

# 图形数据库 NEO4J 与关系数据库的比较研究

王余蓝

(西安交通大学 外国语学院, 陕西 西安 710049)

**摘要:** 为了更为全面地揭示图形数据库与关系数据库在内部机理与应用场景方面的不同, 采用理论分析与实例验证相结合的方法, 从模型成熟度、安全性、可扩展性 3 个方面进行分析对比, 结果显示图形数据库更新非常方便, 更新复杂度仅为关系数据库的 1/4。由此可知, 图形数据库在成熟度、安全性等方面劣于关系数据库, 但在处理复杂数据关联方面远优于关系数据库, 适合存储关联关系复杂、关系动态变化等社交性数据。

**关键词:** 图形数据库; 关系数据库; 扩展性; NEO4J

**中图分类号:** TN919-34

**文献标识码:** A

**文章编号:** 1004-373X(2012)20-0077-03

## Comparison of graphic database NEO4J and relational database

WANG Yu-Lan

(School of Foreign Languages, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** In order to reveal the difference of graphics database and relational database, theoretical analysis and experimental validation were carried out in three aspects: maturity, security and scalability. The experimental results show that the complexity of relational database is 1/4 of graphical database. This means that graphical database is inferior to relational database in maturity and security, but it is superior in dealing with the complex association of data, and suitable for storage of social networks, academic networks and other inter-connecting data.

**Keywords:** graphic database; relation database; expansibility; NEO4J

## 0 引言

在现代社会中, 对信息的管理已经变得越来越重要, 如交通信息<sup>[1]</sup>、文献检索<sup>[2]</sup>、金融信息<sup>[3]</sup>等, 都需要处理大量的数据。数据库技术已成为信息系统的核心和基础。在数据库技术发展过程中, 出现过众多的数据模型, 比较常用的有 3 种<sup>[4]</sup>, 分别为层次模型、图模型和关系模型。关系模型建立在严格的数学基础上, 具有较高的数据独立性和安全性, 使用简单。关系数据库是目前应用最为广泛的数据技术。但是随着数据规模的膨胀与数据复杂性的增加, 关系模型已经无法满足领域需要, 以社交网络<sup>[5-7]</sup>为例, 采用关系数据库将导致数据冗余, 并且不能适应社交数据的动态性, 也不能很好地支持类似“好友的好友”这样的多层复杂查询。针对数据间内在关系复杂且动态变化的问题, 人们再次将目光转向图形数据库<sup>[8]</sup>, 图形数据库能够有效的存储、管理、更新数据及其内在关系, 并能高效执行多层复杂操作<sup>[9]</sup>。

在实际应用中选择何种数据库与应用需求紧密相关, 下面将以图形数据库 Neo4j 为例从产品成熟度、模型安全、可扩展三个方面与关系数据库加以比较, 并为实际应用中的数据库选择提供一些建议和帮助。

## 1 图形数据库

图形数据库就是将数据存储在有向图 (Graph) 结构中<sup>[2-3]</sup>。如图 1 所示是一个简单的有向无环图。其中, 节点表示一个实体。例如人或商品。边表示点与点之间的连接关系, 可以是有方向和无向的。如用户 A 买了商品 B 表示  $A \rightarrow B$ ; 如果用户 A 与用户 C 相互都认识, 这种关系就是双向的, 表示为  $A \leftrightarrow C$ 。属性表示点和边所附带的属性。例如用户姓名、年龄等。需要注意的是每个点或边的属性是动态可变的。

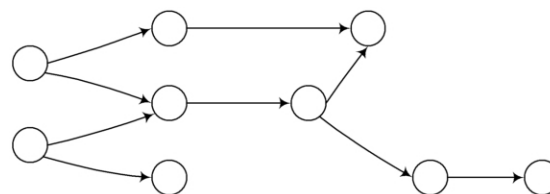


图 1 一个 DAG 例子

图形数据库可以看作是结点与关系的集合<sup>[10]</sup>, 图形数据库就是将数据存储在有属性的结点中, 并用关

收稿日期: 2012-06-11

基金项目: 国家自然科学基金资助项目 (61100166); 陕西省教育厅科技专项 (11JK1035); 西安交通大学智能网络与网络安全教育部重点实验室开放课题 (1030326)

系将这些结点组织起来,如图2所示。

数据存储的重要目的是为了检索。图的查找与搜索可以通过遍历算法完成,根据算法,从开始结点到与之相连的结点查询诸如“我好友的好友是哪些人”等问题。所以通过遍历算法可以对图进行导航与操作,从而确定结点之间的路径。如图3所示。

