

白皮书

图数据库技术十大案例

互联数据, 无限潜能

Jim Webber, Neo4j首席科学家

目录

介绍	1
案例1: 欺诈检测	2
案例2: 实时推荐引擎	4
案例3: 知识图谱	6
案例4: 反洗钱	8
案例5: 主数据管理	10
案例6: 供应链管理	12
案例7: 增强网络和IT 运营管理能力	14
案例8: 数据谱系	16
案例9: 身份和访问管理	18
案例10: 材料清单	20
结论	21

图数据库技术十大案例

互联数据, 无限潜能

Jim Webber, 首席科学家

介绍

“大数据”每年都在增长, 但如今的企业领导者不仅需要管理更大规模的数据, 还迫切需要从现有数据中得出深刻见解。那么, 首席信息官和首席技术官应该如何获得这些见解呢?

引用Seth Godin的说法, 企业需要摒弃仅仅收集数据点的做法, 开始着手建立数据之间的关联关系。[数据点之间的关系](#)甚至比单个点本身更重要。

讽刺的是, 传统的[关系数据库管理系统\(RDBMS\)](#)并不擅长处理数据之间的关系。那些表状数据模式和呆板的结构难以添加新连接或不同类型连接。为了利用这些数据关系, 您的组织需要一种能将关系信息存储为一等实体的新型数据库技术, 这种技术就是[图数据库](#)。

[图技术就是未来](#)。图数据库不仅能有效存储数据点之间的关系, 而且非常灵活, 适合添加新的关系类型, 并使[数据模型适应](#)新的业务需求。

除了数据库本身, 图技术还涵盖了[图数据科学](#)的新兴领域, 适合预测分析和机器学习, 而所用的[图数据可视化](#)适合针对特定目的的数据发现和探索。

那么, 您的企业应当如何利用图数据技术, 通过连接数据形成具有竞争优势的深刻见解, 成就更高业务价值呢?

以下是图技术的十大使用案例:

案例1: 欺诈检测

银行和保险公司每年因欺诈导致数十亿美元损失。

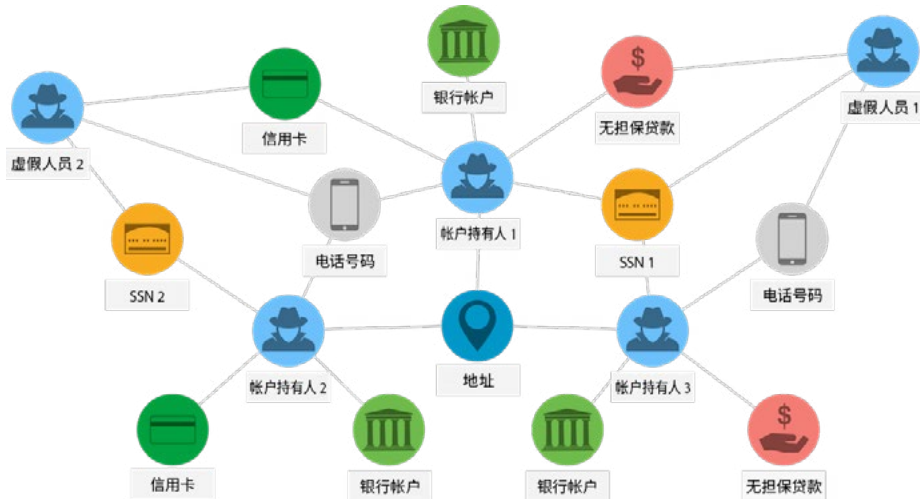
传统的欺诈检测方法往往无法将这些损失降到最低，因为它们执行离散分析，易受误报和漏报影响。越来越老练的欺诈者在认识到这一点后，开发了多种方法来利用离散分析的弱点。

图技术提供了新的方法，通过先进的上下文分析，能够准确[发现欺诈团伙](#)和其他复杂诈骗。这使图数据库能够实时阻止高级欺诈手段。

为什么要使用图技术进行欺诈检测？

虽然没有一个欺诈预防措施是完美的，但是如果您不再只关注单个数据点，而是关注关联数据点的连接时，就可以大大改进效果。

理解数据之间的联系并通过这些链接获得有意义的结论并不需要收集新数据。通过将问题重新构造为图，您就可以从现有数据中获得重要见解。



图与其他查看数据的方法不同，它正是为展现相关性而生的。对于使用传统的表格等表示方式难以检测的模式，图技术可以大显身手。越来越多的公司使用图数据技术来解决各种关联数据问题，[包括欺诈检测](#)。

复杂关联分析

要发现欺诈团伙，您需要遍历计算复杂度很高的数据关系，而随着欺诈团伙的不断壮大，这个问题愈发困难。

侦查和防止欺诈

为了防止欺诈集团，您需要对互联数据集进行实时关联分析，必须覆盖从创建虚假帐户到发生欺诈交易整个时间段。

欺诈团伙不断演进且动态变化

欺诈团伙的形态不断变化、规模不断增长，您的应用程序需要在这种高度动态变化且前所未有的环境中检测这些欺诈模式。

例子: 财富500强金融服务公司

一家**财富500强金融服务公司**收集大量需要进行实时分析的数据以批准交易。虽然自动欺诈检测系统能够立即批准或拒绝其中大部分请求, 但会将潜在的欺诈请求提交给分析员进行手动审查。

分析员拥有专用交易审查工具, 其中整合了所有相关第三方数据。在使用**Neo4j**之前, 分析员必须查询Microsoft SQL Server数据库, 以查看客户的欺诈活动历史记录。

“运行一次查询需要至少5分钟,”该公司欺诈检测解决方案的一位产品经理表示: “而且, 由于我们的分析员每天需要审查上万笔交易, 这种方式难以持续。此外, 关系数据库不是执行关联分析查询的合适解决方案, 因此它给我们的数据库带来了沉重负担。”

该公司需要找到一种更有效的方法来分析数据, 节省客户以及分析员的等待时间。具体来说, 他们需要减少处理欺诈检测查询所需的时间, 并为分析员提供简单的数据可视化功能。

就在此时, 他们发现了Neo4j, 它能够提供关联数据的实时结果, 实现了数据可视化, 使分析员能够做出更快、更准确的决策。这样就使更广泛的搜索成为可能, 公司将搜索深度从4层扩展到10层。

