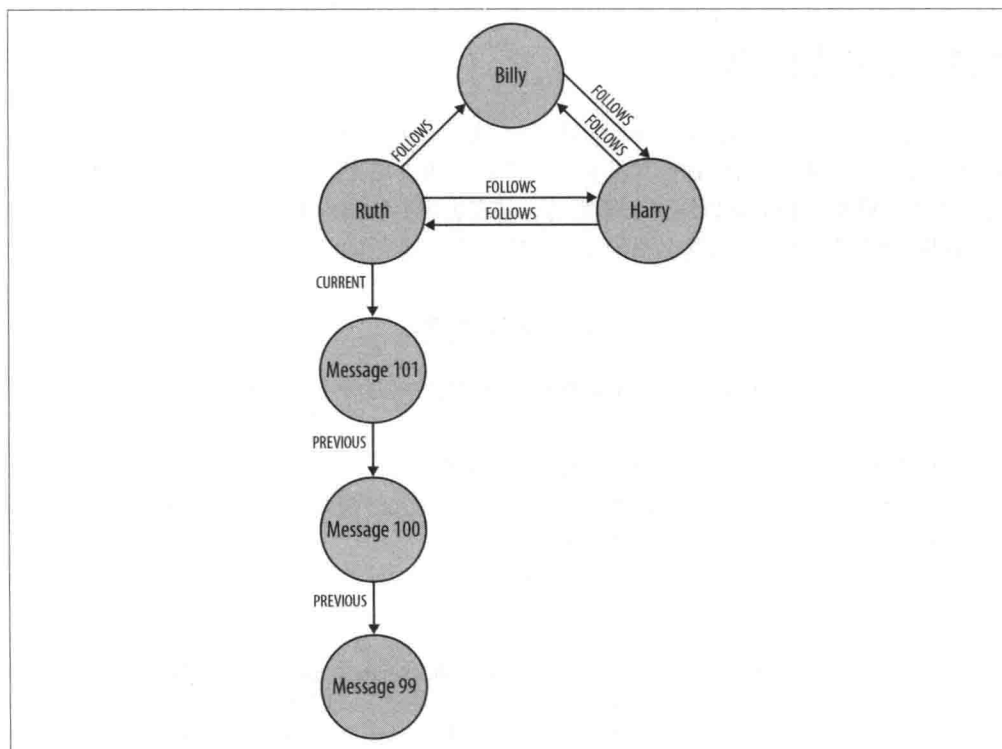


图 1-2 发布消息

尽管图很简单，但图 1-2 还是展示出了图模型的表现力。我们很容易从中看出 Ruth 发布了一连串的消息。通过标记为 CURRENT 的联系可以找到最新的一条消息；PREVIOUS 联系建立了消息时间线。

属性图模型

在讨论图 1-2 的过程中，我们也顺便提一下一个最流行的图模型变体——属性图（property graph）（在附录 A 中我们会更详细地讨论其他可替代的图数据模型）。属性图具有如下特征。

- 它包含节点和联系。
- 节点上有属性（键值对）。
- 联系有名字和方向，并总是有一个开始节点和一个结束节点。
- 联系也可以有属性。

对于大部分人来说，属性图模型是直观且容易理解的。不过简单归简单，使用图将有价值的见解融入到数据中的大多数场景却都可以用属性图来描述。

1.2 图领域概览

近年来，无数用于管理、处理和分析图的项目和产品纷纷涌入市场。技术选择的陡增使我们难以跟进这些工具并摸清它们之间的区别，即便对我们这些一直活跃在这个领域的人来说也是如此。本节的内容对理解新兴的图领域提供了一个“高空俯览”。

从 1 万英尺高空往下看，我们可以将图领域划分成以下两部分。

用于联机事务图的持久化技术通常直接实时地从应用程序中访问

这类技术被称为图数据库，正是本书主要讨论的内容。它们和“通常的”关系型数据库世界中的联机事务处理（online transactional processing, OLTP）数据库是一样的。