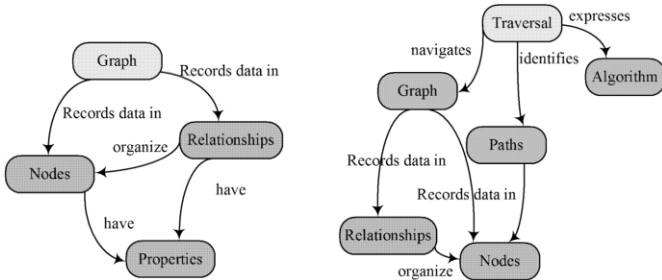


图2 图模型

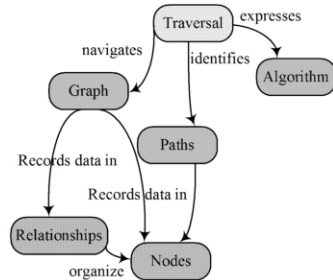


图3 图的遍历

如图4所示,通过建立索引,可以更快、更高效地查找某个结点。通常情况下,可能只想通过属性去查找一个确定的结点或关系,而不是遍历整图。在这种情况下,就可以通过索引来查找某个结点,如“根据用户名定位用户节点”等。

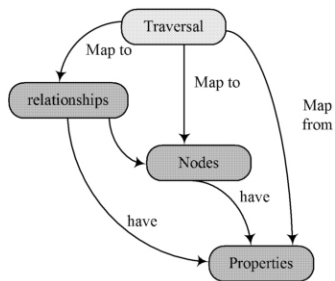


图4 图的索引

把点、边(关系)、属性联系到一起就能描述出一个图。图5展示了一张作者合作关系社交图谱(该图基于MDLayout改进算法生成<sup>[4]</sup>)。每个作者代表一个点,边用白色实线表示,表明了两者间关系。边也可以拥有属性。很多个点关联起来就构成了一个很复杂的学术合作网。使用图形数据库很容易存放这种人际关系网,而传统的关系数据库则不能最优化的存储上述社会关系数据。

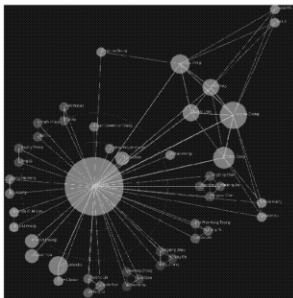


图5 学术合作图谱示例

## 2 图形数据库 NEO4J 与关系数据库比较

### 2.1 理论与技术成熟度

关系数据库理论经过几十年的发展已经日趋完善,相比较图形数据库来说,关系数据库的理论体系更加坚实,而且关系数据库也有相当成熟的实现,例如 Oracle 与 SQL Server 等。关系数据库使用 SQL 语言进行数据查询,SQL 语言的标准比较规范,各种数据库系统都大同小异。Neo4j 研究时间较短<sup>[11]</sup>,尚没有得到广泛的应用,而且 Neo4j 的数据查询语言并不统一,目前就有 SPARQL, Gremlin, Cypher Query 等多种规范。因此目前 Neo4j 的成熟度不高,还在不断的发展与完善中。

### 2.2 安全性

关系数据库普遍支持多用户并发,然而 Neo4j 不具任何内置的安全管理机制来限制与管理多个用户,它假定系统处于一个受信任的温室环境中,而把安全访问控制发在应用程序层处理。关系数据库具有密钥、权限、分层等多种安全保障机制,在安全性方面 Neo4j 相比关系数据库来说有待进一步提高。

### 2.3 可扩展性

虽然在成熟度和安全性上,关系数据库比 Neo4j 图形数据库更加优秀,但关系数据库严格的模式约束使得对已有数据库的扩展变得非常困难,下面通过一个示例来比较 Neo4j 与关系数据库在可扩展性上的优劣。假如给定以下信息:

Marko is a human and Fluffy is a dog.  
Marko and Fluffy are good friend.  
Human and dog are subclass of mammal.

