

大数据时代的新型数据库 — 图数据库 Neo4j 的应用

微云数聚（北京）科技有限公司

zhizh@we-yun.com

内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵

内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵



什么是图数据库：理论

图数据库是基于数学里图论的思想和算法而实现的高效处理复杂关系网络的新型数据库系统。



什么是图数据库：擅长

图形数据库善于高效处理大量的、复杂的、互连的、多变的数据。其计算效率远远高于传统的关系型数据库。



什么是图数据库：应用

图数据库在**社交网络、实时推荐、征信系统、人工智能**等领域有着广泛的应用。



什么是图数据库：元素

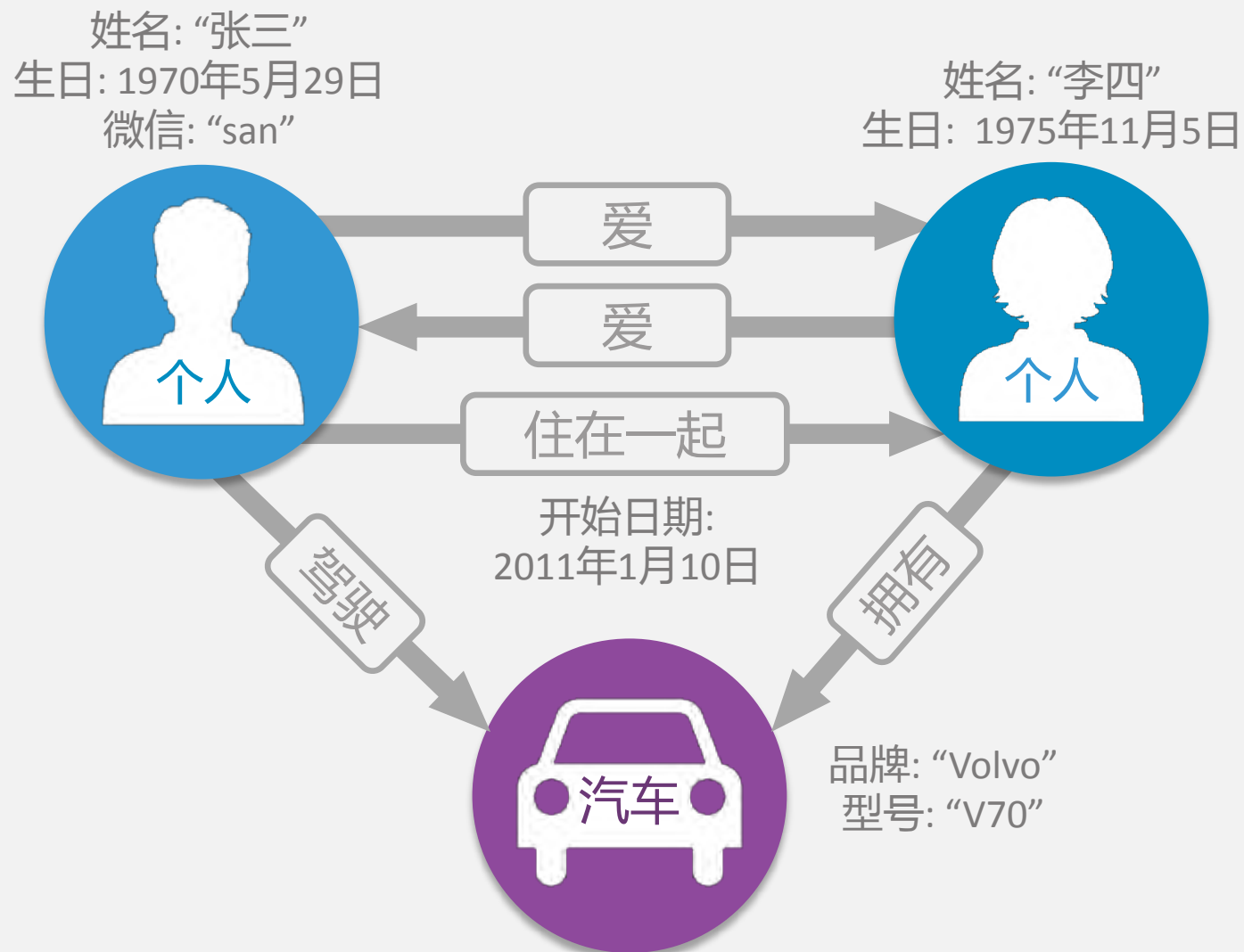
图中每个**节点**代表一个对象，节点之间的连线代表对象之间的**关系**。节点可带**标签**。节点和关系都可以带若干**属性**。

节点

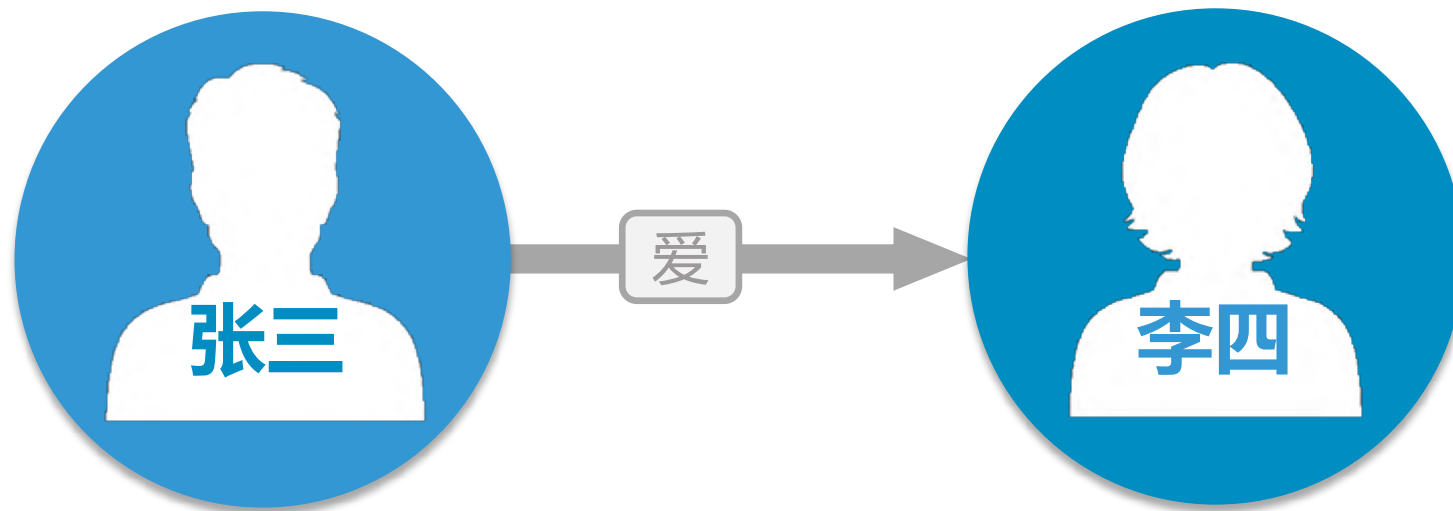
- 图中的对象
- 可带若干名-值属性
- 可带标签

关系

- 连接节点（有类型、带方向）
- 可带若干名-值属性



图的基础：用 Cypher 创建节点、关系



节点

节点

```
CREATE (:个人 { 姓名:"张三" }) -[:爱]-> (:个人 { 姓名:"李四" })
```

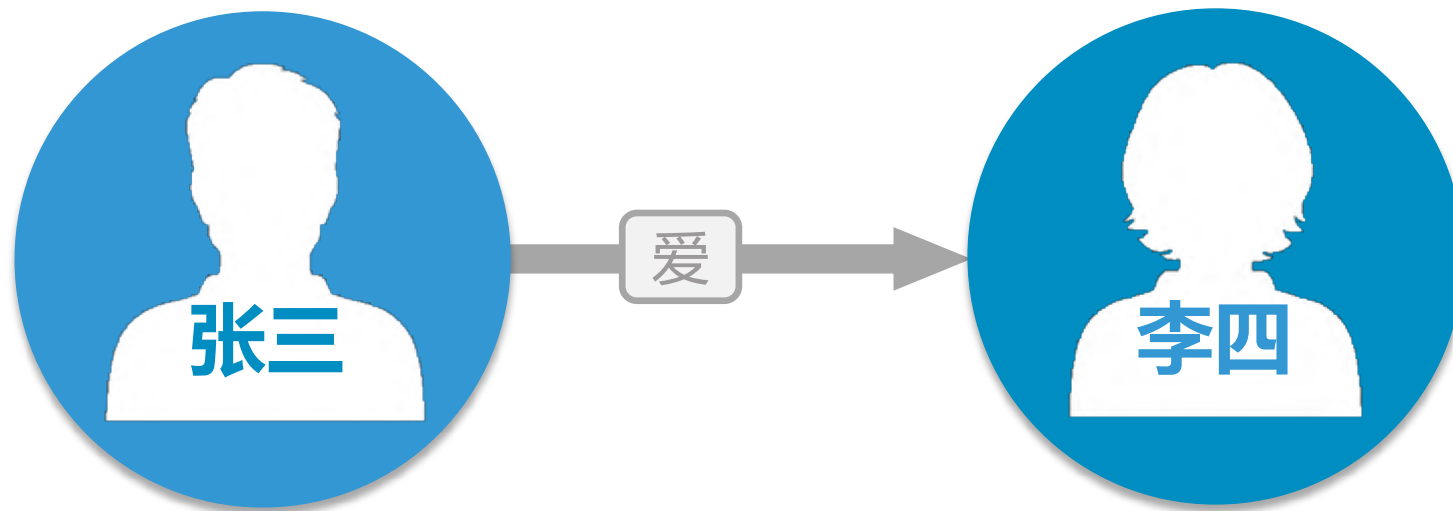
标签

属性

标签

属性

图的基础：用 Cypher 查询



节点

节点

```
MATCH (:个人 { 姓名:"张三" }) -[:爱]-> (:个人 { 姓名:"李四" })
```

标签

属性

标签

属性

内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

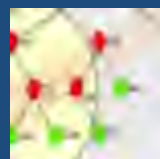
图片化

数据驱动

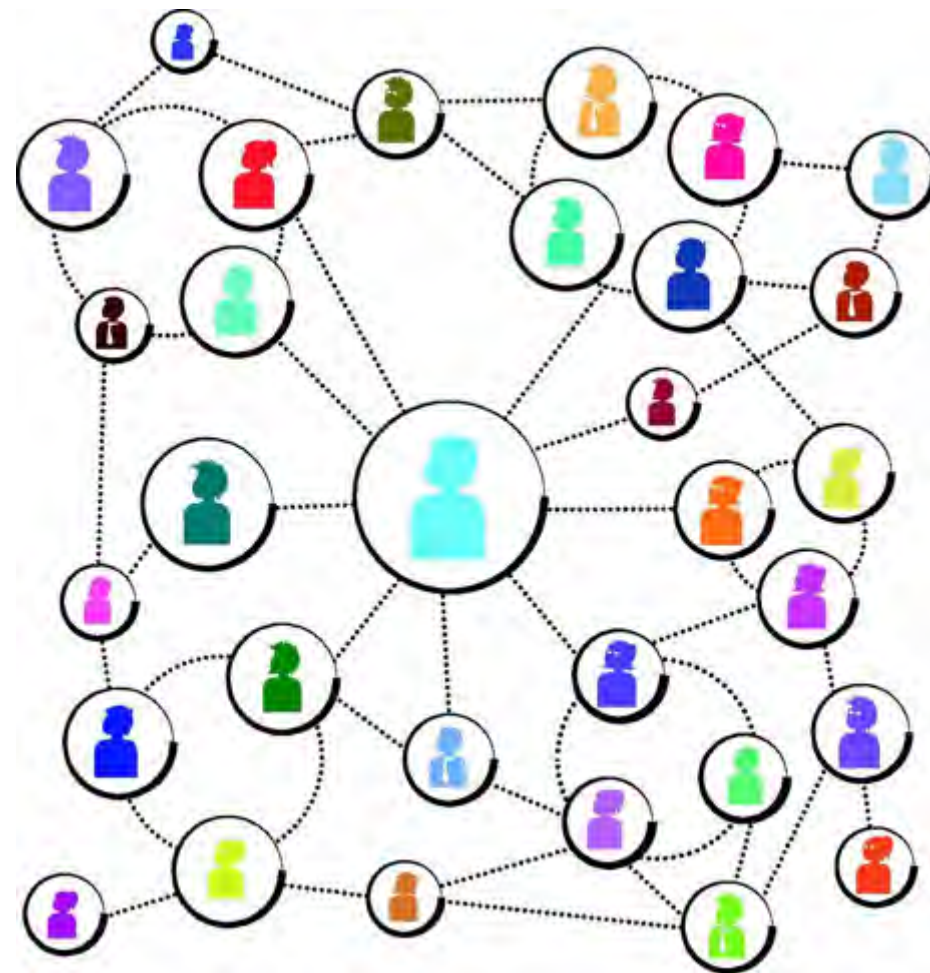
智能查询

导入精灵

为何要用图数据库



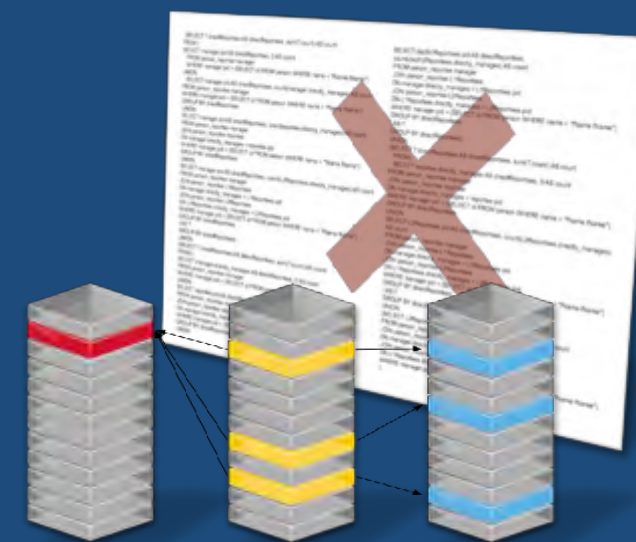
- 世界本来就是由关系组成的
- 关系型数据库处理不好关系
- 图数据库最适合处理关系