公司分析员也开始注意到数据之间的集群和关系, 从而发现以前没有注意到的全新潜在欺诈联系, 这使得更准确的实时诈骗团伙检测成为可能。

Neo4j能够实现数据可视化, **将分析员的手动审核时间缩短了一半**, 使他们能够更快阻止欺诈交易, 并减少合法客户的等待时间。不仅如此, 他们开始看到更多的可能性, 远远超出其初始的数据可视化用例。

该公司目前正在将Neo4j与他们的实时决策平台集成, 从而可以立即停止欺诈交易, 从而为公司每天节省数千美元。

结论

当涉及到基于图的欺诈检测时, 您需要通过链接分析来增强欺诈检测能力。因此, 有两点很明确:

- 随着业务流程加速、自动化程度提高, 检测欺诈的时间间隔正在缩小, 对实时解决方案的需求愈发迫切。
- 传统技术不适合用来检测复杂的诈骗团伙。图数据库通过分析关联数据点来增加价值。

要发展高效且易于管理的欺诈检测解决方案, 图技术是理想选择。无论面对诈骗集团、勾结团伙还是高知罪犯, 图数据库技术可以实时揭露各种重要诈骗模式。

案例2: 实时推荐引擎



无论您的企业是在零售、服务、媒体还是社交领域运营，为您的用户提供具有极强针对性的实时建议对于实现客户价值最大化和保持竞争力至关重要。与其他业务数据不同，这种建议必须能够进行归纳并和上下文关联，最终用户才会认可其相关性。

通过图技术，您可以将客户的浏览行为和人口统计特征与购买历史相结合，立即分析他们当前的选择，并马上[提供相关建议](#)，而且是赶在潜在客户点击竞争对手网站之前提供所有这些建议。

那么，为什么要使用图技术来增强实时推荐引擎？

实现实时推荐的关键技术是图数据库，在关联大量买家和产品数据（或统称关联数据）方面，它超越了其他数据库技术。

要作出有效的实时建议，数据平台必须了解实体之间的关系，以及这些关联的质量和强度。只有图技术能够根据用户购买、交互和评论有效跟踪这些关系，以提供对客户需求和产品趋势最有意义的深刻见解。

图驱动推荐引擎通常采用两种主要方法：识别个人感兴趣的资源，或识别可能对具体资源感兴趣的个人。无论采用哪种方法，图数据库技术都会进行必要的关联和连接，以便为相关的个人或资源提供最相关的结果。

大量的数据和关系

协同过滤和基于内容的过滤等目前流行的推荐算法依赖快速遍历不断增长和高度互联的数据集。

实时相关推荐

提供实时上下文洞察是一项挑战。因此，推荐系统的强大之处在于它能够利用当前数据实时提出建议。

新的数据和关系

添加更多数据点的同时，建议的准确性和范围也会增加。这种快速增长意味着推荐系统需要同时满足当前和未来的需求。

例子: eBay

电子商务巨头eBay一直谋求[改进购物者搜索物品的方式](#)。然而, eBay高级副总裁兼首席产品官RJ Pittman表示, 现有的产品搜索和推荐目前无法在购物请求中提供或推断上下文信息。

eBay的目标是建立一个实时推荐引擎, 它能够理解并学习购物者提供的上下文语言, 并快速聚焦具体产品推荐。具体来说, 要构建谷歌助手eBay应用, 需要将[知识图谱](#)与自然语言理解(NLU)和人工智能相结合, 以存储、记忆过去与购物者的交互并从中学习。

eBay选择Neo4j作为[原生图数据库](#), 该数据库包含概率模型, 有助于理解会话购物场景。Neo4j图数据库包含产品目录和顾客在寻找产品时的交互属性。

此外, 要将聊天机器人部署到其用户群, 需要在互联网规模中提供高度弹性和可用性、毫秒级的可预测响应以及在部署类型方面有经验的图专家的支持。Neo4j具有高可用性集群能力和优异的读写性能, 即使有数百万个节点也能应对自如, 它对此类用户请求的响应能力也很强。

由此产生的应用程序包括Neo4j图数据库和NLU, 它们不仅可以理解语音, 还可以在解析这些会话的意义和上下文时包含拼写和语法意图。

此应用程序当前正在云中的Docker容器内运行。接下来, eBay团队希望通过Slack和Microsoft等插件跨多个平台部署聊天机器人。

结论

实时推荐引擎为零售、物流、招聘、媒体、情感分析、搜索和知识管理等领域的企业提供了关键的差异化功能。通过图数据技术存储和查询推荐数据, 应用程序能够[提供实时结果, 而不是提供预先计算的过时数据](#)。

随着消费者期望值提高, 他们的耐心降低, 提供相关的实时建议将成为比以往任何时候都更强大的竞争优势。

案例3: 知识图谱



使用传统的基于关键词的搜索时, 获得的结果随机性强、效果有限、质量不高。您无法真正提出更精确、更有用的问题, 并获得最相关、最有意义的信息。

同样, 面对未来变化, 关系数据库也不灵活: 如果要添加新类型的内容或进行结构更改, 则必须重新修改关系模型。

为什么要将图技术用于知识图谱?

标准的搜索方法是一个繁琐的过程, 需要反复重新定义您的搜索词, 直到最终找到感兴趣的东西。随着用户逐步习惯[能够理解用户意图的上下文搜索](#), 上述方式的局限性就非常明显。

此外, 还要应对未来型搜索:

- 语音搜索意味着我们进行询问, 而设备做出反应。
- 无处不在的搜索功能将出现在新兴应用程序和更多的设备中。
- 付费搜索将需要更深入的产品知识和购买力数据。

为了增强企业搜索能力, 您需要一个[具有基于图的搜索功能的知识图谱](#), 从而只提供相关的上下文结果。

例子: 美国宇航局

美国宇航局(NASA)首席知识架构师David Meza需要一种[更有效的方法来检验过去50年太空探索中各种经验教训之间的数据关系](#)。

他说, “我们一直在使用标准的键列表搜索方法”, 这种做法局限性明显, 因为他们的信息存储在许多分散和相互独立的系统中。

不断增长的数据资产规模

数据集每秒都在扩大。如果缺乏能够处理不断扩展的数据和关系的有效数据存储, 搜索将始终受到限制。

精确搜索查询

任何无法深入上下文(通过精确定位用户历史、意图和上下文实现)的搜索解决方案都无法保持竞争优势。

消除搜索查询的歧义

随着数据量的增加, 重叠但不相关的搜索结果也会增加。搜索质量从根本上会受到大量结果的影响。

David和他的工程师及科学家团队无法通过关系模型关联各个数据点，从而导致项目非常耗时、成本高昂。不仅如此，这些项目的一个共同目标是找到模式并解决问题，以避免将来重蹈覆辙。

