

# 图数据库算法和分析高级应用实战

讲师 俞方桦

## 【课程概要】

### 1. 让你从零开始认识一种全新的数据表示、建模和分析方法

这是一门从理论基础开始、针对大数据的特征和挑战、全面讲解图计算及其应用的深度培训课程。图数据库是近年来兴起的一种针对大数据及其内在联系进行高效、复杂、发现型分析的重要方法。图数据库的领先厂商，Neo4j，仅仅在过去五年中，其产品和技术就在欺诈检测、实时推荐、主数据管理、网络管理、知识图谱等众多领域得到广泛和成功的应用，行业用户覆盖财富一百强的七成。

图计算针对数据及其之间的关系，其重要方法在网络管理、搜索引擎、社交人群分析、公共安全、舆情分析等众多领域已有成熟的理论框架、方法和实践，而且随着大数据时代的到来更是每天都有创新型的应用。掌握图计算和分析方法无疑是在大数据和人工智能时代推动企业不断创新、挖掘数据价值、以及提升个人职业发展空间的一个新机遇。

### 2. 基于使用最广泛的图数据库开源项目

本次课程将使用图数据库领域应用最广泛、最成功的 Neo4j 社区版，全面、完整和深入地讲解其原生图数据库的设计理念、架构、图数据查询语言 Cypher 和图算法扩展库 ALGO 的使用。

### 3. 强化实践环节、针对实际业务数据和应用场景

本次培训，不仅会讲述图的数学和计算机科学理论基础，还会在 Neo4j 数据库平台上、基于样例数据进行丰富的案例研究和实操。

#### 4. 持续学习、不断提高

为期 2 天的深度培训，已经能够让学员掌握图计算和核心内容、以及 Neo4j 图数据库的实际应用经验。在随后，我们会在 ITPub 上持续推出新的培训课程，并在 Neo4j 中文社区、QQ 和微信用户群以及其他线上、线下媒体上为大家提供交流、学习的渠道内容。

### 【课程简介】

当今的时代是大数据的时代。大数据为政府和企业的信息化带来了新的机遇，同时也为数据存储、处理、分析和挖掘提出了新的挑战。图论和图数据库正是顺应这一机遇和挑战而兴起的创新的数据处理和分析方法。

Neo4j 是全球领先的图数据库平台。Neo4j 的原生图数据库引擎、和 Cypher 查询语使之在处理非结构化、大规模、关联型数据方面拥有传统关系型数据库所无法比拟的高性能、灵活性和可扩展性优势。

本课程从图论基础出发，结合大数据的特征全面介绍图的基本分析方法以及在不同大数据领域的应用实践。实战环节则以 Neo4j 图数据库为工具，介绍其架构、Cypher 查询语言以及算法包，通过现场操作和实际案例展示图数据的查询和分析方法、技术和最佳实践。

#### 整套课程大体上分为四个部分：

1. **图理论、技术和应用**，以及在大数据时代的重要作用。介绍数学和计算机科学中“图”的理论起源、内容和经典应用；通过对生活中常见的四个图的分析 and 讲解，让学员快速建立对“图问题”的认识、以及图分析基本方法的了解；通过对大数据的四个 V 的解释揭示大数据的特征，以及图数据分析如何与之有机结合。

2. **Neo4j 原生图数据库及其应用开发基础**。包括 Neo4j 的架构、Cypher 查询语言和常用操作；学员在此过程中不仅会对声明型的图查询语言有基本掌握，还能实现对数据的导

入、映射和建模。

**3. 经典图算法及其应用**：深入介绍并结合实际操作对图计算的重要算法全面和深入地讲解：包括图特征计算、路径寻找、社区检测、中心性算法、相似度检测。课堂上将基于样例数据、就各个算法的含义、适用领域、执行方法以及对结果的解释进行全面和细致的讲解和实操。

**4. 实践环节和讨论**：在一个有百万节点、千万关系的社交网络数据集上运用学到的算法进行分析、并对得到的结果进行阐释和讨论。

### 【课程对象】

面向企业信息化主管、数据架构师、设计建模工程师、开发人员和数据科学家，分享 Neo4j 图数据库的基础概念、Cypher 查询语言及其使用、图数据查询分析应用实例、查询优化；理论结合实践，帮助学员建立基于图数据库的建模设计和实践开发经验，在大数据分析应用领域具备最具创新、最前沿的技术能力。

**根据讲师的经验与授课体系，有意报名的参课者最好有如下的一些准备或者经历：**

1. 拥有关系型数据库的使用和开发经验(中等以上、至少 3 年)、了解 SQL 以及数据建模的基本方法；
2. 对常用的数据分析技术、方法和工具有一定了解和使用经验；
3. 扎实的数学和计算机科学领域基础知识，包括离散数学、算法、数据结构、图论等；
4. 有 Neo4j 的使用和开发经验更佳。

### 【课程收益】

1. 掌握图理论和图数据的主要内容、特征、应用；
2. 掌握大数据时代数据分析的要求、趋势、方法和技术；
3. 掌握 Neo4j 图数据库的重要内容：安装、配置、启动、导入数据以及安装扩展包等；
4. 了解并能熟悉 Cypher 图数据查询语言、实现 Neo4j 图数据库中节点和关系的增、删、改、查；
5. 掌握图计算的五大类重要算法，以及如何在 Neo4j 中运行和使用它们；
6. 掌握对实际(开源)数据集、运用相应算法进行图计算和分析、并对结果进行诠释和讨论的实际能力。