用于离线图分析的技术通常都是按照一系列步骤执行的

这类技术被称为图计算引擎。它们可以和其他大数据分析技术看做一类，如数据挖掘和联机分析处理（online analytical processing, OLAP）。



我们可以从另一个视角去划分图领域，去观察各种技术使用的图模型。主流的图模型有 3 种，分别是属性图、资源描述框架（Resource Description Framework, RDF）三元组和超图。我们将会在附录 A 中对它们进行详细的说明。市场上常见的大多数图数据库使用的都是属性图模型，因此，在本书的剩余部分我们也将使用这一模型。

1.2.1 图数据库

图数据库管理系统（以下将简称图数据库）是一种在线的数据库管理系统，它支持对图数据模型的增、删、改、查（CRUD）方法。图数据库一般用于事务（OLTP）系统中。相应地，它们也对事务性能进行了优化，在设计时通常考虑了事务完整性和操作可用性。

在研究图数据库技术时，有两个特性需要多加考虑。

底层存储

一些图数据库使用原生图存储，这类存储是优化过的，并且是专门为了存储和管理图而设计的。不过并不是所有图数据库使用的都是原生图存储，也有一些图数据库会将图数据序列化，然后保存到关系型数据库或面向对象数据库，或是其他通用数据存储中。

处理引擎

一些定义要求图数据库使用免索引邻接，这意味着，关联节点在数据库中物理意义上“指向”彼此^①。如果看得更远一点：站在用户的角度，任何看起来像是图数据库的都可以称为图数据库（比如说，提供了对图数据模型的 CRUD 操作的数据库）。然而，我们得承认这个事实，免索引邻接带来巨大的性能优势是其他数据库无法比拟的，因此我们使用原生图处理来代表使用免索引邻接的图数据库。



需要注意的是，原生图存储和原生图处理并不一定比其他方式更好或更差——这不过是典型的工程取舍而已。原生图存储的好处是，它的栈是专门为性能和扩展性设计建造的。但相对的，非原生图存储通常建立在非常成熟的非图后端（如 MySQL）之上，运维团队对它们的特性烂熟于心。原生图处理（无索引邻接）虽然在遍历查询时性能优势很大，但代价是一些非遍历类的查询会比较困难，而且还要占用巨大的内存。

^① 参考 Rodriguez, M. A. 和 Neubauer, P. 的《The Graph Traversal Pattern》（2010）。

与那些需要额外增加像外键这样的属性或者使用 map-reduce 这样的额外处理来推测实体间关联的数据库管理系统不同，联系在图数据模型中是“一等公民”。图数据库通过将节点和联系的简单抽象组装为相互关联的结构，使我们能够建造任意复杂的模型，来形象地映射我们的问题域。比起那些传统的关系型数据库和其他 NoSQL 存储，我们所得到的模型更简单，也更具表现力。

根据其存储和处理模型不同，图 1-3 生动形象地展示了现在市场上的一些图数据库。



图 1-3 图数据库概览

1.2.2 图计算引擎

图计算引擎技术使我们可以在大数据集上使用全局图算法。图计算引擎主要用于识别数据中的集群，或是回答类似于“在一个社交网络中，平均每个人有多少联系？”这样的问题。

因为偏重于全局查询，图计算引擎通常为扫描和批处理大规模信息做过优化，在这个方面，它们和其他批分析技术（如在关系型数据库世界中大家都很熟悉的数据挖掘和 OLAP）类似。只有一部分图计算引擎有自己的图存储层，其他的（几乎可以说大部分）则只完全关注于如何处理外部传入的数据和返回结果。

图 1-4 展示了一个通用的图计算引擎部署架构。该架构包括一个带有 OLTP 属性的记录系统（SOR）数据库（如 MySQL、Oracle 或 Neo4j），它给应用程序提供服务，请求并响应应用程序在运行中发送过来的查询。每隔一段时间，一个抽取、转换和加载（ETL）