大卫说：“我们遇到了‘大数据’问题，不仅要处理结构化数据，还要处理非结构化数据，我们还在不断收集更多的数据。”“我需要一种产品或应用程序，用于跨越并发展数百万甚至数十亿个节点，以极快的速度连接这些信息。”

David和他的团队发现[Neo4j正是他们梦寐以求的可扩展工具](#)，可以用来查看数以百万计的节点。

他说：“以前，这些信息存储在分散和相互独立的系统中，现在我们将这些点联系起来，并揭示多年来的趋势，了解我们在空间站进行的实验可能会对哪些系统和子系统产生影响。”

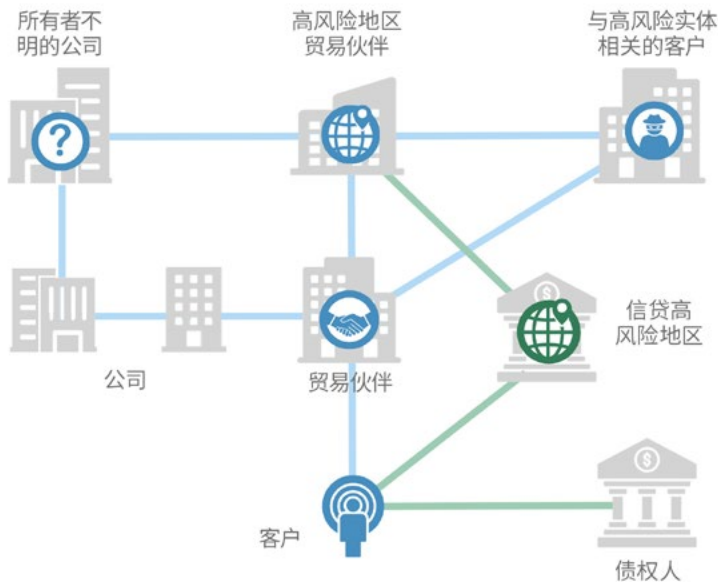
在美国宇航局的“经验教训”项目中，使用Neo4j发现了一些猎户座飞船和阿波罗飞船时代的相关信息，从而避免了一个严重问题，节省了超过两年的工作时间和100万美元的纳税人资金。

结论

当用户得到最相关搜索结果（可能是意外结果）时，参与度和满意度都会更高。但是，如果没有关系上下文和元数据，搜索将无法提供用户所需的精确结果。

随着数据资产的大量增长，您需要一个知识图谱来容纳数据集的内在关系。通过使用图数据库，您可以提高信息访问能力，用户和客户可以从中找到他们最需要的产品、服务或数字资产。

案例4: 反洗钱



如今的反洗钱(AML)方案日趋复杂,往往涉及繁复的间接操作,以甄别误导和掩盖欺诈的可疑活动。然而,传统技术的设计初衷并不是为了跨越许多中间步骤,连接那些中间步骤。

检查员通常会花费大量时间仔细研究大量数据,常需要数月,导致日常待审查交易严重堆积。

为什么要使用图技术来反洗钱?

许多传统技术的设计并不能将洗钱行为中的关键点连接起来,因此检测洗钱的方法需要大量的艰苦努力。检查员小组需要手工检查大量数据,工作负担沉重。

而如今出现了各种各样的洗钱策略:

- 结构化洗钱(又名拆分洗钱)
- 大宗现金走私
- 现金密集型业务
- 依托贸易洗钱
- 空壳公司和信托
- 往返交易
- 银行股权收购
- 赌场和其他赌博
- 房地产
- 现金工资
- 人寿保险业务

误报

误报警报继续激增,动摇了人们对数据准确性的信心,并使金融业白白付出数十亿美元的无谓努力。

监管格局

监管规范要求改进监测和分析。数据报告透明度对于避免违规处罚和起诉至关重要。

传统解决方案

就打击洗钱而言,时间就是一切。目前的反洗钱流程降低了生产力,并存在效率障碍。

随着数据资产的大量增长，能够适应数据集内大规模多样化关联的技术解决方案就成为迫切需求。

[图数据库能够识别复杂的数据关系，并具有实时查询能力](#)，这是对抗洗钱和非法资金挪用的强大武器。

例子: 转账服务

作为转账服务公司的合规经理，他很清楚他们需要一种软件工具，让反洗钱调查人员能够看到汇款的模式，这种模式会影响到数十万笔日常交易。

这家每年转移近6000亿美元资金的转账服务公司需要一种方法来检测“拆分洗钱”活动，这种活动非常猖獗，其方式是将大额非法资金分配给一个隐藏的受益人网络。

因此，国际货币转移行业受到高度监管，公司在全球150个国家运营，必须遵守每个国家的反洗钱(AML)要求。

“我们一直在寻求改进我们的系统和流程，”这位合规经理表示：“我们研究了FIU（欧洲金融情报部门）的地图投射技术，立即发现这种工具对我们业务可能具有潜力。”

他们说：“当我们看到Neo4j时，我们知道它是[最好的工具，因为它提供了一种动态的方式来查看数据](#)。我们（传统的）工具都是基于SQL技术的，在调查过程中，无法获得Neo4j能实现的动态方法。”

只需点击几下，Neo4j就可以动态、实时处理该公司的大量数据，使这家资金转账公司在合规和调查过程中都能获得更高效率。

合规经理说，当处理一起跨国和至少两大洲的重大案件时：“Neo4j让整个情况一览无余……怀疑的关键因素是过程，当使用Neo4j进行查看，流动过程显而易见。”

