

2018PostgreSQL中国技术大会



图数据库及应用场景

邵宗文

腾讯云数据库 - 邵宗文

个人简介

邵宗文, CCF,IEEE会员,2016年开始学人工智能, 2017年6月份带队获得搜狐图文匹配大赛第10名。

2009年加入腾讯, 现为CSIG腾讯云数据库专家产品经理。之前曾负责为OMG事业群构建数据库平台, 部署, 规划及运维支持,为腾讯网, 新闻客户端, 快报, 视频, 财经, 体育, 等提供了稳定的服务。

之前曾任新浪数据库专家, 数据库平台主管。有非常丰富的大型项目的经验: 如统一通行证, 发布系统, 论坛, 财经, 体育等重大项目的数据库架构改造和性能优化。



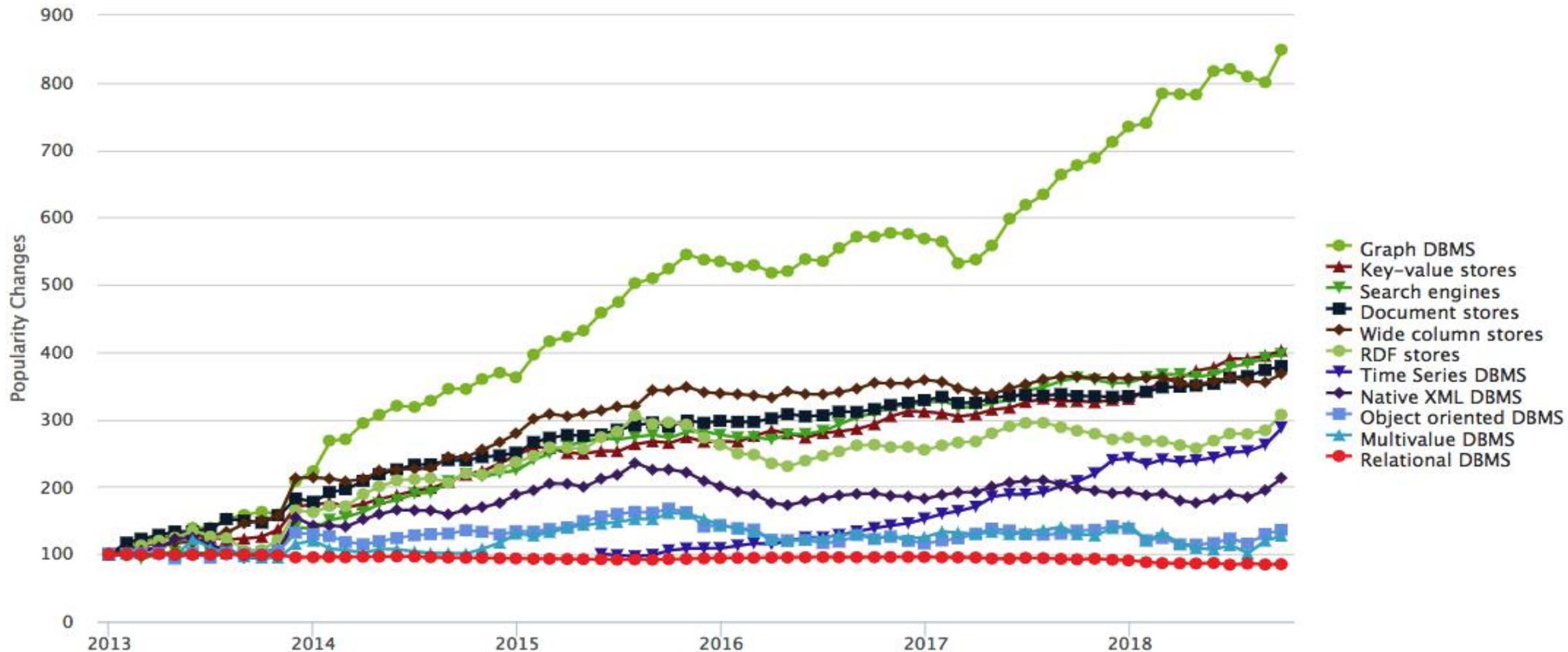


目录

1. 市场分析
2. 应用分析
3. 优劣对比
4. 行业展望



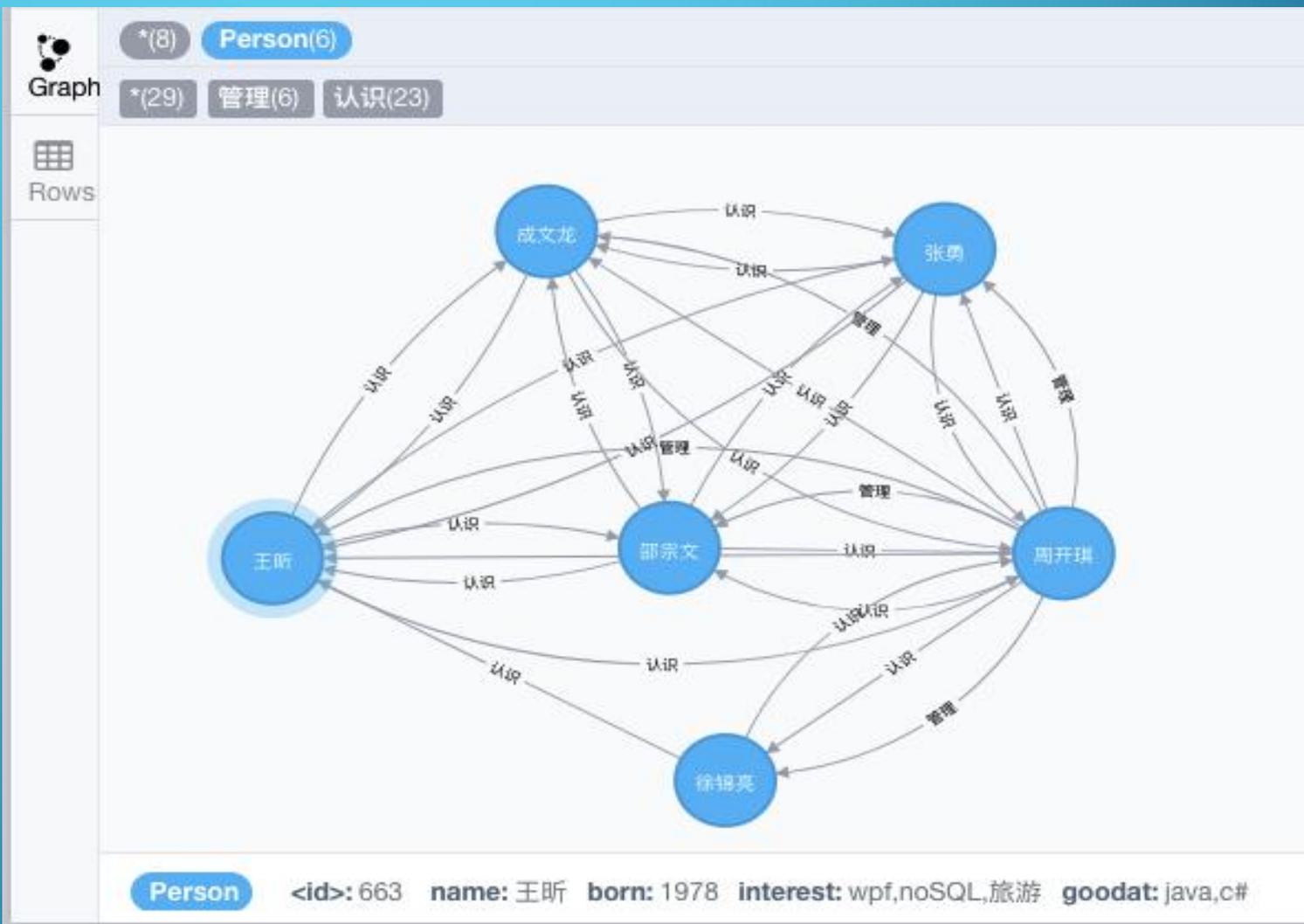
1. 市场分析

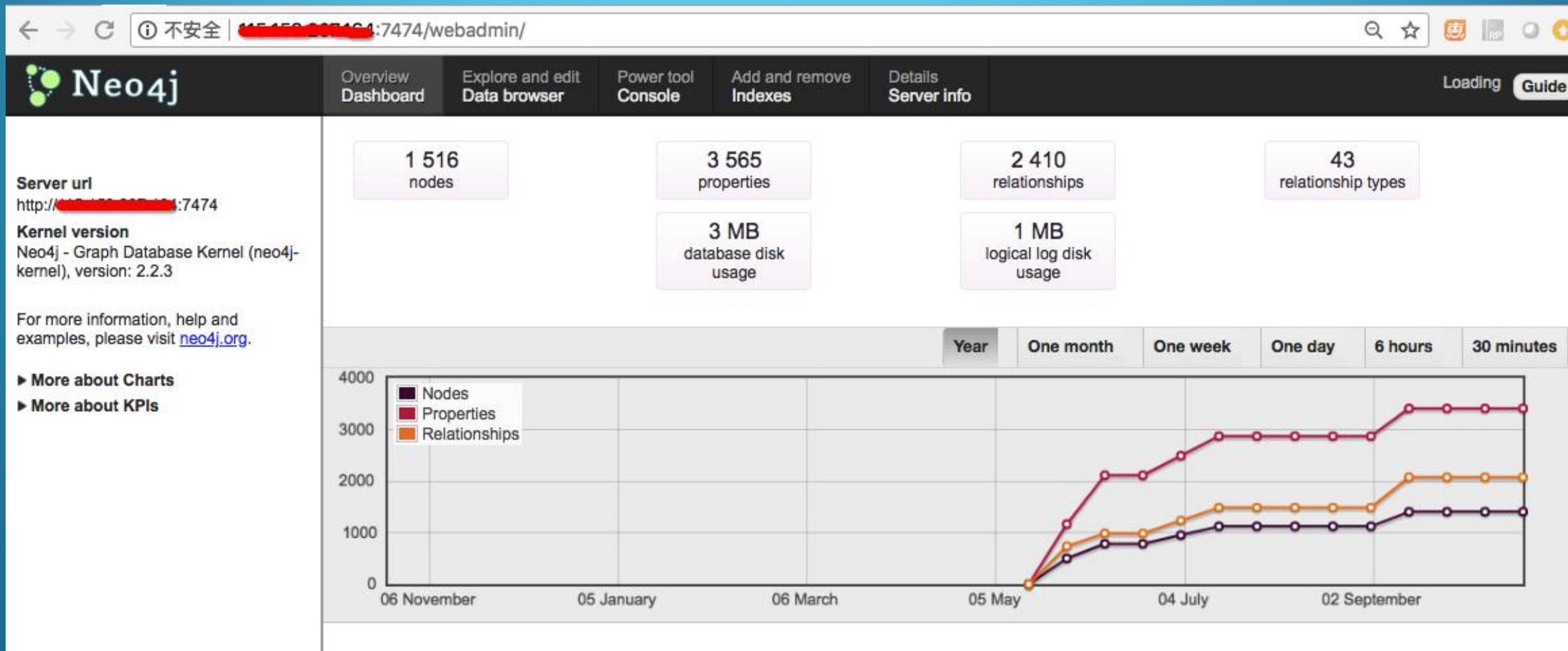




一图胜过千言万语

比起传统的信息存储和组织模式，图数据库能够很清晰揭示复杂的模式，尤其在错综复杂的社交，物流，金融风控行业效果更为明显。







Operational 图数据库

Neo4j

Titan

JanusGraph

OrientDB

RDF图数据库

AllegroGraph

Blazegraph

Virtuoso

Stardog

多模式图数据

Azure Cosmos DB

ArangoDB

Sqrrl

RedisGraph

分析及大图数据库

Apache Giraph

Turi

TigerGraph

CayLey

Dgraph



1 国外的图数据库情况



2017年11月底, 亚马逊在 AWS 2017 全球峰会发布了全新的图数据库产品: Neptune.