关系型数据库不能很好地处理关系

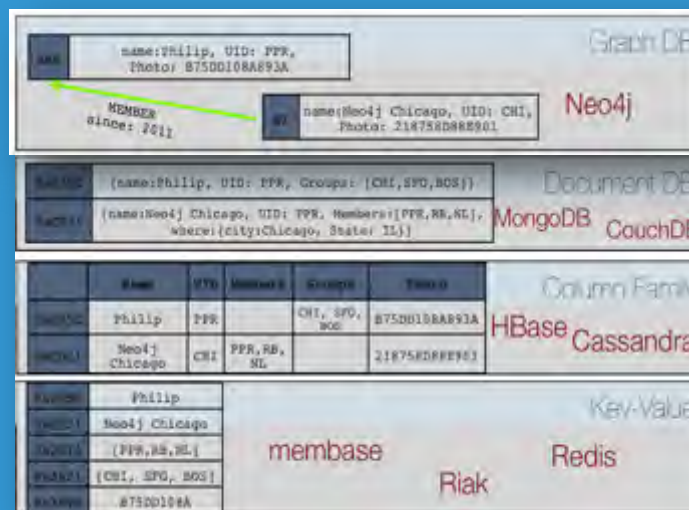
建模难	性能低	查询难	扩展难
<ul style="list-style-type: none"> 不复杂就不能建模和存储数据和关系 	<ul style="list-style-type: none"> 随着关系数量和层次的增加，数据库尺寸的增加 性能降低 	<ul style="list-style-type: none"> 需要 JOIN 操作，查询复杂性增加 	<ul style="list-style-type: none"> 增加新类型的数据和关系 需要重新设计模式 增加上市时间