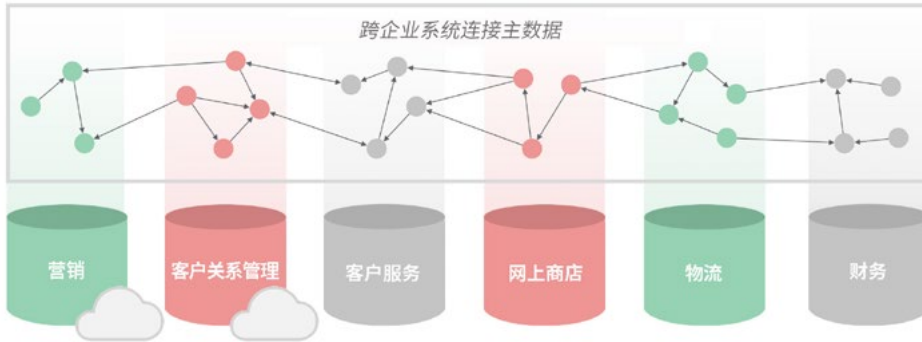
自从全面部署Neo4j以来，该公司能够以比传统工具快20倍的速度追查刑事案件，而且他们已经“确定了行业内调查合规性的基准”。

结论

使用图数据库，您可以通过采用新数据源和新类型，不断改进AML检测，且无需重写[数据模型](#)。内置的高可用性功能确保您的任务关键型AML引擎始终可以访问用户数据。

与关系数据库不同，图存储相互关联的数据，使检测欺诈活动不受数据的深度或形状的影响，更加便捷。

案例5: 主数据管理



主数据是企业的命脉, 包括以下数据:

- 用户
- 客户
- 产品
- 账户
- 合作伙伴
- 地点
- 业务单位

许多业务应用程序使用主数据。主数据经常保存在许多不同位置, 存在大量重叠和冗余, 采用不同格式, 质量参差不齐。主数据管理(MDM)是识别、清理、存储和管理(最为重要)此类数据的操作。

MDM最佳实践各异, 有的将所有主数据合并到一个位置, 有的则通过部署一个服务或应用对数据资产进行管理并实现方便的访问。在这两种情况下(或任何混合解决方案), 企业数据架构师都需要[数据模型能够随业务需求变化而提供任务特定的、可变的和特殊的结构](#)。这种快速发展的模型最适合图数据库。

为什么使用图技术来进行主数据管理?

由于主数据是高度关联和共享的, 如果MDM系统的构建方式不佳, 会降低业务灵活性, 并波及整个企业。大多数传统MDM系统依赖一个关系数据库, 这种数据库并未针对遍历关系或快速响应进行优化。

随着业务分析的发展, 主数据集内的这些数据连接和关系对于竞争优势至关重要。好消息是, 图数据库是建模、存储和查询主数据中的层级、元数据和连接的理想选择。

与构建关系型解决方案相比, 使用图数据库, 您的主数据更易于建模, 所需资源(建模人员、架构师、DBA和开发人员)更少。此外, 您不必将所有主数据迁移到一个位置。图关系可以轻松关联CRM系统、库存系统、会计和销售点系统之间的孤立数据, 提供企业数据的统一视图。

分层的复杂数据集

组织数据和产品数据等主数据具有自上而下、横向和对角连接的深层层次结构。使用关系数据库管理这样的数据模型需要臃肿庞大的复杂代码, 导致运行缓慢、构建成本高且维护时间长。

实时存储和查询性能

主数据存储必须与企业内的大量应用程序集成并提供数据, 有时需要实时进行。然而, 通过遍历复杂且高度互联的数据集来提供实时信息是一项严峻挑战。

动态结构

主数据本质上是动态的, 这使得开发人员很难设计出适应其发展的系统。

例子: 爱彼迎

在[大型、复杂的组织 \(如爱彼迎\(Airbnb\)\)](#)中, 不断增长的内部和外部数据资源 (尤其是分散在各种平台上的数据资源) 最终变得难以管理和限制。

爱彼迎软件工程师John Bodley认识到爱彼迎的数据极其孤立、难以访问且缺乏适当的上下文。他还注意到, 员工依赖局部知识来回答问题, 这大大损害了生产力。

“我们经常进行员工调查,”他说:“对‘我工作所需的信息很容易找到’这个问题, 我们的得分一直很低”。Bodley的团队知道他们需要使数据真正容易被所有员工获取, 这样, 无论员工的数据处理水平如何, 他们都能够找到资源, 并可以完全相信结果相关且可靠。

他的团队开始开发Dataportal, 这是一个自助服务的集成数据空间, 为员工提供爱彼迎数据的上下文、整体视图, 以便他们在需要访问或响应日常工作需求时导航数据环境。

Bodley和他的团队很快意识到他们的整个数据生态系统最好用一个图来表示。他们由此注意到Neo4j图数据库, 这种数据库提供了每秒数百万个数据连接的最快搜索方式。Neo4j还能够与爱彼迎首选的编程语言完美整合, 同时还允许它们利用图拓扑来丰富搜索排名。

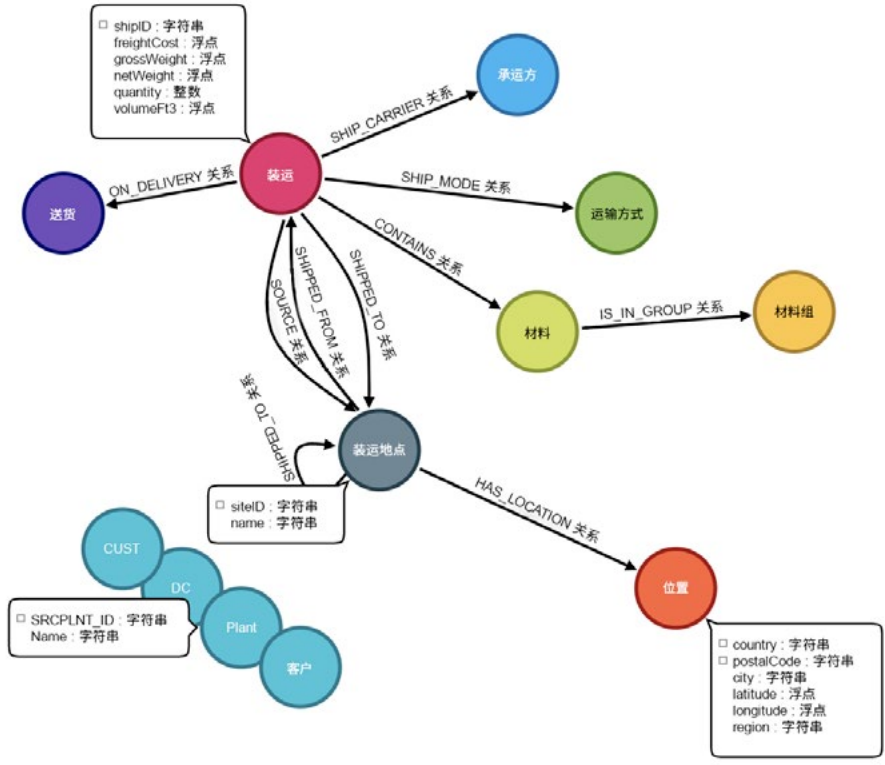
有了Neo4j, 爱彼迎能够将他们的整个数据生态系统联系起来, 使其具有可搜索性、相关性和可信赖性。爱彼迎遍布各地的员工不再需要依靠局部知识资源, 而是将Dataportal作为寻找相关数据的一站式资源, 尤其是对日常绩效至关重要的以员工和团队为中心的信息, 这种改变至关重要。