2017年7月, SQL Server 2017 rc1发布, 引入了图数据库的支持, 2017年10月正式发布。

2017年3月, Oracle 12.2开始支持属性图和图分析

2 国内图数据库情况

2017年10月, 华为云推出了图引擎服务

阿里云蚂蚁金服Geabase图数据库已开始宣传, 但还未在阿里云发布

2018年8月, 百度推出了hugeGraph图数据库



图数据库名称	Neo4j	Orientdb	Arangodb	dgraph	janusGraph
语言	java	java	c++	go	java
version	3.3.5 (April 2018)	2.2.24 (July 2017)	3.3.11 / (June 28, 2018)	1.09 (Oct 2018)	0.3.1 (October 2, 2018)
stars数	5546	3624	6506	6501	1798



从下载数量来看还是前景很大
有7M是基于下的neo4j的docker版



通过多年培训和逐步推广
目前大概有5万多名



图数据库相关技术会议
GraphConnect大会已有上千人参会。



2. 应用分析



社交网络

金融风控

知识图谱

关系分析

车联网应用

网络&IT运维

推荐引擎

社交网络

数十亿个关系查询延迟降低到秒级，目前传统关系型数据库基本无法实现超过3层好友关系的查询。

推荐引擎

通过用户兴趣、好友和阅读历史记录等信息之间的关系，向用户提供推荐。

电商，短视频，新零售等都有强烈需求。

网络&IT运维

基础设备规模庞大、结构复杂，帮助深入了解设备状态、设备之间的关系，实现全网络设备智能监控与管理

金融风控

提供实时的用户行为检测，识别敏感用户，及时识别欺诈风险，错综复杂的人物关系分析，进行用户分群，识别异常群体



银行风险压力

- 合规监管风险
- 黑产欺诈风险
- 内部员工风险

利用图谱网络开展风险管理

员工管理

- 员工与亲属、员工与外部企业

客户管理

- 财务状况、征信、行业、资金流动性、重大事件

关系管理

- 亲属关系、股权关系、集团关系、业务关系、供应链、行业关系

业务管理

- 业务流程、业务合规、业务资金、业务进展、业务数据

如何向用户推荐最符合需求和兴趣的聚会活动A

Meetup活动推荐 – 推理和发现



- 无效的推荐(已经参加的活动)


```
graph LR; User((User)) -- [RSVP] --> Event((Event))
```
- 有效的推荐: 匹配感兴趣的主题


```
graph LR; User((User)) -- [INTERESTED] --> Topic1((Topic)); Group((Group)) -- [HAS_TOPIC] --> Topic1; Group -- [HOSTS_EVENT] --> Event((Event))
```
- 接近度: 越远的关系相关度越小


```
graph LR; User((User)) -- [INTERESTED] --> Topic1((Topic)); Group1((Group)) -- [HAS_TOPIC] --> Topic1; Group1 -- [HAS_TOPIC] --> Topic2((Topic)); Group2((Group)) -- [HAS_TOPIC] --> Topic2; Group2 -- [HAS_TOPIC] --> Topic3((Topic)); Group3((Group)) -- [HOSTS_EVENT] --> Event((Event))
```

如何向用户推荐最符合需求和兴趣的聚会活动B

Meetup活动推荐 – 推理和发现(续)



注：以下的查询均用Cypher 语言实现。

- 发现最热门的主题

```
1 MATCH (t:Topic)<-[:HAS_TOPIC]-()  
2 RETURN t.name, COUNT(*) AS count  
3 ORDER BY count DESC
```



- 发现已经参加的聚会

```
1 MATCH (m:User)-[:MEMBER_OF]->(group)  
2 WHERE m.name = {name}  
3 RETURN group
```