➔ 导致传统数据库不适用于有实时价值的关系数据



NoSQL 数据库不处理关系

- 没有数据结构建模或存储数据关系
- 没有查询结构支持数据关系
- 在应用中连接数据需要 “JOIN 逻辑”
- 对事务没有 ACID 支持



导致 NoSQL 数据库不适用于有实时价值的关系数据

图数据库 Neo4j 是专为数据关系而生的

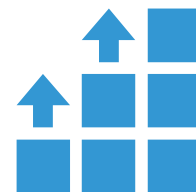
开发优势

- 模型维护容易
- 查询简单

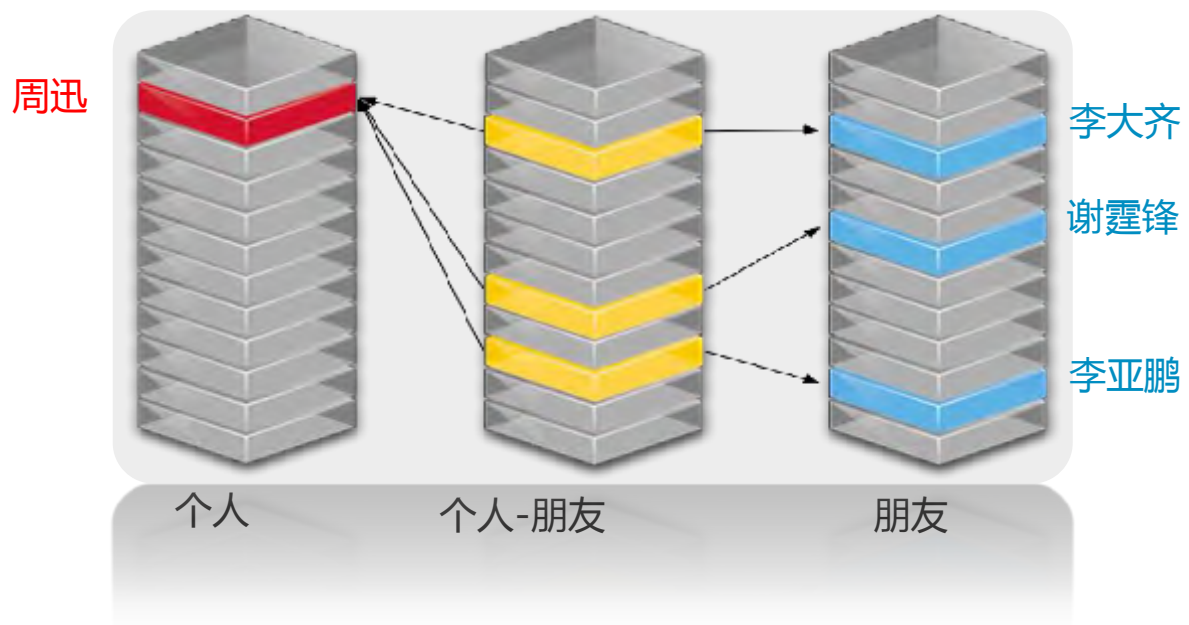


部署优势

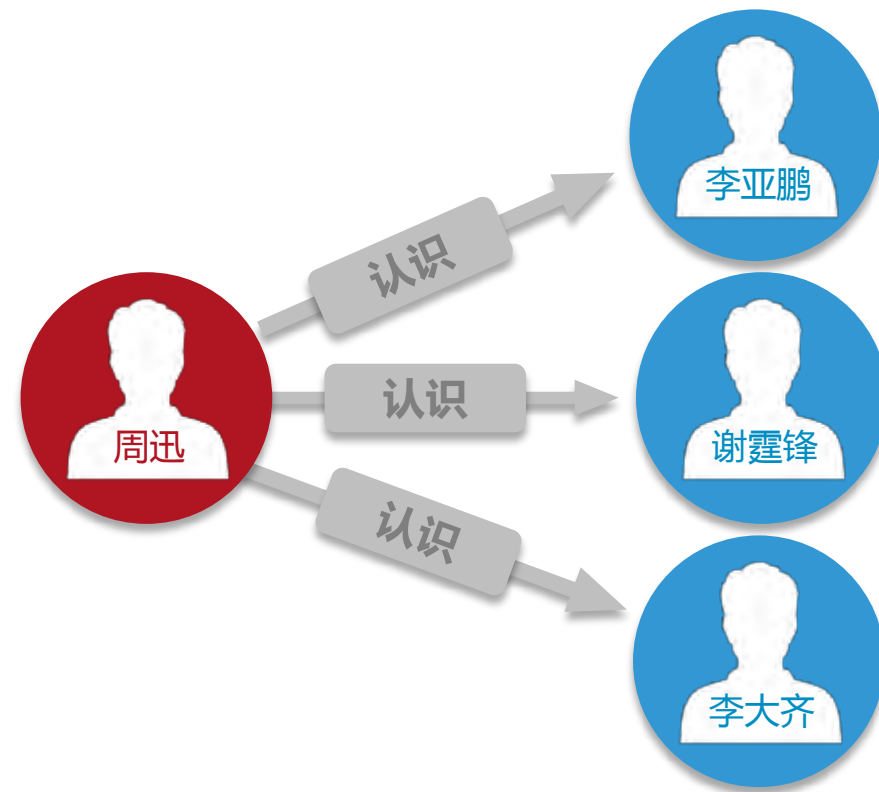
- 超高性能
- 使用最少的资源



关系模型与图模型对比

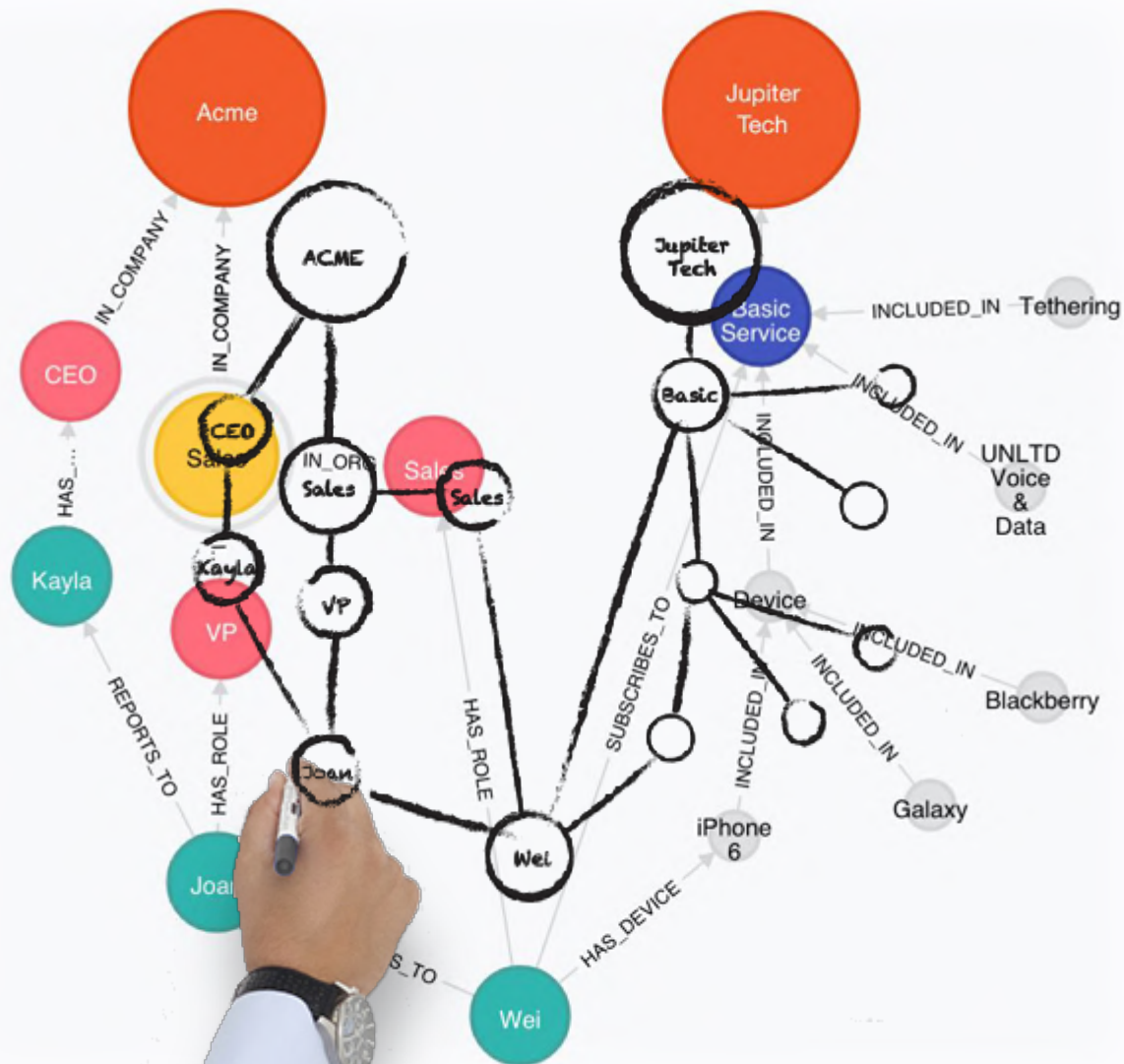


关系模型



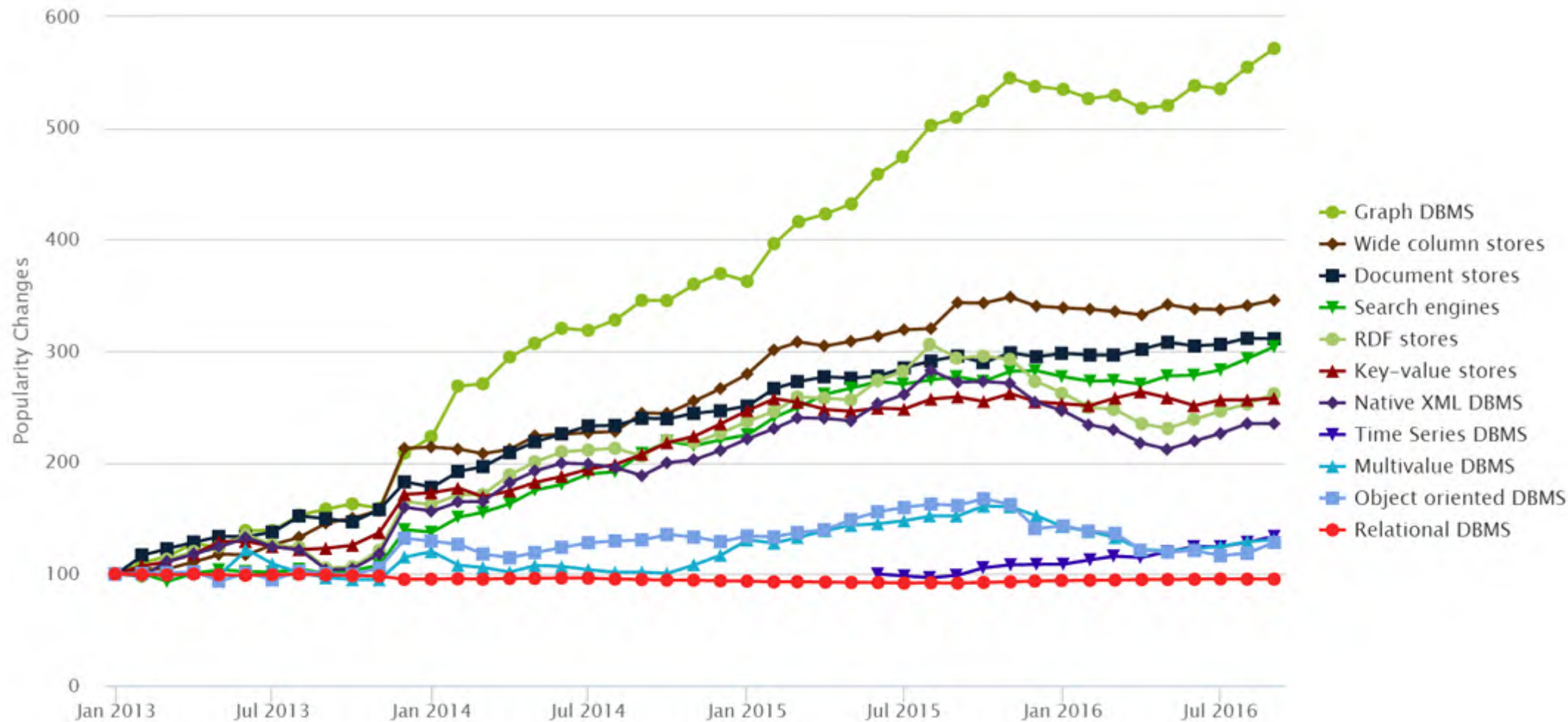
图模型

建模简易: 白板模型即物理模型



数据库按类别人气趋势

2016年6月



内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵

为何要用 Neo4j



世界领先

世界领先的、最好的图数据库。Neo4j 已被数千家机构用于关键产品生产应用中，其中包括全球2000强中的50多家机构。



社区活跃

地球上最大的和最活跃的图形社区。Neo4j 拥有最大的和最活跃的图数据库爱好者社区，这些爱好者撑起了一个稳健、活跃的生态系统。



可伸缩性

不折不扣的高性能读写伸缩性。Neo4j 在提供超高速读写性能的同时，还能保证数据的完整性。



超高性能

依靠完全本地存储和处理达到高性能。索引自由邻接，大大缩短了读取时间，即使数据不断增长，也能提供超并行吞吐量。











最易学习

成熟的用户界面，内置丰富的学习和教育资源，使你能轻松愉快地学习和掌握 Neo4j。

图数据库人气排名





































2015年2月 26.80分

17 systems in ranking, February 2015

Rank	Last Month	DBMS	Database Model	Score	Changes
1.	1.	Neo4j	Graph DBMS	26.80	+2.38
2.	2.	Titan	Graph DBMS	3.38	+0.47
3.	3.	OrientDB	Multi-model 	2.93	+0.39
4.	4.	Sparksee	Graph DBMS	0.96	+0.04
5.	5.	Giraph	Graph DBMS	0.79	+0.03
6.	6.	ArangoDB	Multi-model 	0.73	+0.11
7.	7.	InfiniteGraph	Graph DBMS	0.28	-0.01
8.	8.	Sqrrl	Multi-model 	0.26	+0.03
9.	9.	GraphDB	Multi-model 	0.17	+0.00
10.	10.	InfoGrid	Graph DBMS	0.15	+0.00
11.	11.	FlockDB	Graph DBMS	0.13	+0.00
12.	 13.	HyperGraphDB	Graph DBMS	0.02	-0.00
13.	14.	Amisa Server	Multi-model 	0.00	±0.00
13.	14.	Blazegraph	Multi-model 	0.00	±0.00
13.	12.	GlobalsDB	Multi-model 	0.00	-0.05
13.	14.	GraphBase	Graph DBMS	0.00	±0.00
13.		VelocityGraph	Graph DBMS	0.00	

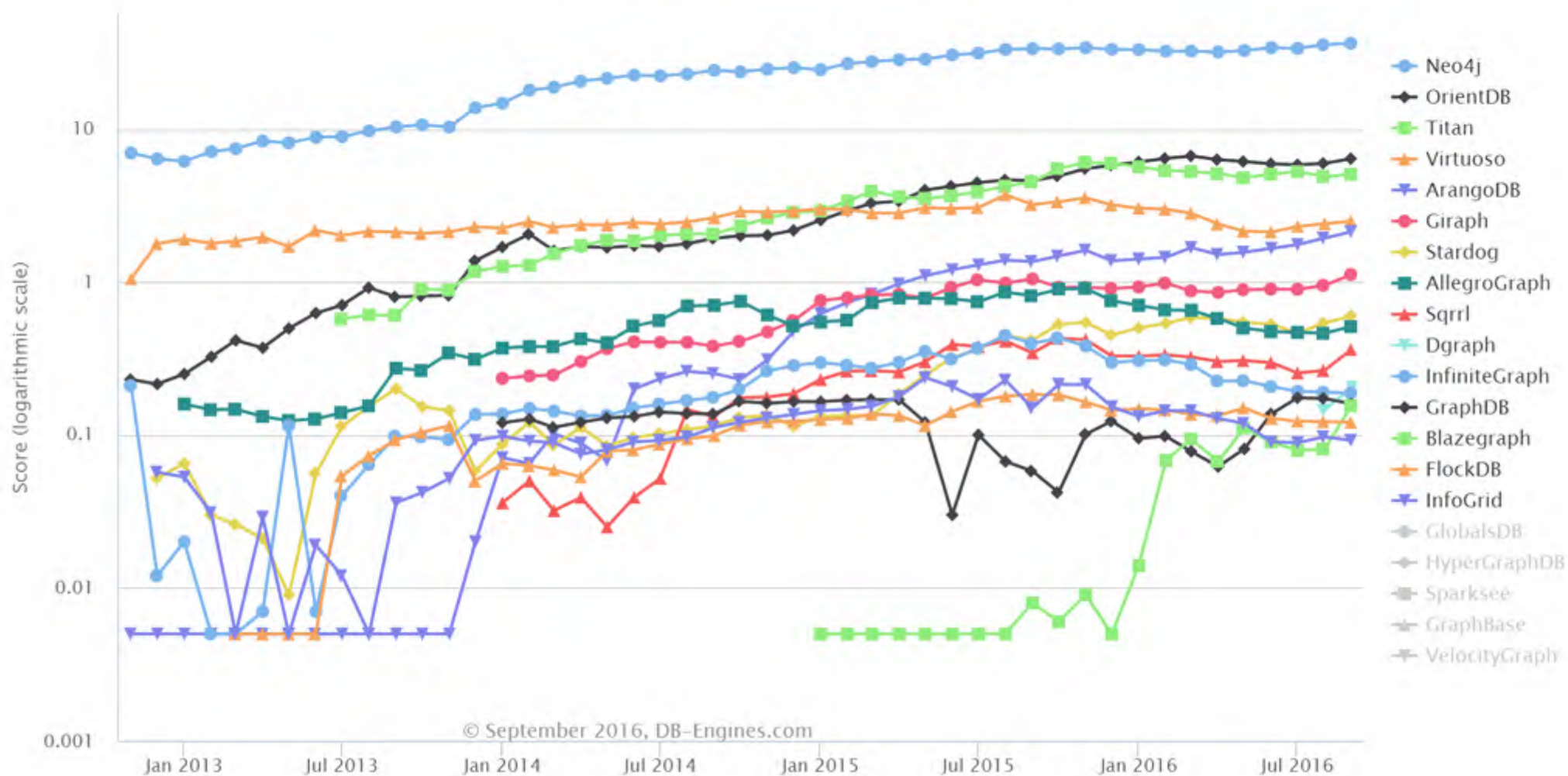
2016年9月 36.37分

21 systems in ranking, September 2016

Rank			DBMS	Database Model	Score		
Sep 2016	Aug 2016	Sep 2015			Sep 2016	Aug 2016	Sep 2015
1.	1.	1.	Neo4j 	Graph DBMS	36.37	+0.80	+2.83
2.	2.	2.	OrientDB	Multi-model 	6.40	+0.43	+1.80
3.	3.	3.	Titan	Graph DBMS	5.06	+0.17	+0.55
4.	4.	4.	Virtuoso	Multi-model 	2.50	+0.10	-0.71
5.	5.	5.	ArangoDB	Multi-model 	2.13	+0.20	+0.77
6.	6.	6.	Giraph	Graph DBMS	1.12	+0.17	+0.07
7.	7.	 8.	Stardog	Multi-model 	0.60	+0.06	+0.18
8.	8.	 7.	AllegroGraph 	Multi-model 	0.51	+0.05	-0.30
9.	9.	 10.	Sqrrl	Multi-model 	0.36	+0.10	+0.02
10.	 12.		Dgraph	Graph DBMS	0.20	+0.06	
11.	 10.	 9.	InfiniteGraph	Graph DBMS	0.19	-0.00	-0.21
12.	 11.	 17.	GraphDB	Multi-model 	0.16	-0.01	+0.10
13.	 15.	 19.	Blazegraph 	Multi-model 	0.16	+0.07	+0.15
14.	 13.	 13.	FlockDB	Graph DBMS	0.12	-0.00	-0.06
15.	 14.	 14.	InfoGrid	Graph DBMS	0.09	-0.00	-0.06
16.	 18.	 15.	GlobalsDB	Multi-model 	0.08	+0.04	-0.05
17.	 16.	 12.	HyperGraphDB	Graph DBMS	0.08	-0.00	-0.11
18.	 17.	 11.	Sparksee	Graph DBMS	0.04	-0.03	-0.22
19.	19.	 18.	GraphBase	Graph DBMS	0.03	+0.00	-0.03
20.	 21.	20.	Amisa Server	Multi-model 	0.00	±0.00	±0.00
20.	20.	 16.	VelocityGraph	Graph DBMS	0.00	-0.02	-0.09

图数据库人气趋势

DB-Engines Ranking of Graph DBMS



© September 2016, DB-Engines.com

© http://db-engines.com/en/ranking_trend/graph+dbms

拥有众多图爱好者的最大生态系统



超过1百万的下载量



超过2万培训师



超过1.8万聚会成员



超过100个技术和服务伙伴



超过150个企业版订购客户
其中全球2000中超过 50 家



Neo4j 领导图数据库革命

Gartner

“对于在数据捕获设计之后，追求数据驱动运营和决策的组织而言，图分析可能是最有效的竞争优势。”

FORRESTER

“Forrester 估计，到 2017 年，超过 25% 的企业 将使用图数据库。”

EMA

“Neo4j 是当前市场图数据库的领袖。”

Neo4j 获奖及头条报道

InfoWorld

年度技术 2013 2014 2015 2016
博西大数据奖 2013

database
TRENDS AND APPLICATIONS

数据领域中最重要 的 100 家公司
2016

SD Times
SOFTWARE DEVELOPER

最佳展示 100
2014

Gartner

ODBMS 魔力象限 2014
Who's Who in NOSQL DBMSs 2013

CRN

数据管理领域中大数据 100
2013

DB-ENGINES

“图数据库管理系统中的领先者是 Neo4j” 2014

451 Research

Neo's GraphConnect shows graph databases coming into their own
Matt Aslett 2013

BG

Neo Technology – 图数据库的兴起
– Robin Bloor 2013



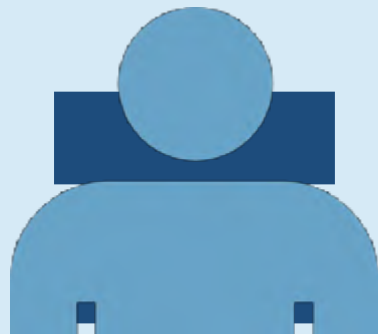
O' Reilly出版 – 图数据库
由 Neo Technology 职员编著

Neo4j 关键产品特征

 <p>确保数据一致性和性能</p>	 <p>实时下每秒达数以百万跳</p>	 <p>数据建模更加自然</p>	 <p>完全支持 ACID 事务处理</p>
原生的图存储	原生的图处理	“白板友好” 数据建模	数据完整性高
 <p>比 SQL 要少 10 到 100 倍的代码</p>	 <p>图的垂直和水平伸缩优化</p>	 <p>其它数据库的无缝导入</p>	 <p>流行语言的驱动程序及 API</p>
功能强大，富有表现力的查询语言	可伸缩性和高可用性	内置的 ETL	集成