结论

您无法依靠陈旧、分散和相互独立的信息系统做出由数据驱动的理想业务决策, 所以需要包含有关数据关系信息的实时主数据。

图数据库就是为支持这些关系应运而生的。通过更高效的建模和查询, [将主数据组织在图数据库中](#), 您能够以前所未有的效率和灵活性获得相关答案。

案例6: 供应链管理



物流路线, “最后一英里”

需要更深入了解阻碍物流过程中最后一段的所有挑战、难点和成本。

透明度和可靠性

仅估算与每个新客户相关的时间和成本难以满足任何一方的需求。

简化配送网络

由于有一系列的仓库、工厂和配送中心,它们之间的相互关系面临许多挑战。

确定未来需求

市场需求瞬息万变,如果没有大数据解决方案,就无法真正洞察客户、供应商和传感器以及外部因素的信息。

从本质上而言,供应链管理是动态的,活动部分众多,瓶颈可能在任何位置出现。挑战在于,传统数据库生成的数据卷和详细信息缺乏实时、准确的信息处理能力。

如今,大数据正在推动许多行业甚至所有行业快速发展,在其中,供应链管理是数据驱动型见解、预测分析和生产力保障的完美实例。由于数据只会增加,跟踪和保持控制的挑战将越来越大。

为什么要使用图技术进行供应链管理?

根据定义,由于主数据是高度共享的,这种争夺往往会波及整个组织,降低业务敏捷性。我们的架构一直以来专注于获取数据,更符合事实单一性定义,然而我们大多数人都意识到,从长远来看,这不是一个可行的解决方案。

供应链管理数据段

- 分配网络资源以满足需求
- 时间和运输成本的路线规划
- 通过客户管理使他们再次光顾
- 营销以获得新客户并推动收入增长
- 潜在事故和非常事件的风险分析
- 维修工作、翻新和资产购买
- 配送中心和仓库管理,以适当分配容量和可用空间

例子: Transparency-One

如今的供应链庞大而广泛, 导致风险无处不在。这使得透明度变得更加重要和更复杂。

认识到这个问题, [Transparency-One开发了一个平台](#), 允许制造商和品牌所有者了解、监控、分析和搜索他们的供应链, 并共享有关生产地点和产品的重要数据。

2013年之后, 该公司认为有必要扩大这一数字地图的范围, 将供应链中所有要素的更详细信息(产品、供应链映射等)纳入其中。

Transparency-One首席执行官Chris Morrison表示: “因为这一领域对我们的客户来说是全新领域, 而且没有任何市场解决方案能提供真正的完全透明化, 所以我们面临的挑战尤为巨大”。

为了开发解决方案, Transparency One最初转向了经典的SQL数据库类型解决方案。但是, 它很快意识到, 要处理信息的数量和结构将对性能产生重大影响, 带来严重问题。

因此, Transparency-One团队开始研究图数据库。

Morrison表示: “当我们决定建立一个图数据库时, 我们研究了市场领导者使用的这类数据库。” “答案非常明确: Neo4j是这一领域的世界领先者, 久负盛名。”

Transparency One之所以选择Neo4j, 是因为只有图数据库才能满足平台的要求; Neo4j管理大量数据的能力非常突出, 而且是世界上使用最广泛的此类数据库, 用户小到初创公司, 大到巨型企业, 规模各异, 这些成为决定因素。

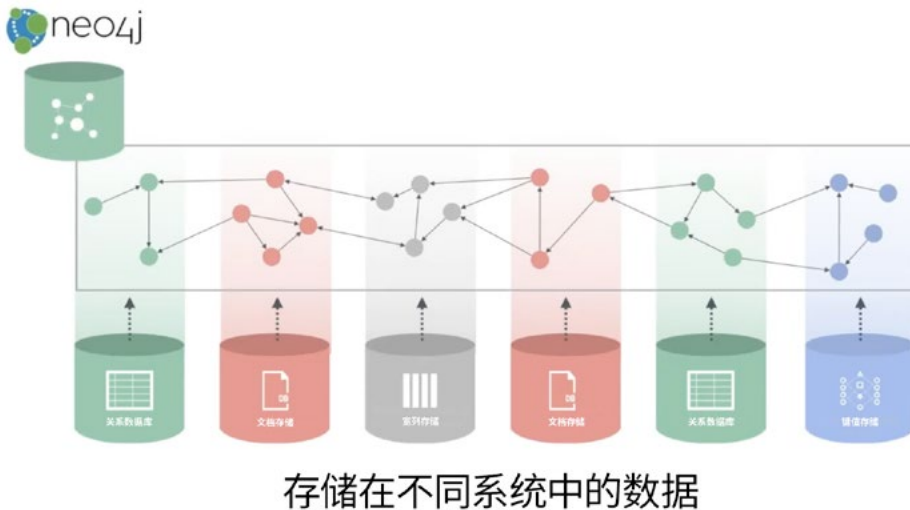
这个项目很快就投入运行, 不到三个月就建成了一个原型。此后, Transparency-One通过添加新模块扩展了解决方案, 目前正由多家公司部署。

结论

传统的关系系统有自身优点, 但是穿越主数据网络非常昂贵、缓慢、缺乏敏捷度。

使用图数据库, 您可以实现[360度视图的全局存储, 从而有效管理和利用供应链管理数据关联](#)。数据源随时间变化时, 图模型更易演化, 从而实现灵活性能且易于扩展, 无论数据集规模有多大, 都能实时决策。

案例7: 增强网络和IT运营管理能力



本质上说, 网络是图。因此, 图数据库非常适合于[建模、存储和查询网络和IT操作数据](#), 不论受您的业务运行在防火墙的哪一边。

如今, 图数据库已成功应用于电信、网络管理、影响分析、云平台管理、数据中心和IT资产管理等领域。

在所有这些领域中, 图数据库存储配置信息。该系统可以实时提醒操作人员注意基础设施中的潜在设计缺陷和共享故障模式, 并将问题分析和解决时间从数小时减少到数秒。

为什么要在网络和IT操作中使用图技术?