在关系数据库中,这些信息存储方式如表1~表3所示。

表1 数据对象表

ID	NAME	TYPE
001	Marko	Human
002	Fluffy	Dog

表2 朋友关系表

ID1	ID2
001	002
002	001

表3 类型分类表

TYPE1	TYPE2
Human	Mammal
Dog	Manmal

而在 Neo4j 图形数据库中,这些信息可以组织成图的形式,如图6所示。

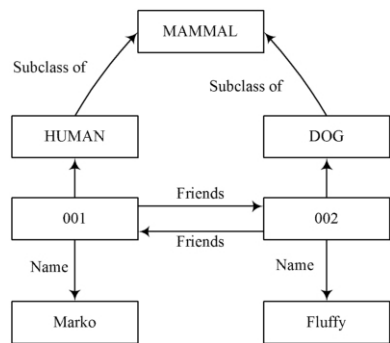


图 6 图形数据库实现

现在如果再增加一些信息,如:Marko and Fuffy are both mammals. 对于关系数据库来说,为了能存储增加的信息必须重构上面的数据表,如表 4~表 7 所示:

表 4 数据对象表

ID	NAME	TYPE
001	Marko	<del>Human</del>
002	Fluffy	<del>Dog</del>

表 5 朋友关系表

ID1	ID2
001	002
002	001

表 6 类型分类表

TYPE1	TYPE2
Human	Mammal
Dog	Manmal

表 7 对象类型表

ID	TYPE
001	Human
002	Dog
001	Mammal
002	Mammal

然而在图形数据库 Neo4j 中,则不需要重构整个数据库,仅仅需要动态增加几个结点和表示结点间关系的边而已,如图 7 所示。

3 结 语

与图形数据库相比,关系模型数据库存在四点缺陷:对象关系的不匹配使得把面向对象的“圆的对象”挤到面向关系的“方的表”中;关系模型静态、刚性、不灵活的本质使得改变 schemas 以满足不断变化的业务需求非常困难;关系模型很不适合表达半结构化的数据——而业界的分析家和研究者都认为半结构化数据是信息管理中的下一个重点;关系模型可以表达面向网络的数

据,但是在遍历网络并抽取信息的能力上关系模型非常弱。通过以上分析比较,可以得出结论:当数据量较小且数据对象间关联关系固定时,关系数据库可以很好的工作,然而当数据规模庞大,数据对象间的关系复杂且动态变化时,图形数据库则更为合适。

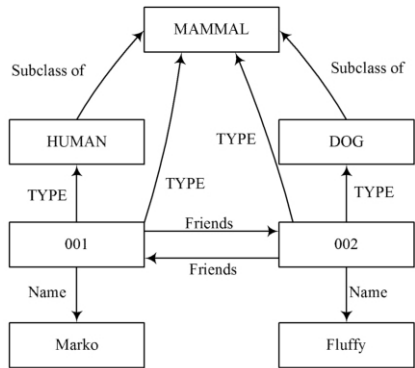


图 7 改进后的图形数据库

参 考 文 献

[1] 仰燕兰,叶桦,费树岷. 面向关系数据库的道路识别算法研究[J]. 地理信息世界,2011(5):58-62.

[2] 刘焱,贾君枝. 中文信息处理中的语义关系表示探析[J]. 现代图书情报技术,2006(10):25-29.

[3] 李平,文凤华. 计算金融的应用前景分析[J]. 时代经贸, 2010(23):90-91.

[4] HULL R, KING R. Semantic database modeling: survey, applications, and research issues [J]. ACM Computing Surveys (CSUR), 1987, 19(3): 201-260.

[5] 乔秀全,杨春,李晓峰,等. 社交网络服务中一种基于用户上下文的信任度计算方法[J]. 计算机学报,2011,34(12): 2404-2411.

[6] 付晓燕. 社交网络服务对使用者社会资本的影响:社会资本视角下的 SNS 使用行为分析[J]. 情报杂志,2010(4): 10-15.

[7] 于婷婷,窦光华. 社交网络服务兴起的社会学意义[J]. 当代传播,2011(6):11-15.

[8] R. Angles, C. Gutierrez. Survey of graph database models [J], ACM Computing Surveys (CSUR), 2008, 40(1): 1-6.

[9] 陈锐. 网络结构与数据库模式在陕西科技信息网中的应用[J]. 情报杂志,2004,23(10):3-7.

[10] 姚玉斌,叶爽利,吴志良,等. 稀疏矩阵法网络拓扑分析[J]. 电力系统保护与控制,2011,39(23):1-5.

[11] Neo4j Org. The Neo4j Manual v1.7 [DB/OL]. [2012-04-18]. [http:// docs. neo4j. org/ chunked/ 1. 7- SNAPSHOT](http://docs.neo4j.org/chunked/1.7-SNAPSHOT).

[12] PAPADOPOULOS C, VOGLIS C. Drawing graphs using modular decomposition [C]// Proc. of 13th International Symposium on Graph Drawing (GD). Heidelberg, Berlin: Springer-Verlag, 2005:343-354.

作者简介:王余蓝 女,西安交通大学外国语学院。研究方向为信息系统、实验室管理。