集群特征

- 主从复制，重选主服务器和容错
- 每个实例都有自己的本地缓冲
- 水平伸缩及灾难恢复

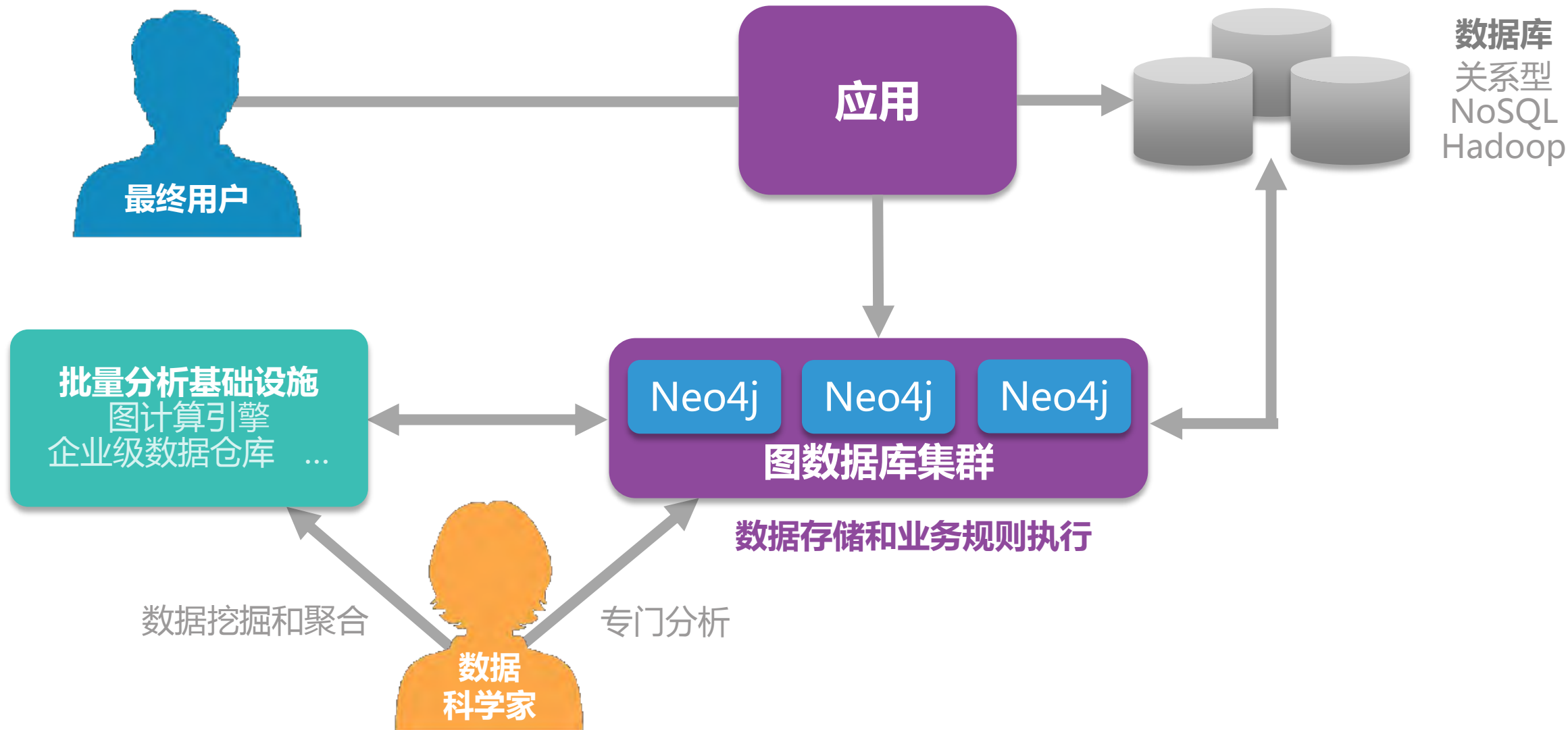


性能优势

- 查询内不跨越网络
- 实时操作，具有快速和一致的响应时间
- 缓冲分区，对于非常大的图，跨集群扩展缓冲



Neo4j 融入你的企业环境



Neo4j: 图数据库领袖



内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵

谁在用 Neo4j



203
国家

10 万
账户

6 万
文件

27.5 万
节点

40 万
联系

ICIJ 调查组织

采用 Neo4j 进行处理

查出成千上万诈骗、贿赂、逃逸事件

内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵

演示：明星关系图（原文）



演示：明星关系图（创建）

创建索引

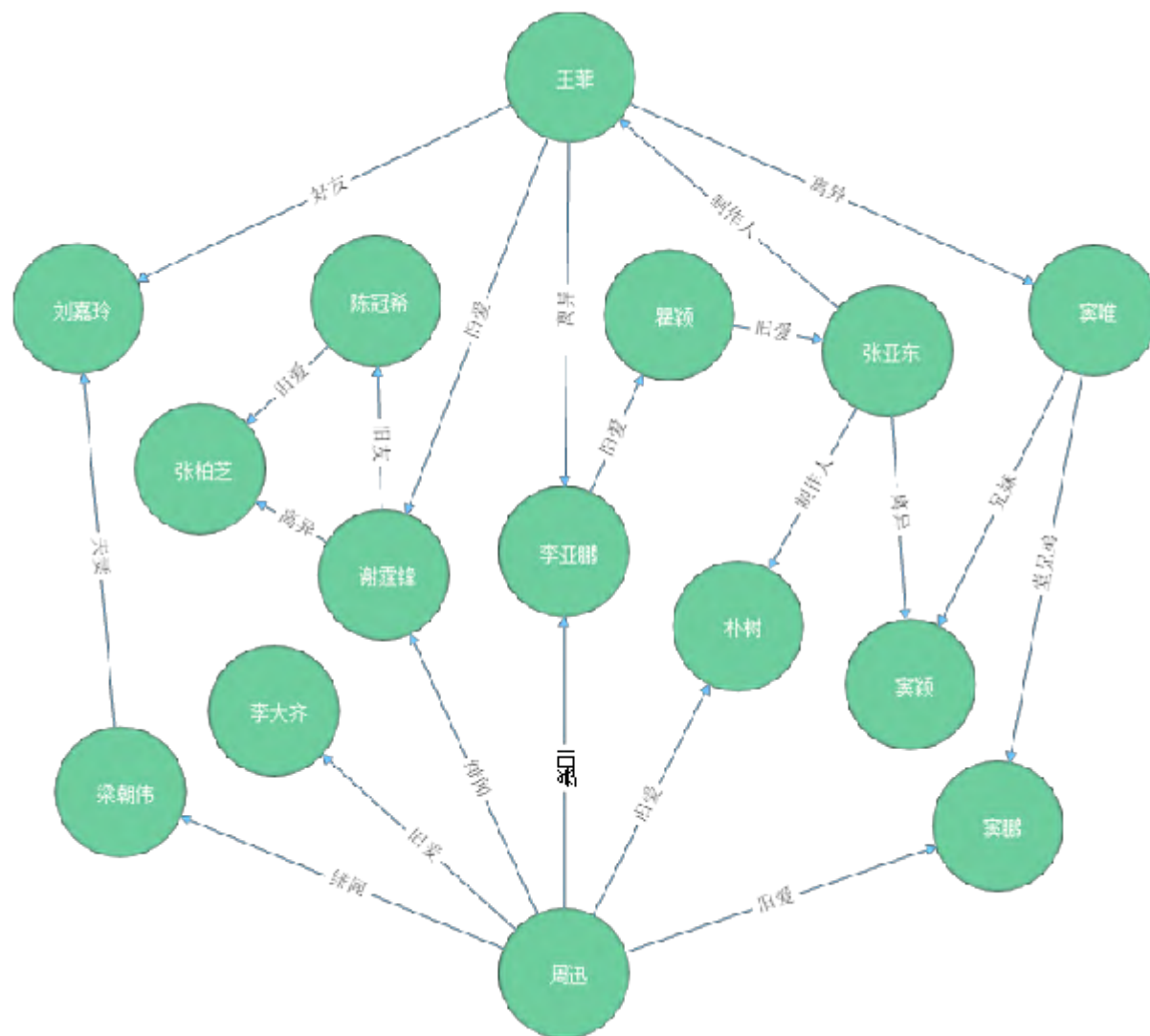
```
create index on :明星(姓名)
```

创建对象及关系

```
CREATE
(陈冠希:明星 { 姓名: "陈冠希"}),
(王菲:明星 { 姓名: "王菲"}),
(李亚鹏:明星 { 姓名: "李亚鹏"}),
(瞿颖:明星 { 姓名: "瞿颖"}),
(张柏芝:明星 { 姓名: "张柏芝"}),
(谢霆锋:明星 { 姓名: "谢霆锋"}),
(周迅:明星 { 名称: "周迅"}),
(张亚东:明星 { 名称: "张亚东"}),
(窦唯:明星 { 名称: "窦唯"}),
(窦鹏:明星 { 名称: "窦鹏"}),
(李大齐:明星 { 名称: "李大齐"}),
(朴树:明星 { 名称: "朴树"}),
(窦颖:明星 { 名称: "窦颖"}),
(梁朝伟:明星 { 名称: "梁朝伟"}),
(刘嘉玲:明星 { 名称: "刘嘉玲"}),
(周迅)-[:旧爱]->(窦鹏),
(周迅)-[:旧爱]->(李大齐),
(周迅)-[:旧爱]->(朴树),
(周迅)-[:旧爱]->(李亚鹏),
(周迅)-[:绯闻]->(梁朝伟),
(周迅)-[:绯闻]->(谢霆锋),
(王菲)-[:离异]->(李亚鹏),
(王菲)-[:旧爱]->(谢霆锋),
```