### 【课程大纲】

序号	课程时间	课程名称	课程提纲
<b>5 月 11 日</b>			
1-1	30 分钟	《什么是图？》	<ul style="list-style-type: none"> <li>- 什么是图？</li> <li>- 数学中的“图”</li> <li>- 计算机科学中的“图”</li> <li>- 图的类型和研究领域</li> </ul>
1-2	60 分钟	《为什么图是大数据分析的重要方法》	<ul style="list-style-type: none"> <li>- 现实世界中图无处不在</li> <li>- 图的实例 1：社交网络</li> <li>- 图的实例 2：金融网络</li> <li>- 图的实例 3：基因网络</li> <li>- 图的实例 4：知识网络</li> <li>- 数据库技术的演变和 NoSQL 的兴起</li> </ul>
1-3	30 分钟	《图和大数据的特征》	<ul style="list-style-type: none"> <li>- 大数据的四个 V</li> <li>- 图分析的挑战：规模，复杂度，建模的考虑</li> <li>- 图数据处理平台</li> </ul>
1-4	70 分钟	《图的分析方法》	<ul style="list-style-type: none"> <li>- 图的特征：连通性、模块化、演变</li> <li>- 路径分析</li> </ul>

			<ul style="list-style-type: none"> <li>- 连接度分析</li> <li>- 社区检测</li> <li>- 中心性计算</li> <li>- 相似度分析</li> </ul>
1-5	40 分钟	《图数据库概述》	<ul style="list-style-type: none"> <li>- 什么是图数据库？</li> <li>- 图数据的模式：属性图和 RDF</li> <li>- 关系型数据库和图数据库的比较：数据库模式，存储模式，查询类型，性能比较，应用类别 OLTP vs OLAP</li> <li>- 图数据库的优势</li> <li>- 图数据库的主要应用场景</li> <li>- Neo4j 图数据库平台介绍</li> <li>- 图数据库比较：原生 vs 混合，属性图 vs RDF，</li> </ul>
1-6	40 分钟	《Cypher 图查询语言基础》	<ul style="list-style-type: none"> <li>- Neo4j 的图数据库模式(LPG)</li> <li>- Cypher 查询语言介绍</li> <li>- 创建节点和关系</li> <li>- 查询节点和关系</li> <li>- 更改节点标签、属性</li> <li>- 定义限制和索引</li> <li>- 删除节点和关系</li> </ul>
1-7	40 分钟	《Neo4j 图数据分析：导入数据》 (含实战操作)	<ul style="list-style-type: none"> <li>- 数据集介绍</li> <li>- 导入数据</li> </ul>
<b>5 月 12 日</b>			
2-1	50 分钟	《Neo4j 图数据分析（一）：基本图分析》 (含实战操作)	<ul style="list-style-type: none"> <li>- 基本图分析</li> </ul>
2-2	80 分钟	《Neo4j 图数据分析（二）：路径分析(Path Finding)》 (包含在小型测试数据集上的实战)	<ul style="list-style-type: none"> <li>- 广度优先搜索</li> <li>- 深度优先搜索</li> <li>- 最短路径</li> <li>- 全图最短路径</li> <li>- 单一源最短路径</li> <li>- 最小生成树</li> <li>- 基于成本的最短路径</li> <li>- 随机漫步算法</li> </ul>

			- 小结
2-2	50 分钟	《Neo4j 图数据分析（三）：中心性算法(Centrality)》 (包含在小型测试数据集上的实战)	- 维度中心性 - 紧密中心性 - 间接中心性 - 页面排行 - 小结
2-3	70 分钟	《Neo4j 图数据分析（四）：社区检测算法 (Community Detection)》 (包含在小型测试数据集上的实战)	- 三角计数和聚类系数 - 强连通分量 - 连通分量 - 标签传播算法 - Louvian 模块度算法 - 小结
8	50 分钟	《Neo4j 图数据分析（五）：相似度算法(Similarity)》 (包含在小型测试数据集上的实战)	- Jaccard 相似度 - 余弦相似度 - 欧几里得相似度 - 重叠相似度 - 小结
7	30 分钟	问答时间和总结	

### 【讲师简介】



俞方桦

博士，英文名 Joshua，Neo4j 亚太地区售前和技术支持主管

Neo4j 亚太地区技术专家，IEEE 和澳大利亚计算机学会会员。二十多年 IT 从业历史，职责包括开发、架构、咨询顾问、技术管理，行业经验跨越政府、零售、金融、科技、制造，专注领域涉及大数据、商业智能、AI 和机器学习、云计算、软件开发以及儿童编程教育。

俞方桦现定居悉尼。业余时间喜爱读书、旅游、美食、网球和任何新颖的科技产品。他的

联系方式：<https://www.linkedin.com/in/joshuayu>。

**【讲师资质】**

2004 PMP – 项目管理专家

2007 (澳大利亚)金融市场高级证书

2016 AWS 认证架构师

2017 Neo4j 认证技术专家

**附录：**

**ITPUB 学院**是为企业和个人提供企业应用方案培训咨询（包括企业内训）、个人实战技能培训（包括认证培训）在内的全方位 IT 技术培训咨询服务新平台，隶属于盛拓传媒 IT168 企业级平台。

IT168 企业级平台包括 ITPUB、ChinaUnix 两大社区；中国系统架构师大会（SACC）、中国数据库技术大会（DTCC）两大会议品牌。其中，ITPUB 目前是中国专业的数据库技术社区，ChinaUnix 是国内专业的 Linux/Unix 系统、开源软件技术社区。中国系统架构师大会、中国数据库技术大会是每届规模超过 1500 人的技术培训大会。

ITPUB 学院旨在提供更多精彩课程，详情请查看官方网站 <http://edu.itpub.net>

联系人：黄老师

电话：13699173494

邮箱：[edu@itpub.net](mailto:edu@itpub.net)

微信：18500940168