与主数据一样, 图数据库用于汇集来自不同库存系统的信息, 提供网络及其使用者的单一视图, 涵盖范围从最小的网络元素一直到使用它们的应用程序、服务和客户。

网络的图表示使IT管理人员能够对资产进行编目, 实现资产部署可视化, 并识别两者之间的依赖关系。图的连接结构使网络管理员能够进行复杂的影响分析, 回答以下问题:

- 在应用程序、服务、虚拟机、物理机、数据中心、路由器、交换机和光纤等网络部分中, 哪些部分是特定客户所依赖的? (自上向下分析)
- 反之, 如果某个特定的网络元素发生故障, 网络中的哪些应用程序、服务和客户将受到影响? (自下而上分析)
- 对于最重要的客户, 整个网络是否有冗余?

故障排除

无论您在管理重大网络更改、强化与安全相关的访问, 还是在优化应用程序基础结构, 实体设备和人员的相互依赖关系都是极其复杂的, 这使得故障排除非常困难。

影响分析

网络中各个节点之间的关系既不是纯粹的线性关系也不是层级关系, 因此很难确定网络元素之间的相互依赖关系。当两个或多个系统结合在一起时, 这些关系变得更加复杂。

不断增长的实体和虚拟节点

随着网络规模的快速增长和支持新网络服务和设备的元素飞速增加, 您的IT组织必须开发同时满足当前和未来需求的系统。

网络的图数据库表示也可以丰富基于事件关联的操作智能化。每当事件关联引擎（如[复杂事件处理器](#)）从低级网络事件流推断出一个复杂事件时，它就会评估该事件对图模型的影响，并触发任何必要的补偿或缓解操作。

例子: Comcast

全球公司在家居智能化方面已经取得长足进步。

如果家门未上锁、房屋报警器断开或有人在家门口时设备就可以发出警报。这些连接的设备可以轻松相互交互，但通常缺乏自然语言处理等功能，这就使大量关联设备无法实现自动化。

“一个人并不是一个身份证明(ID)，人是一组与个人信息、位置、人物和设备的关系。”Comcast项目管理总监Jessica Lowing说。

现今家居系统无法执行像“关掉莉莉房间里的灯”这样的任务，因为这些要求需要深入理解复杂的语义和社会关系。个性化同样受到很大限制。为了应对这些挑战，Comcast组建了一个团队，全力创造和完善xFi智能家居标准。

第一步是为Comcast资料图表中的所有术语制定丰富的定义。

例如，“人”的定义不仅需要唯一ID，它必须包含一组与个人信息、位置、人物和设备的独特关系。

“既然人是这些智能家居的中心，也理应成为我们自动化体系的中心，”Lowing说，“这样我们就要重新考虑社会和语义关系的建模。”

团队认识到，这些丰富数据的真正价值在于它们之间的关系，这需要一个本地图形数据库结构。最终，该团队还需要在家庭层面建立一个共享平台。

这就产生了Xfinity资料图表，这是一种能够灵活扩展的多租户用户资料服务，用于跨越各种Xfinity产品扩展个人信息和关系。它对客户的实际生活关系进行建模，并提供上下文，推动Xfinity应用为用户提供更个性化的体验。

结论

发现、捕获和理解复杂的相互依赖关系是有效运行网络和IT操作的关键，这对于经营企业至关重要。无论是优化网络、优化应用程序基础设施，还是提供更高效的安全相关的访问，这些问题都涉及到一组复杂的物理设备和人的相互依赖关系，这对管理相当具有挑战性。

网络和基础设施要素之间的关系很少是线性的或纯粹的层次关系。但是，图数据库正是为了存储这些相互关联的数据而生的，这使得[将网络和IT数据转换为可操作的深刻见解](#)变得非常容易。

案例8: 数据谱系

风险建模是需求的多重组合。它还要求各组织（特别是大型银行、对冲基金和激进的投资机构）在投资、持股、金融工具和精细定价数据的复杂网络中跟踪数据连接。

为什么对数据谱系使用图技术？

银行和其他受监管的金融机构（如贵公司）在获得原始的权威数据源之前，需要在许多复杂的层面上跟踪数据相关性，而大多数系统无法满足这一关键基本要求。BCBS 239等严苛要求给数据谱系带来了挑战，相比传统数据存储系统，这需要系统更加灵活、更为持久。

您需要通过离散的数据系统向后[追溯数据，直到其依赖过程以最终的权威源结束](#)。同样，尽管实体需要标准标识符，不同业务仍然会常常使用自己的术语和算法，有时在同一组织中也是如此。

数据的结构和位置常常使在单一、集中的数据存储中寻址极为困难。具有讽刺意味的是，将所有内容移动到单个存储库中可能会使跟踪数据谱系变得更加困难。

将信息集成到企业范围内的单一逻辑数据模型中需要图技术。使用图数据库，您可以得到单一真实数据源，并在几毫秒内查询和显示复杂或隐藏的数据连接。

例子: 瑞银

瑞银的解决方案架构师Sid Hannif为了[让客户使用体验与最新的实时数据无缝对接](#)，他需要加入数百个系统和孤立的数据。

他说：“我们应该提供及时、准确和完整的数据，之前消费者必须自己加入数据”。

瑞银成立于1854年，公司现有员工约6万人，分为财富管理、投资银行、资产管理、企业中心和数据服务等部门。

在数据服务部门中，Sid和他的团队负责12个数据域和子数据集，其中的数据通过多个通道发送出去。帐户数据保存在关系数据库中。整合的数量庞大，构成严重障碍，瑞银客户参考数据分发平台因此受到影响。