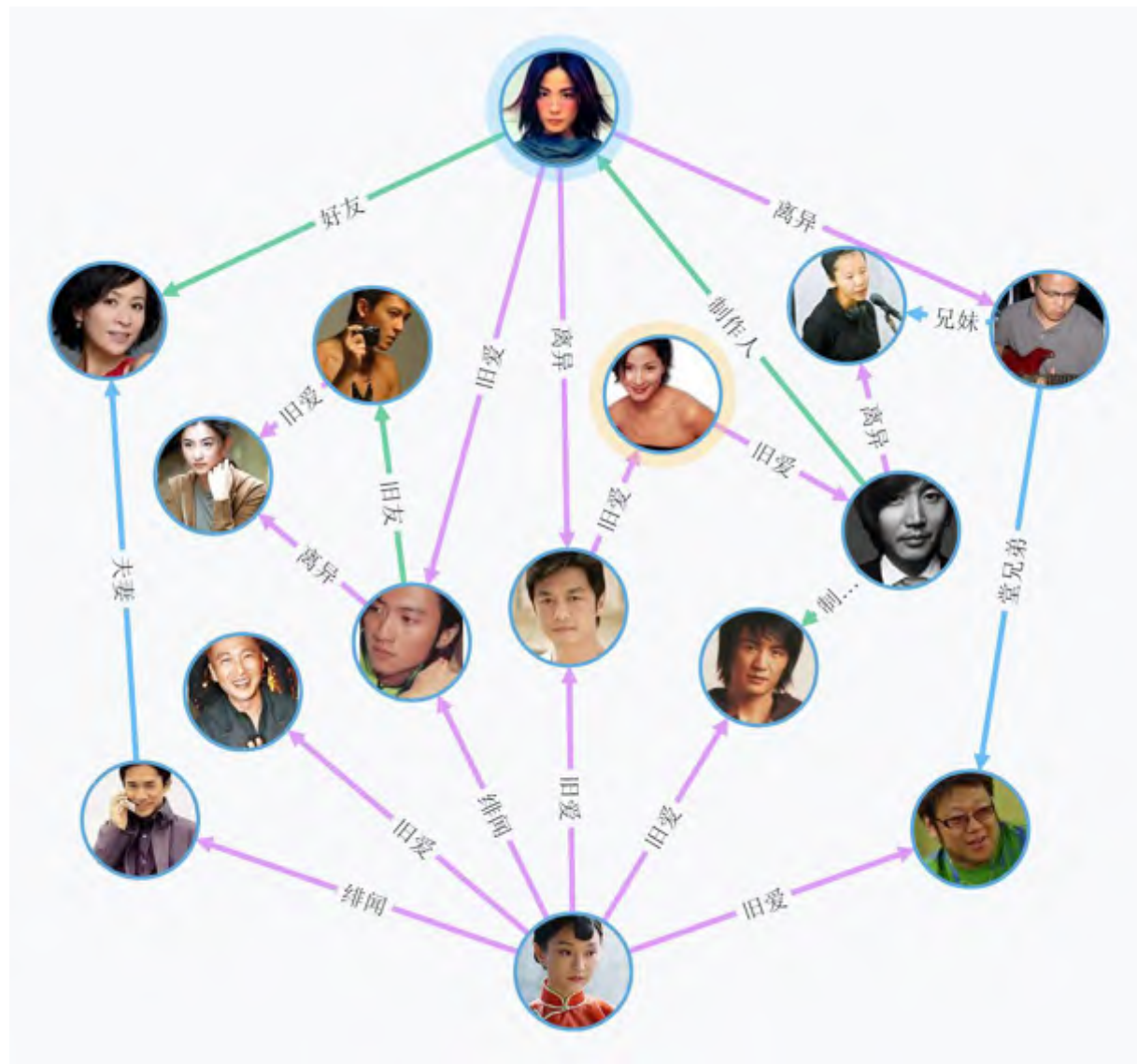
演示：明星关系图（原版）

MATCH (n:明星) RETURN n



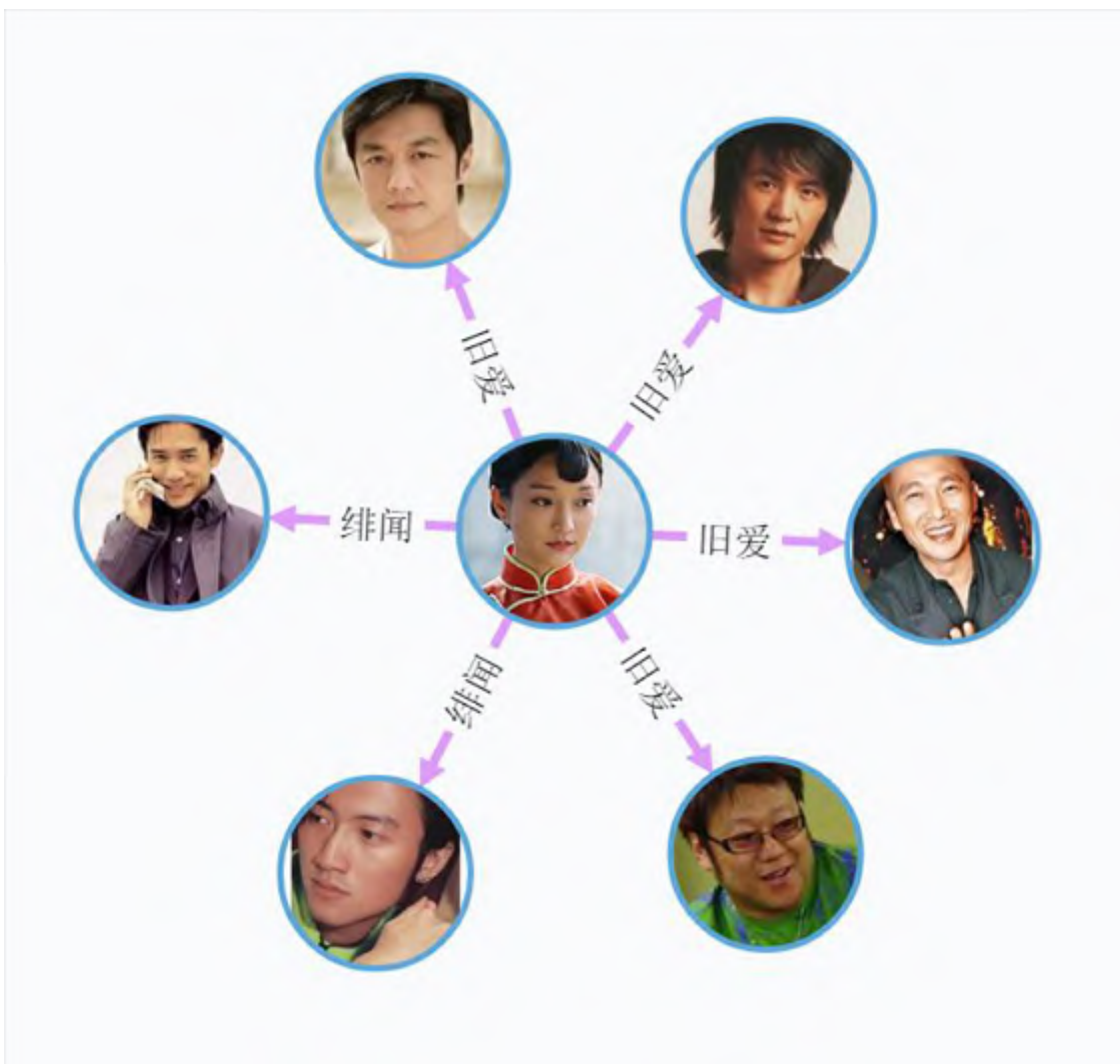
演示：明星关系图（国产版）

MATCH (n:明星) RETURN n



演示：查询周迅的直接关系

```
match m=(:明星 {名称:"周迅"})-[*..1]-() return m
```



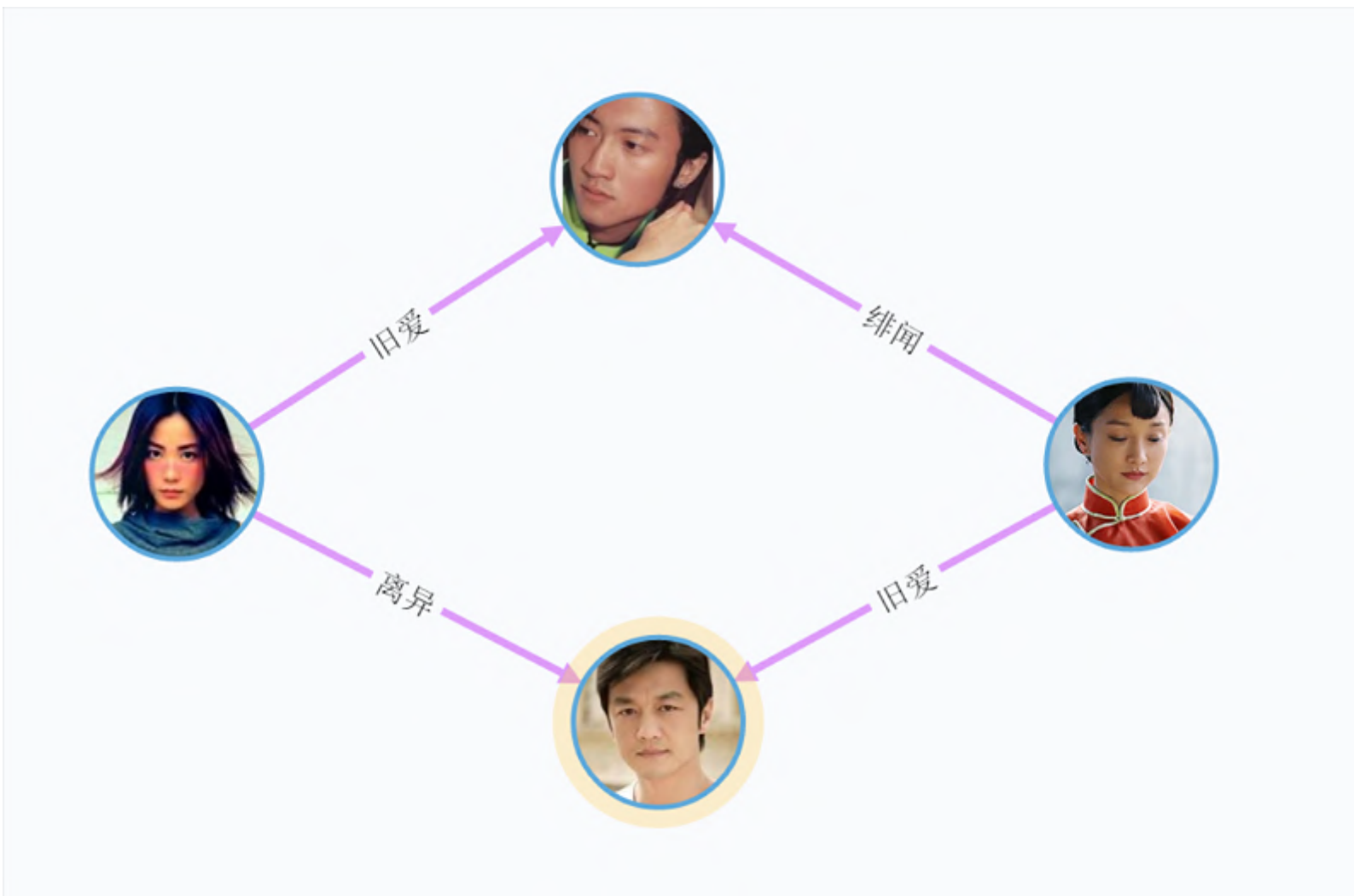
演示：查询周迅与王菲的最短路径

```
MATCH p = shortestPath((周迅:明星 {名称:"周迅"})-[*..6]-(王菲:明星 {名称:"王菲"}))RETURN p
```



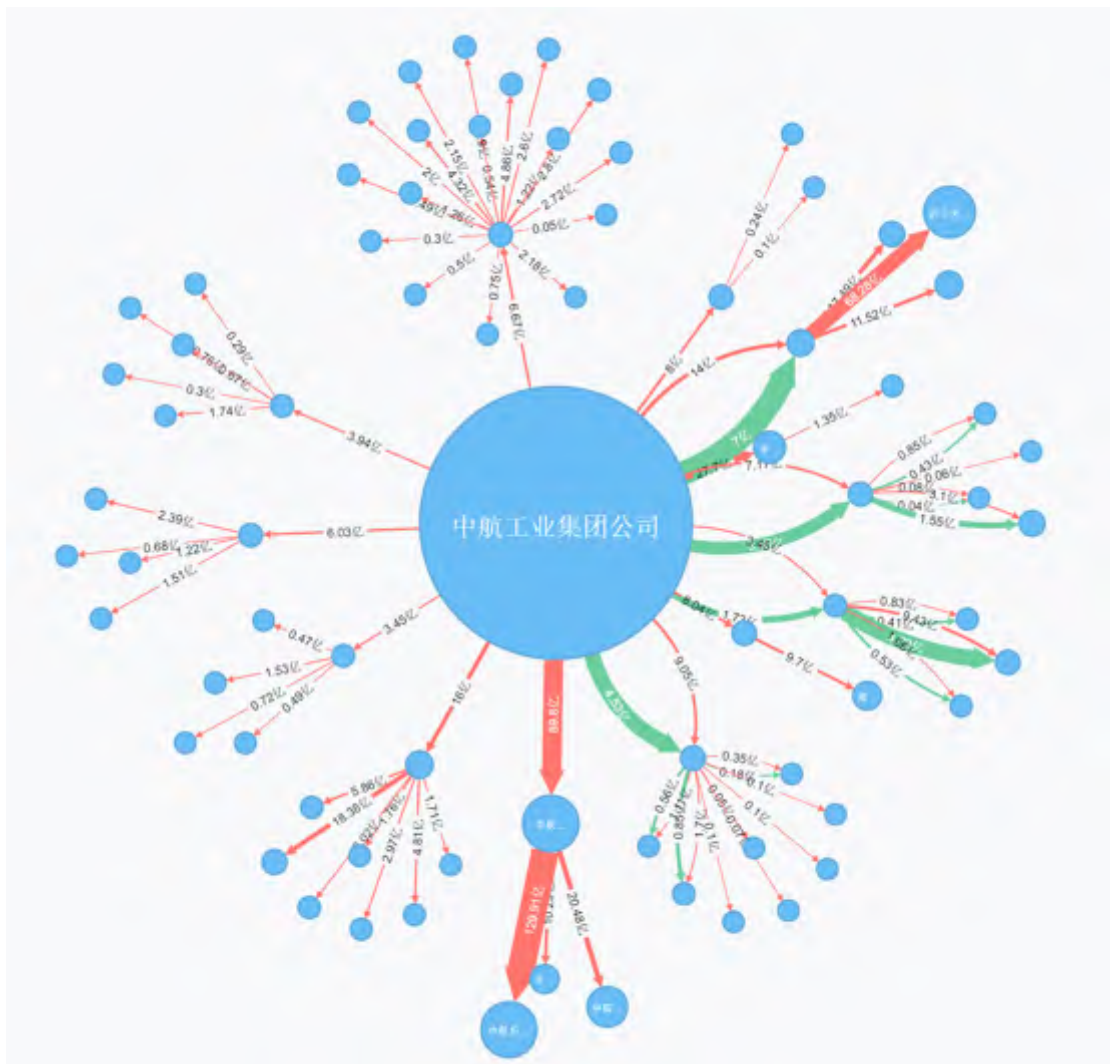
演示：查询周迅与王菲的所有最短路径

```
MATCH p = allShortestPaths((周迅:明星 {名称:'周迅'})-[*..6]-(王菲:明星 {名称:'王菲'})) RETURN p
```



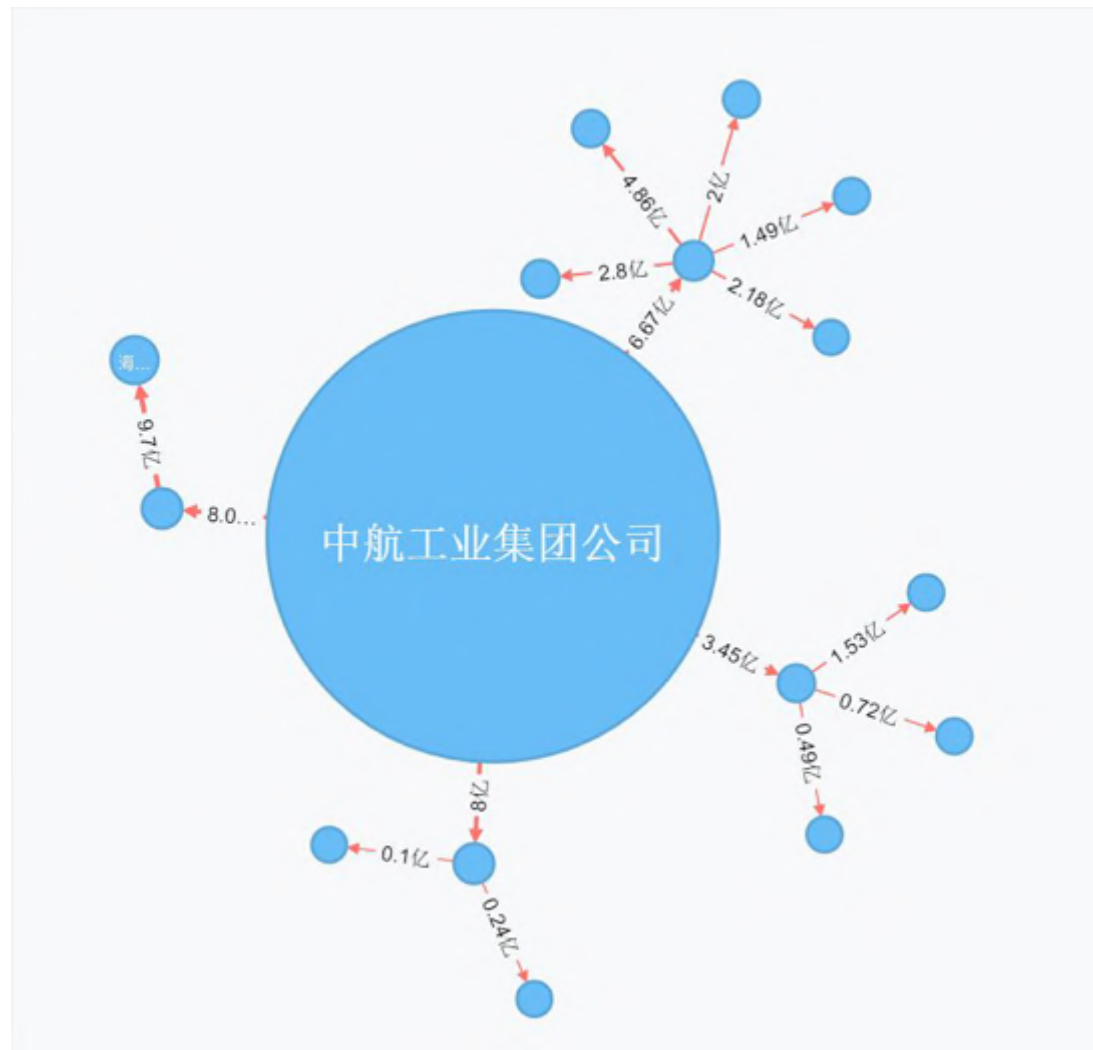
演示：投资关系图（投资与担保）

MATCH a=(:公司 {名称:'中航工业集团公司'})-[r*]->() RETURN nodes(a)



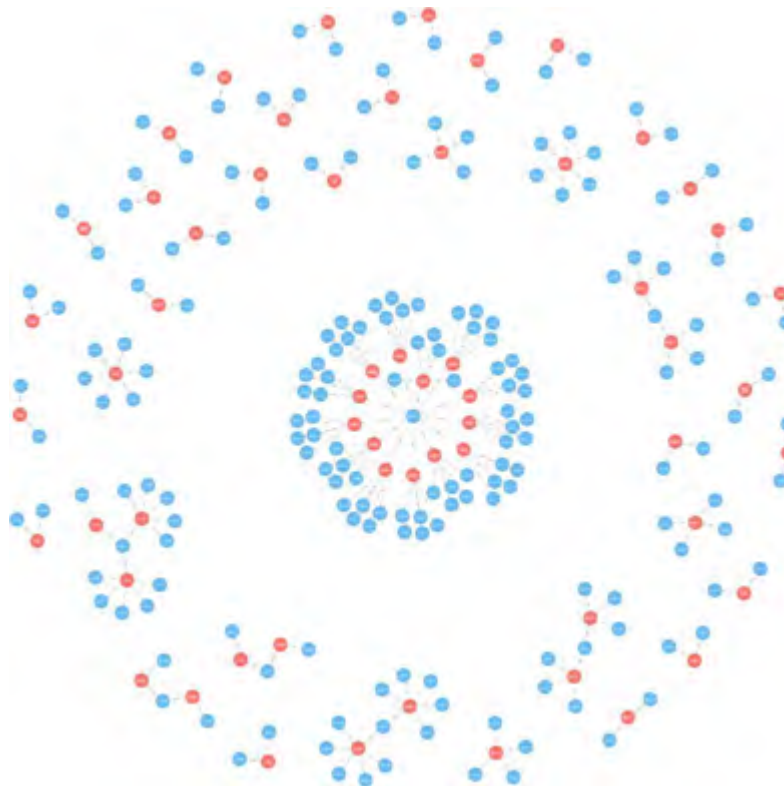
演示：投资关系图（亏本）

MATCH a=(:公司 {名称:'中航工业集团公司'})-[r1:投资]->(c1:公司)-[r2:投资]->(c2:公司) where c2.净利润<0 RETURN nodes(a)



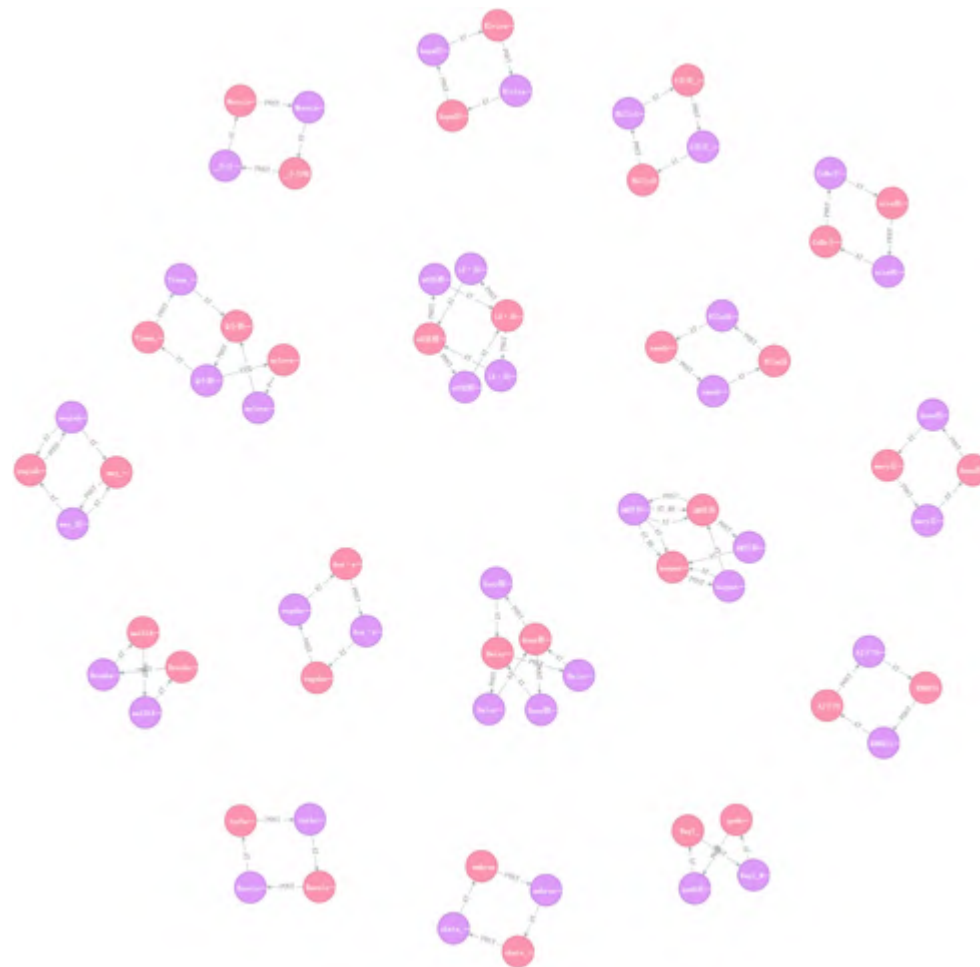
演示：查找标题中包含"性福"的邮件的传递路径

```
match m=(s:Person)-->(e:Email)-->(r:Person) where e.title=~".*性福.*" return m
```



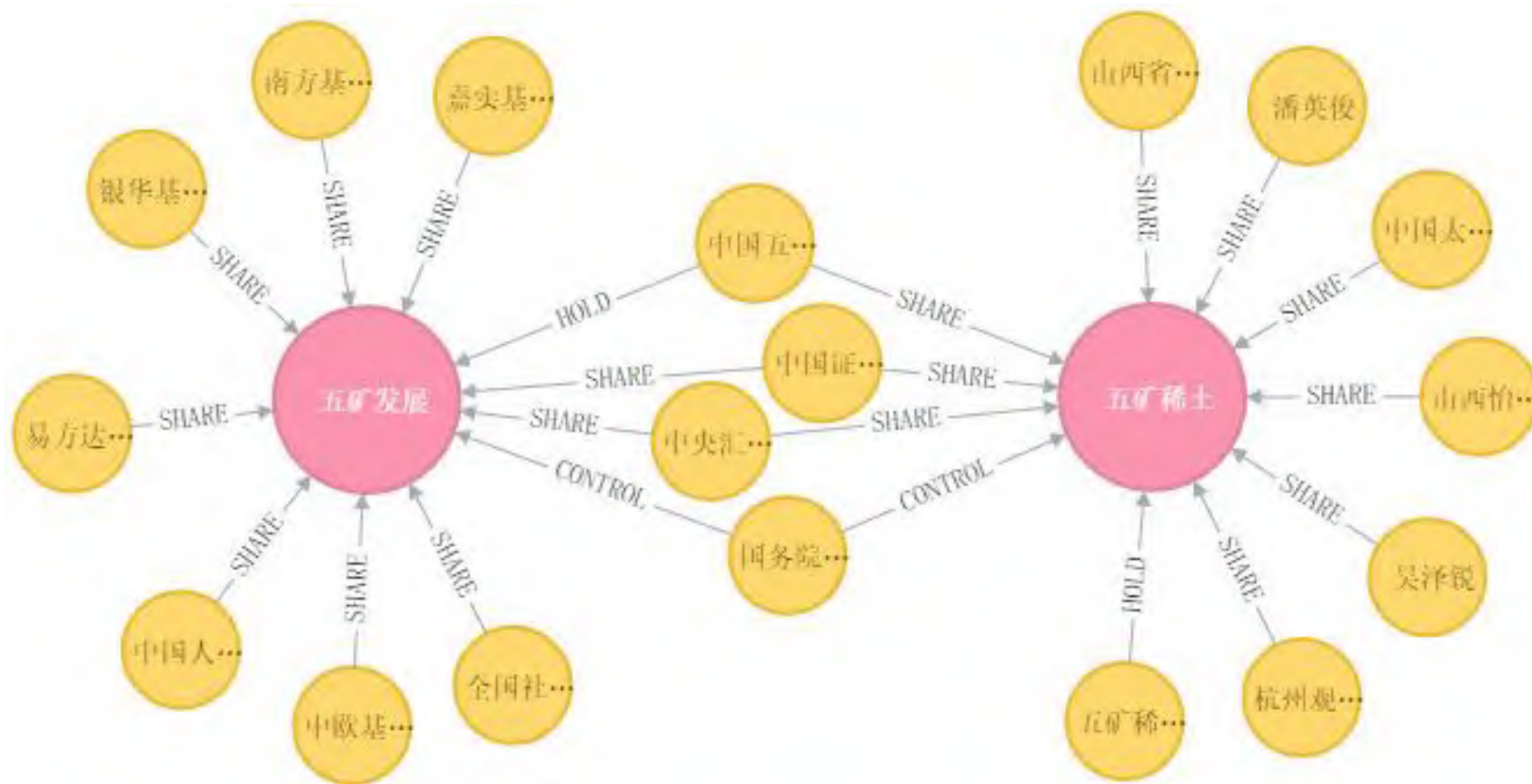
演示：查找微博博主团伙关系

```
match r=(p1:Poster)-[:POST]->(:Post)-[:AT]->(p2:Poster)-[:POST]->(:Post)-[:AT]->(p1) where p1.userName<>p2.userName return r
```

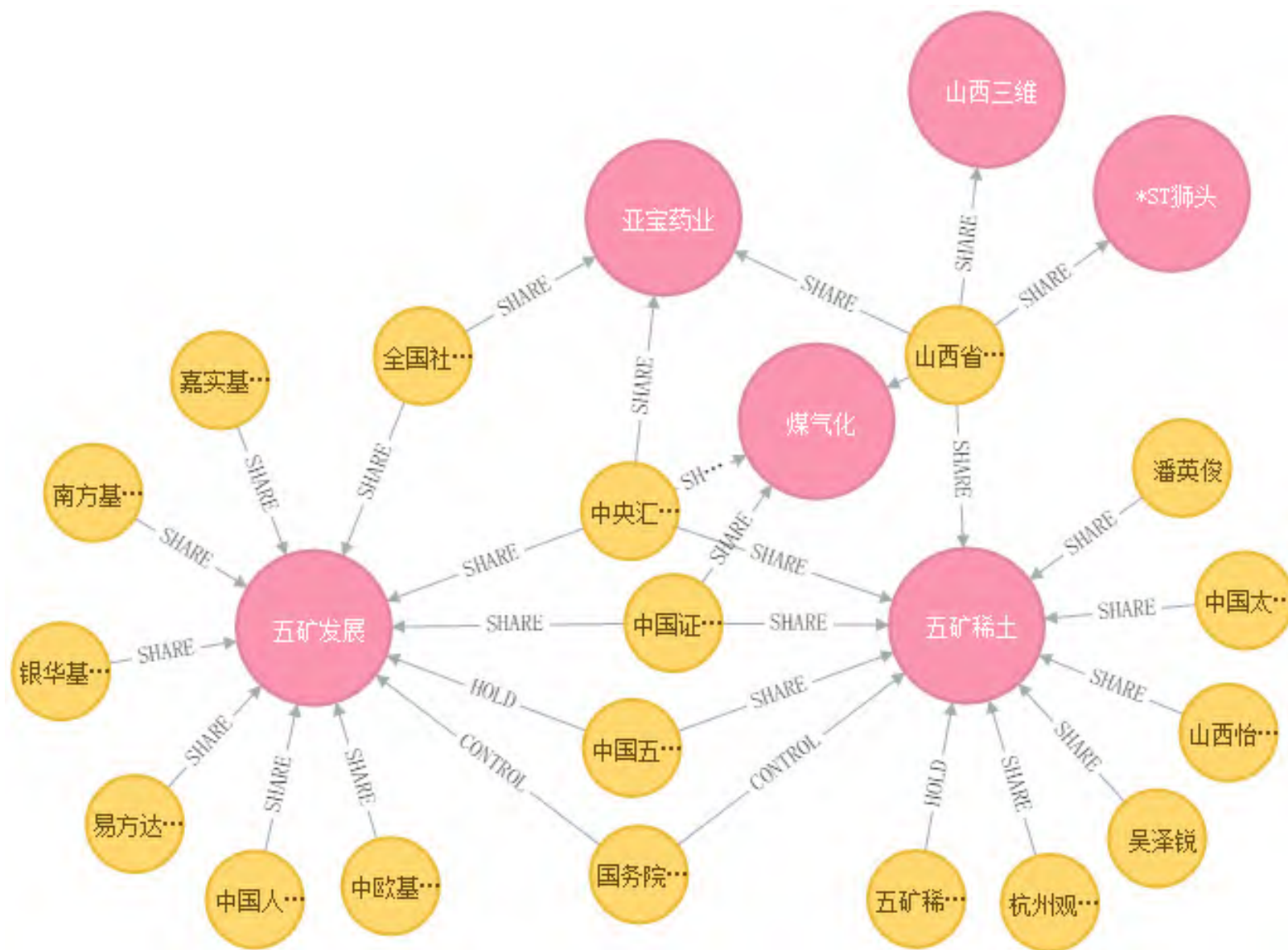


演示：查询股票与股东的关系

```
match m=(h:Holder)-->(s:Stock) where s.name=~".*五矿.*" return m
```



演示：双击一下山西省...



内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵



实时推荐



主数据管理



欺诈检测



基于图的搜索



IT网络管理



身份和访问管理

Neo4j 典型案例



实时推荐



主数据管理



欺诈检测



基于图的搜索



IT网络管理



身份和访问管理

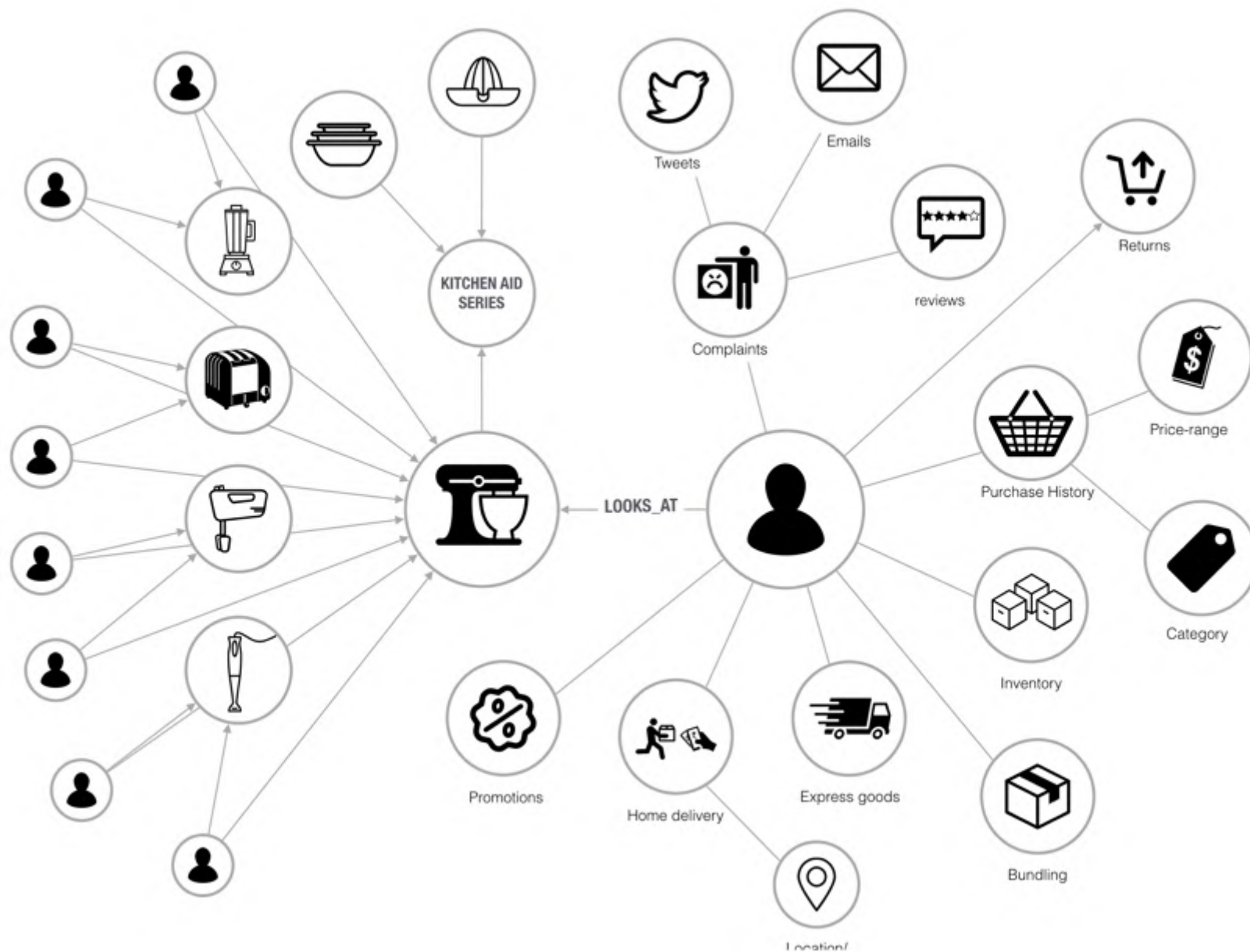


Walmart

“Neo4j 是当今市场图数据库的领袖,它具有可伸缩性和高可用性等企业级功能,是满足我们需求的最佳选择。”