数据服务团队希望使用元数据让引擎查询数据的形成方式以及与其他数据关联的方式，然后进行整合。这恰恰是知识图谱作为系统的组件可以实现的。

正因如此，瑞银选择了Neo4j的图数据库来为他们的数据建模。

Sid说：“针对数据迁移这样的数据治理工作大大得到简化，我们了解谁在获取我们的数据，以及通过什么渠道获取数据。但更重要的是，我们的消费者能够了解我们提供的数据集以及它们与其他数据集的关系。”

治理和基础设施

银行必须利用数据治理和集成数据分类，以及集团范围的元数据，包括实体、交易对手、客户和账户的标识符。

风险数据汇总

银行必须生成准确、一致和可靠的风险数据，同时保持权威数据源完全可见。这些数据集必须满足管理者和监管者的各种要求。

风险报告

管理和监管报告必须以精确和可审计的方式反映风险，并能适应银行风险模型和运营的复杂程度。

监督审查

银行风险监管机构必须审查其机构对BCBS 239规范的持续遵守情况。

结论

相比传统系统,使用符合隐私和风险报告法规的图数据库有以下优势:

- 追踪风险因素的沿袭,追溯到它们最初的权威数据来源
- 将定价、持仓、现金管理和其他离散的数据系统扩展到一个统一的数据集中
- 与监管机构合作,实现风险模型图可视化并进行修改
- 能够轻松修改风险模型,以跟上不断变化的市场条件、组织变革和投资策略
- 处理影响交易部门历史及未来运营和业绩的合并、剥离和重组事务

Neo4j图数据库还支持依靠专业服务的全球金融术语标准,这些服务可确保成功,并使您更明确了解合规工作和日常运营。

案例9: 身份和访问管理

身份和访问管理(IAM)解决方案存储有关各方(例如管理员、业务部门、最终用户)和资源(例如文件、共享、网络设备、产品、协议)的信息,以及管理对这些资源访问的规则。IAM解决方案应用这些规则来确定谁可以访问或操作资源。

传统上,身份和访问管理通过使用目录服务或通过应用程序后端构建自定义解决方案来实现。然而,层次目录结构无法处理多方分布式供应链中的复杂依赖结构。使用非图数据库存储标识和访问数据的自定义解决方案随着其数据集的增大而变得缓慢且难以响应。

为什么使用图技术进行身份和访问管理?

图数据库可以存储复杂的、紧密关联的访问控制结构,这些结构跨越数十亿个参与方和资源。它丰富多变的结构化数据模式支持层次结构和非层次结构,而其可扩展的属性模式允许捕获系统中每个元素的丰富元数据。

有了一个每秒可以遍历数百万个关系的查询引擎,大型复杂结构的图数据库访问查找将在数毫秒内执行,而不必耗费几分钟或几小时。

与网络和IT操作一样,图数据库访问控制解决方案允许自上而下和自下而上查询:

- 特定管理员可以管理哪些资源(公司结构、产品、服务、协议和最终用户)?(自上而下)
- 给定特定资源,谁可以修改其访问设置?(自下而上)
- 最终用户可以访问哪些资源?

图数据库支持的访问控制和授权解决方案特别适合内容管理、联合授权服务、社交网络个人资料和软件即服务(SaaS)产品等领域,相比早先关系数据库几分钟处理时间,新数据库将性能大幅提高到几毫秒。

例子: 挪威电信

挪威电信是一家国际通信服务公司。几年来,它已经为其规模居前的商业客户提供了自助服务帐户的能力。使用基于浏览器的应用程序,每个客户组织中的管理员都可以代表其员工添加和删除服务。

为了确保用户和管理员只查看和更改有权管理的组织部分和服务,应用程序使用了一个复杂的身份和访问管理系统,该系统将权限分配给数千万个产品和服务实例中的数百万个用户。

内在互联的权限数据

为了准确验证身份及其访问权限,系统需要遍历高度互连的数据集,而数据集的规模和复杂性不断增加。

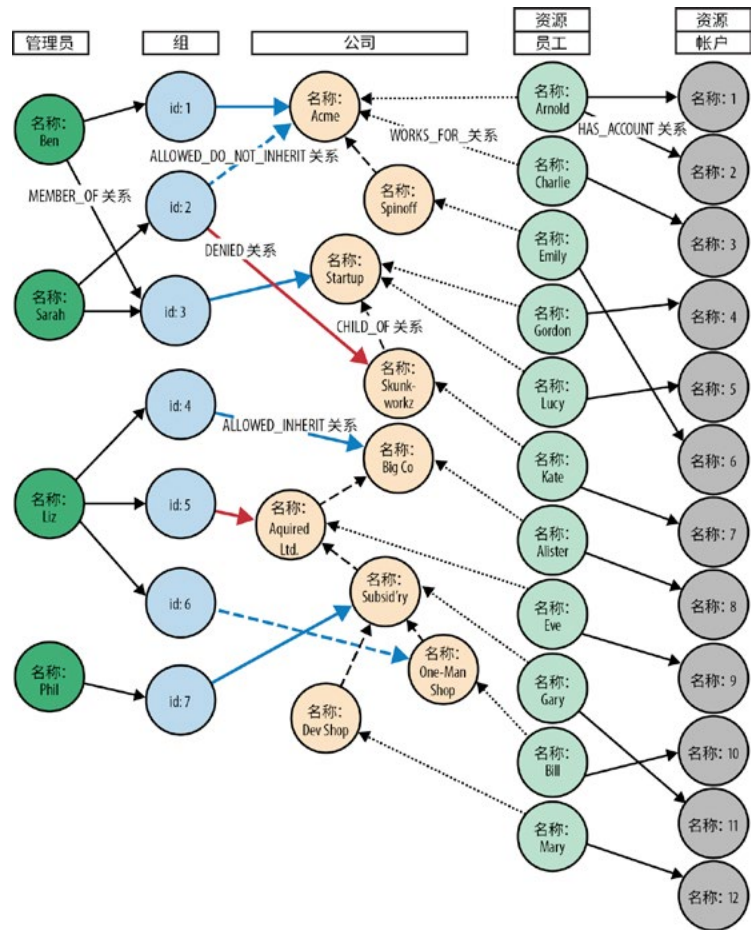
生产力和客户满意度

随着用户、产品和权限的增长,传统系统无法继续提供能够及时响应的查询性能,从而损害了用户体验,无法让用户满意。

动态结构与环境

随着用户规模及其相关元数据的快速增长,您的应用程序需要同时满足当前和未来的身份管理需求。

下面是挪威电信数据模型的一个例子。



说明：挪威电信公司的数据模型示例，展示其身份和访问管理应用(IAM)。

由于性能和响应问题，挪威电信决定用图数据库解决方案取代现有的IAM系统。他们最初的系统使用关系数据库，该数据库采用递归JOIN来对复杂的组织结构和产品层次结构进行建模。由于连接密集型模型，最重要查询的执行速度慢得令人无法接受。