Marcos Wada
Software Developer, Walmart

实时推荐



Neo4j 典型案例



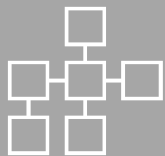
实时推荐



主数据管理



欺诈检测



基于图的搜索



IT网络管理



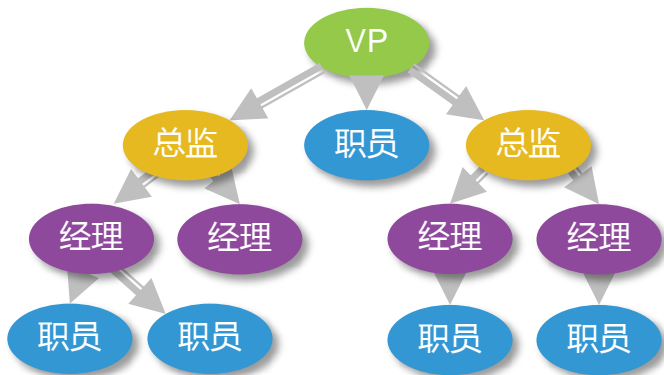
身份和访问管理



Neo4j 是 Cisco HMP 的心脏: 用于管理和单一可信来源, 以及 Cisco 所有层次结构的一站式商店

-Prem Malhotra,
Director of Enterprise Architecture at Cisco Systems

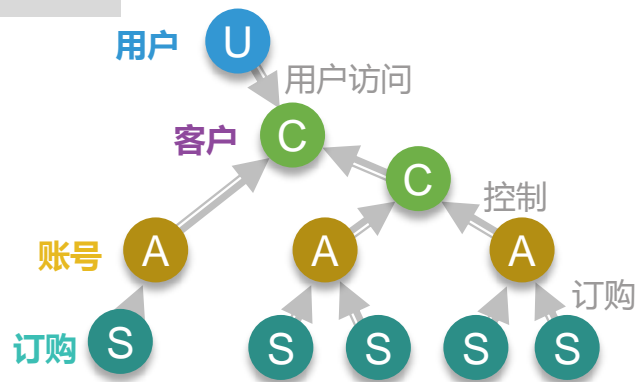
组织架构



社交网络



产品订购



IT网络



Neo4j 典型案例



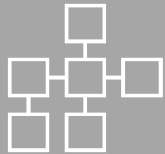
实时推荐



主数据管理



欺诈检测



基于图的搜索



IT网络管理

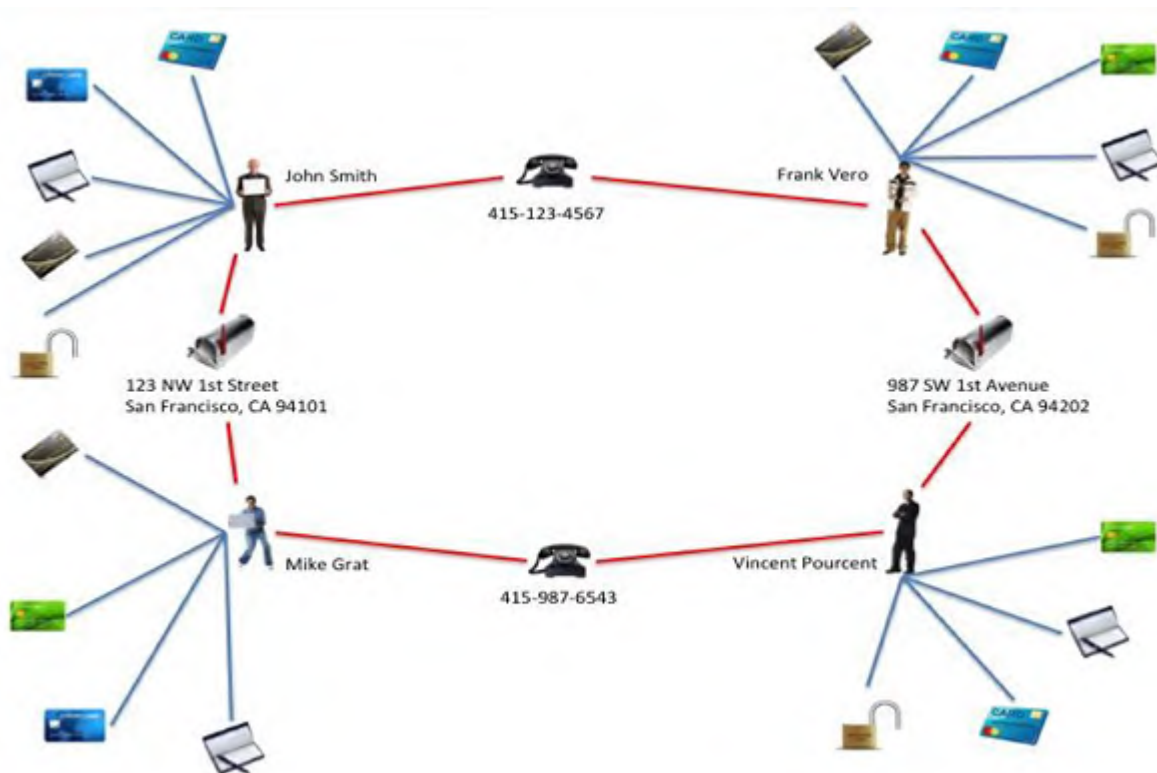


身份和访问管理

“图数据库为揭露诈骗环、高级骗局提供了高度精确的新方法, 并且能够实时终止高级欺诈行为。”

Gorka Sadowski
Cyber Security Expert

合成身份的诈骗环



4 个合成身份的诈骗环,
每个身份有 4-5 账号, 总共 18 个账号

 \$ 4K

 \$ 72K

假定每个账号平均贷款 \$4K, 银行的损失则高达 \$72K

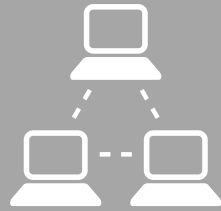
Neo4j 典型案例



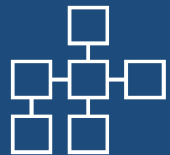
实时推荐



主数据管理



欺诈检测



基于图的搜索



IT网络管理



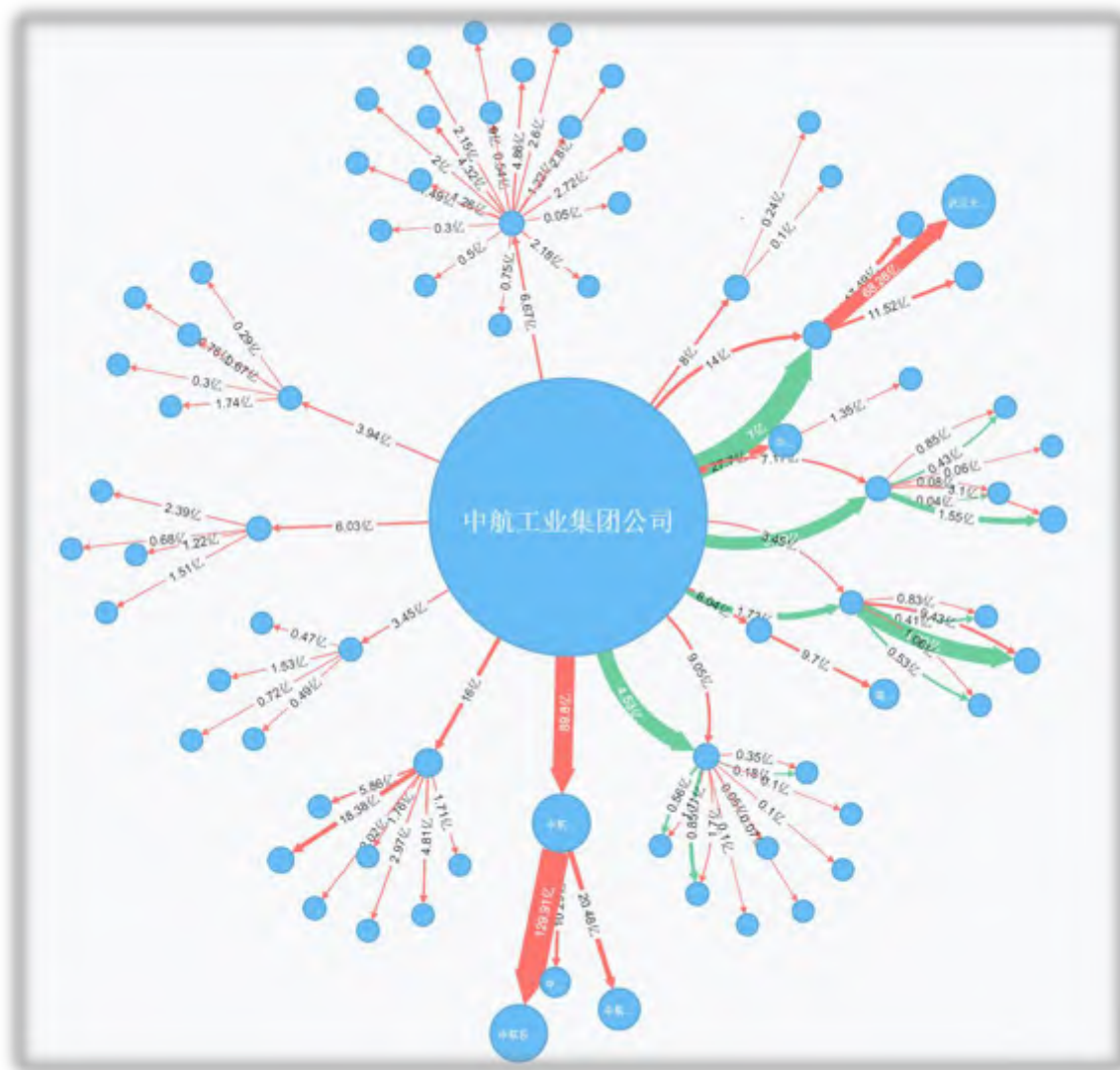
身份和访问管理



Lufthansa

使用 Neo4j 来管理数字资产及其新一代机上娱乐系统

基于图的搜索: 公司投资担保关系网



Neo4j 典型案例



实时推荐



主数据管理



欺诈检测



基于图的搜索



IT网络管理



身份和访问管理



使用 Neo4j 为大型电信服务提供商进行
网络拓扑分析



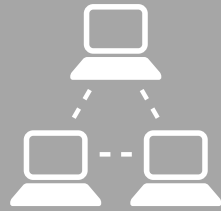
Neo4j 典型案例



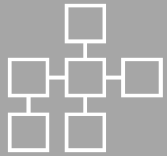
实时推荐



主数据管理



欺诈检测



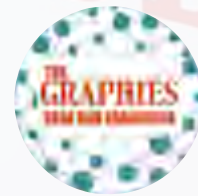
基于图的搜索



IT网络管理

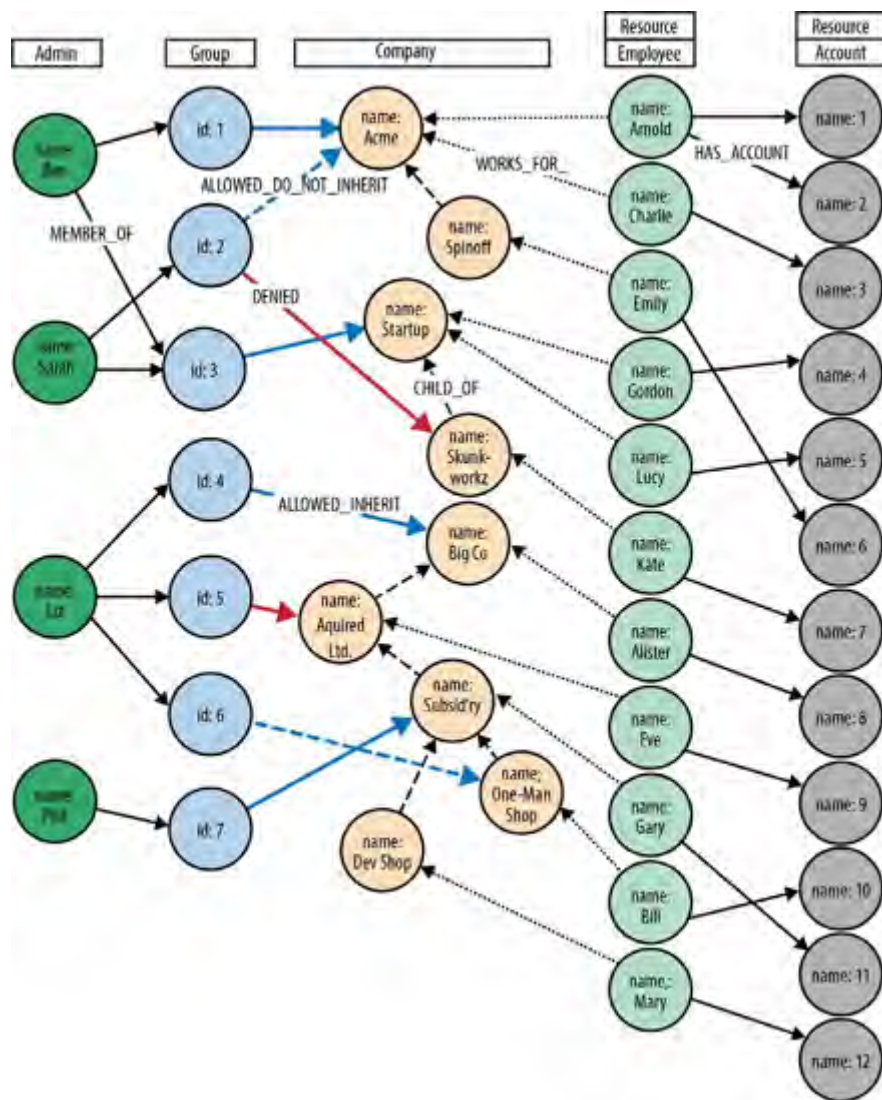


身份和访问管理



瑞银 (UBS) 获得 2014 “最佳身份和访问管理 App” 奖

身份和访问管理



管理组
● 公司
● 员工
● 账号



更多成功案例

社交	推荐	搜索与发现	网络和数据中心	主数据管理	身份和访问	GEO

垂直应用精选

金融服务



通讯



健康和生命科学



HR和招聘



媒体和出版



社交网



工业和物流



游戏娱乐



消费品零售



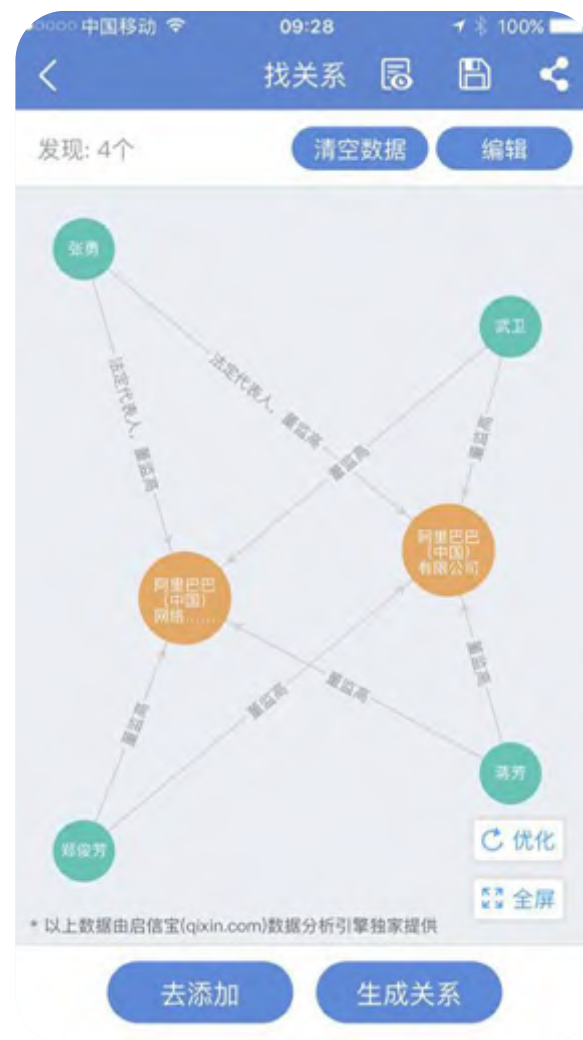
商业服务



信息服务



国内案例: 征信



内容



基本概念

什么是图数据库

为何要用图数据库

为何要用 Neo4j

谁在用 Neo4j

演示 Neo4j



应用案例

实时推荐

主数据管理

欺诈检测

IT网络管理

身份和访问管理



国产化进程

本地化

图片化

数据驱动

智能查询

导入精灵



The screenshot shows the Neo4j web interface in Chinese. The browser address bar shows 'we-yun.com:7070'. The main content area features the Neo4j logo and version '3.1.0 - COMMUNITY - 简体中文版 by 微云数聚'. Below the logo are three main sections: '了解 Neo4j' (Learn Neo4j), '跳进代码' (Jump into code), and '监控系统' (Monitoring system). Each section has a list of topics and a '开始学习' (Start learning) button. At the bottom, a banner reads '每一个优秀的图都开始于 Neo4j.' (Every great graph starts with Neo4j.)

Neo4j

we-yun.com:7070

Neo4j

3.1.0 - COMMUNITY - 简体中文版
by 微云数聚

了解 Neo4j

图将使您顿悟.

- 什么是图数据库?
- 如何查一个图?
- 用 Neo4j 能做什么?

开始学习

跳进代码

使用图查询语言 Cypher.

- 代码演练
- 从关系数据库到图数据库
- 查询模板

编写代码

监控系统

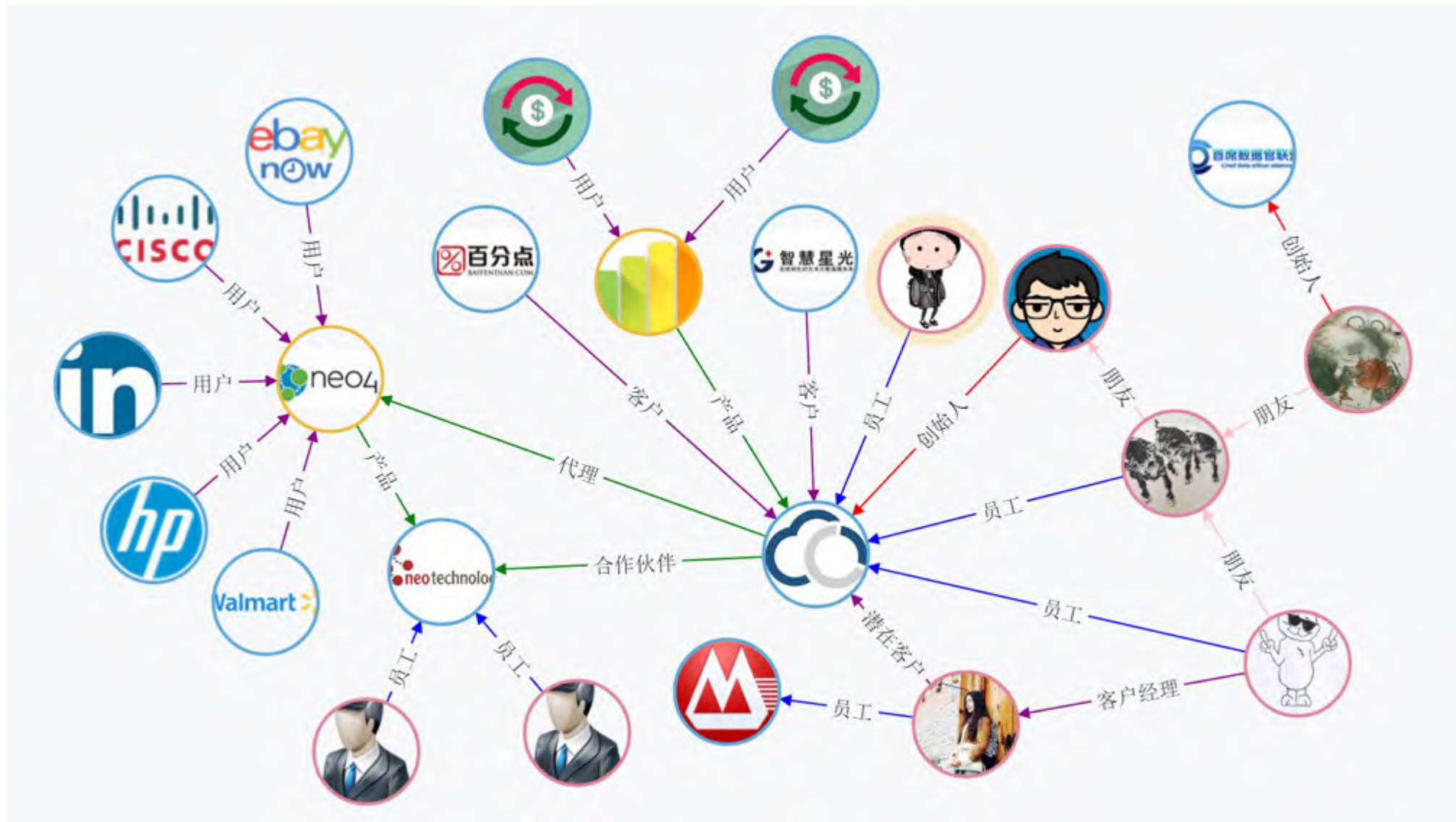
关键的系统健康状态指标.

- 磁盘使用情况
- 高速缓冲活动
- 集群健康现状

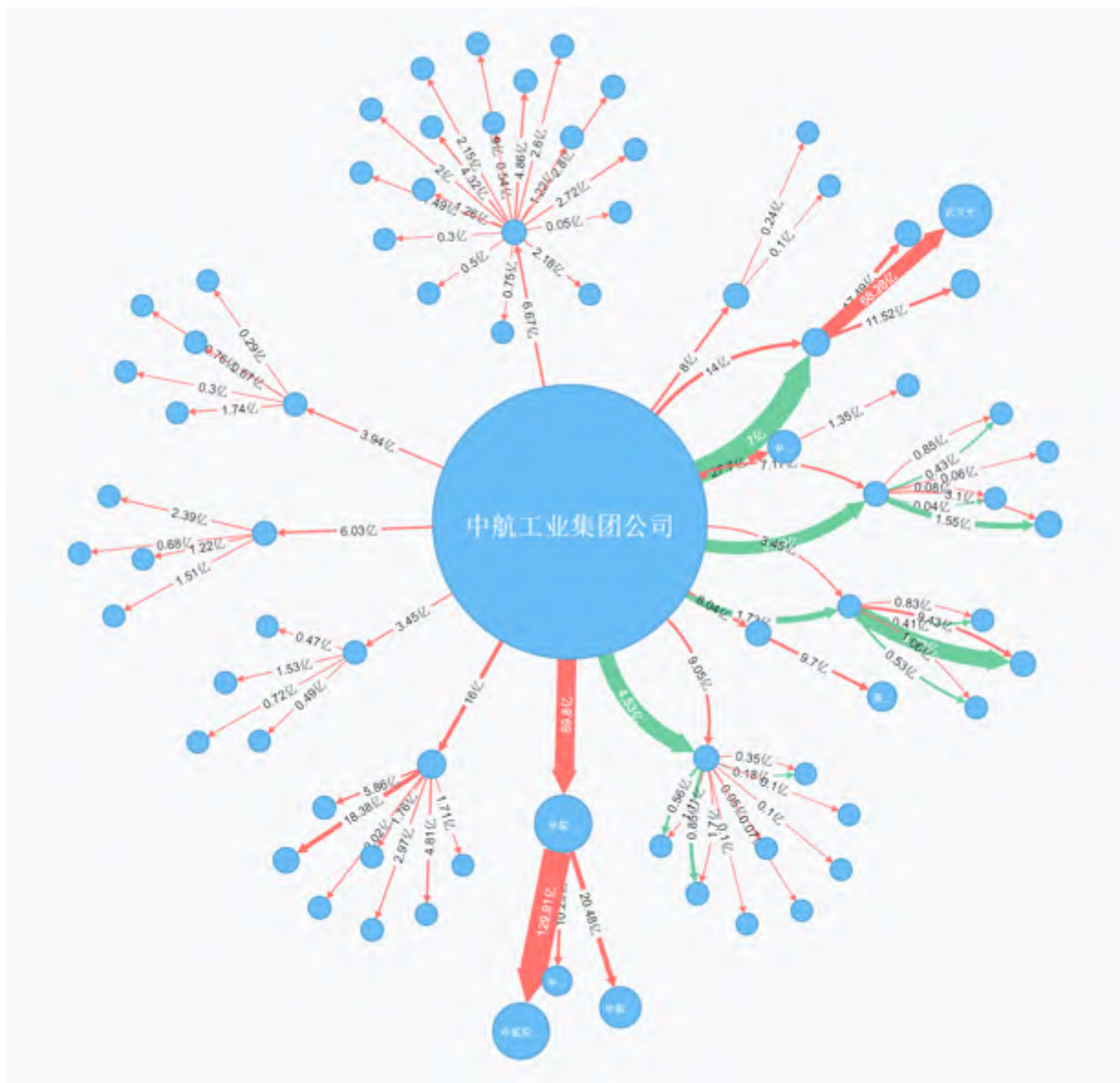
监控系统

每一个优秀的图都开始于 Neo4j.

国产化进程 - 图片化



国产化进程 - 数据驱动



国产化进程 - 智能查询

The screenshot displays the Neo4j web interface in a browser window. The address bar shows the URL `we-yun.com:7070`. On the left, a search sidebar is active with the following fields:

- 开始节点标签: 公司
- 开始节点内容: CMB
- 结束节点标签: 产品
- 结束节点内容: Neo4j
- 操作: `MATCH (n:'公司') WHERE RE`

The main query editor contains the Cypher query: `$ MATCH (n:'公司') WHERE REDUCE(hit...`. The results summary shows:

- *(4) 个人(1) 产品(1) 公司(2)
- *(3) 代理(1) 员工(1) 潜在客户(1)

The graph visualization shows a network of nodes and relationships:

- A yellow circle node labeled "neo4j".
- A blue cloud icon node labeled "代理" (Agency).
- A red circle node labeled "员工" (Employee).
- A red circle node labeled "潜在客户" (Potential Customer).
- A red circle node with a white "M" logo.

Relationships are shown as arrows: "代理" points to "neo4j", "代理" points to "员工", "员工" points to "潜在客户", and "潜在客户" points to the "M" logo node.

ToNeo4j 1.0, Copyright © 2016 微云数聚 (北京) 科技

Neo4j 服务器: 显示细节

用户名:

密码:

Label : <标签>
ID : 名称:<名称>
Relationship:
Label_T :
ID_T :
Property : color:<color>
Cypher :

```
MERGE (n:`<标签>` {`名称`:<名称>})  
CREATE INDEX ON `公司` (`名称`)  
MERGE (n:`公司` {`名称`:`微云数聚`}) SET n.color=`blue`, n.image=`image/公司/微云数聚.jpg`  
MERGE (n:`公司` {`名称`:`Neo Technology`}) SET n.color=`blue`, n.image=`image/公司/NeoTechnology.png`  
MERGE (n:`公司` {`名称`:`百分点`}) SET n.color=`blue`, n.image=`image/公司/百分点.png`  
MERGE (n:`公司` {`名称`:`智慧星光`}) SET n.color=`blue`, n.image=`image/公司/智慧星光.png`
```

完成 26/26

MySQL 设置

服务器:

用户名:

密码:

端口:

数据库:

表:



公司

Neo4j 在中国的战略合作伙伴
由大数据领域博士团队和移动互联网技术精英团队组成



产品

ADI, 敏捷数据智能平台, 傻瓜式BI产品, 已广泛用于互联网金融

Neo4j 国产化, 包括界面本地化、节点图片化、数据驱动化、查询智能化和导入工具化等



合作伙伴

华为、中国首席数据官联盟、瀚莎、百分点、智慧星光、数据堂等



用户

招商银行、人民银行、中科院、中国电信、中国工商总局、互联网金融公司等

THANK YOU !



we-yun.com
we-yun.com:7070