相比之下，在实施图数据库解决方案后，挪威电信就实现了处理其身份和访问管理需求所需的性能、可伸缩性和适应性，将查询时间从几分钟减少到几毫秒。

结论

对于企业组织来说，管理多个不断变化的角色、业务组、产品和授权是一项越来越复杂的任务。由于关系数据库查询太慢而且难以响应，根本不能胜任管理您的身份和访问需求的任务。

使用图数据库，您可以无缝实时跟踪所有的身份和访问关系。通过数据的互联视图，您比以往任何时候都拥有更好的洞察力和控制能力。

案例10: 材料清单

物料清单(BOM)流程并未因数字化转型时代而消亡,它仍然是制造流程的关键所在,但是新的制造数据无处不在,难于管理。

由于电子表格和孤立的数据缺乏与分散人员(承包商、供应商)的实时协作,且面临多专业(机械、机电)产品以及全新设计-制造工作流程的挑战,已经成为障碍。

为什么对物料清单使用图技术?

大多数企业制造商使用供应商应用程序:CRM系统、工作管理系统、应付账款、应收账款、销售点系统等。由于这种方式,您需要将数据存储和建模为一个图,其中原生图存储的[互连主数据](#)既不是纯线性结构,也不是层级结构。

同样,新兴的BOM趋势也带来更多层的数据挑战。合规要求更严格。仿冒零件充斥市场。零部件制造商之间正在进行整合,零部件的生命周期也越来越短。与此同时,技术创新已经证明会增加可用性风险,同时也会导致过时。

而采用灵活的图数据模型,可以在需求随时间变化时方便进行主数据调整。

案例: 日本汽车制造商

车辆测试对于汽车零部件OEM生产商而言是一项耗时但必不可少的过程。在开始大规模生产之前,需要收集和整理测试数据,以便对故障进行排除和修复。

一家大型汽车制造商长期以来一直试图让关键测试信息能够对“长期产品验证管理(PVM)”产生价值。然而,制造商的产品验证生命周期无法正常工作,因为非标准化的车辆测试数据无法有效识别问题的根本原因,更不用说解决问题了。

由于来自不同领域(甚至即使在同一领域内)的工程师以不同的方式进行测试,并使用不同的工具以多种格式存储测试数据。结果,数据变得不一致且彼此隔离,从而使其对其他团队毫无用处。

汽车制造商与iASYS合作,决定在Neo4j中构建知识图谱,并使用BRIX PVM解决方案整合整个产品验证生命周期。iASYS团队使用Neo4j设计了一个知识图谱,该知识图谱连接了所有产品验证生命周期数据,并允许在企业范围内集成各种邻域和功能。他们还在知识图谱中创建了工程词典,其中定义了关键元数据,例如测试类型、测量特征和测量条件。

然后,iASYS将制造商的工程领域映射到新的标准化验证过程中。同时,BRIX PVM为测试、子测试和测量提供了定义明确的语义,使工程师能够跨域和平台进行沟通。

一个问题是领域之间并不总是有明确的界限。更多时候,领域彼此重叠。这意味着所有团队中的信息都是内在联系的,必须一起工作。

例如,在公路测试中观察到的振动问题可能是由于引擎或车辆结构造成的。为了确定问题的根源,工程师需要与跨学科的不同团队合作。在这种情况下,车辆测试小组和动力总成组分享记录的振动数据,然后该动力组将检查负载状况、燃烧数据以及有关发动机扭矩的时间序列ECU数据,通过知识图谱把这些数据关联起来,这样就非常容易找到主要振动源。

物料清单挑战

产品可制造性

保持物料清单与产品开发周期同步。

多专业物料清单

制造业现在已经不再局限于机械零件,还包括电子零件、软件等。

设计 workflow

无法满足全球可追溯和审计的要求,缺乏用于库存管理或按订单设计流程的集中目录。

结论

这十个图数据库的案例远称不上详尽，但确实展现了图技术影响深远、收益明显的应用情况。即便如此，图技术还有很多其他的案例，包括[生命科学](#)、[社交网络](#)、[游戏](#)、[政府](#)、[体育](#)甚至[非盈利组织](#)。

如今的首席信息官和首席技术官面临着越来越大的压力，在数据集越来越大和处理越来越困难的情况下，他们仍然需要从自己的大数据中提供可操作的见解。他们需要技术方案能够确定数据点之间的联系并得出令人信服的合理结论。

图数据库就是这种技术解决方案。数据库允许各个级别的数据专业人员超越单个数据点，挖掘数据关系的潜力。如何利用这些关系，完全取决于数据库用户的想象力。

由此，我们不难得出这样一个结论：图数据库不是过时的时尚，而是这一充满大数据洞见时代的新趋势。

关于Neo4j:

Neo4j是全球图数据库技术的领导者。作为全球部署最为广泛的图数据库，我们帮助了包括：Comcast（康卡斯特）、NASA（美国国家航空航天局）、UBS（瑞银）及Volvo Cars（沃尔沃）在内的国际品牌预测并揭示了人、流程、系统之间如何进行关联和联系的模式和结构。使用这种以关系优先的方法，基于Neo4j构建的应用程序可以解决大数据带来的挑战，例如人工智能与分析、欺诈检测、实时推荐及知识图谱。更多详情，请见neo4j.com官网。

有关于Neo4j的疑问吗？

无论您在全球任何地方，
都可以与我们联系：

china@neo4j.com

neo4j.com/contact-us