如何向用户推荐最符合需求和兴趣的聚会活动C

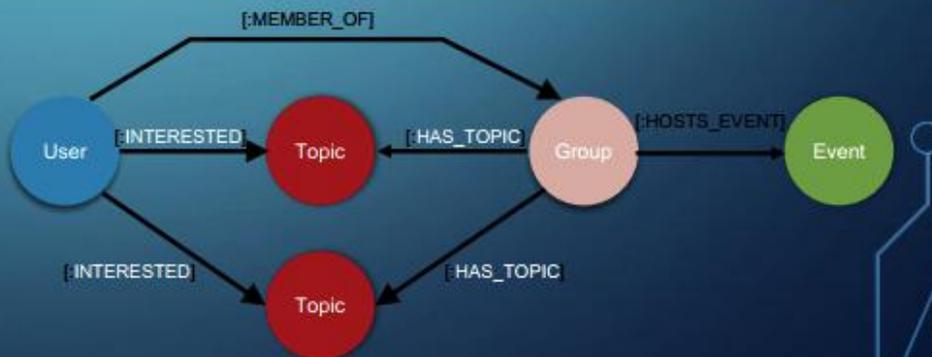
Meetup活动推荐 – 推理和发现(续)



推荐最可能感兴趣的活动

```
1 MATCH (member:Member) WHERE member.name CONTAINS {name}
2 MATCH (futureEvent:Event) WHERE futureEvent.time > timestamp()
3
4 WITH member, futureEvent, EXISTS((member)-[:MEMBER_OF]->(:HOSTED_EVENT)->(futureEvent)) AS myGroup
5 OPTIONAL MATCH (member)-[:INTERESTED_IN]->(:HAS_TOPIC)-(:HOSTED_EVENT)->(futureEvent)
6
7 WITH member, futureEvent, myGroup, COUNT(*) AS commonTopics
8 MATCH (futureEvent)<-[:HOSTED_EVENT]-(group)
9
10 RETURN futureEvent.name, futureEvent.time, group.name, commonTopics, myGroup
11 ORDER BY futureEvent.time
```

1. 找到所有未来的活动;
2. 看看这些活动是不是属于已经加入的兴趣组;
3. 统计共同的兴趣主题;
4. 找到那些组织这些活动的兴趣组;
5. 按照时间对结果排序。



如何向用户推荐最符合需求和兴趣的聚会活动D

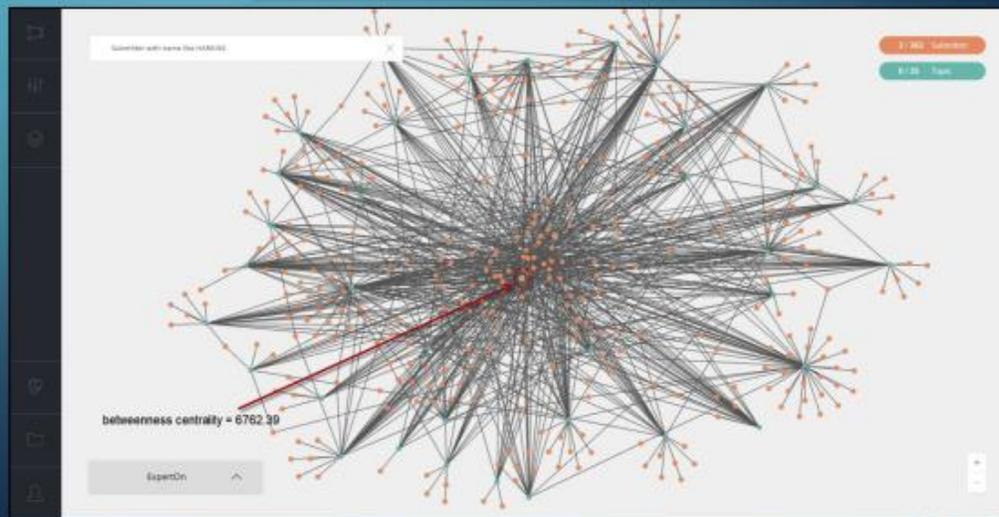
Meetup活动推荐 – 进一步的改进(续)



发现“公众人物” / “领域专家”

- 运用中心性算法寻找“公众人物”
- 推荐公众人物的选择
- 专家对活动的评价

```
1 MATCH (node:User)
2 WITH collect(node) AS nodes
3 CALL apoc.algo.betweenness(['RSVP'], nodes, 'BOTH')
4   YIELD node, score
5 RETURN id(node), node.name, score
6 ORDER BY score DESC
```





3. 优劣对比



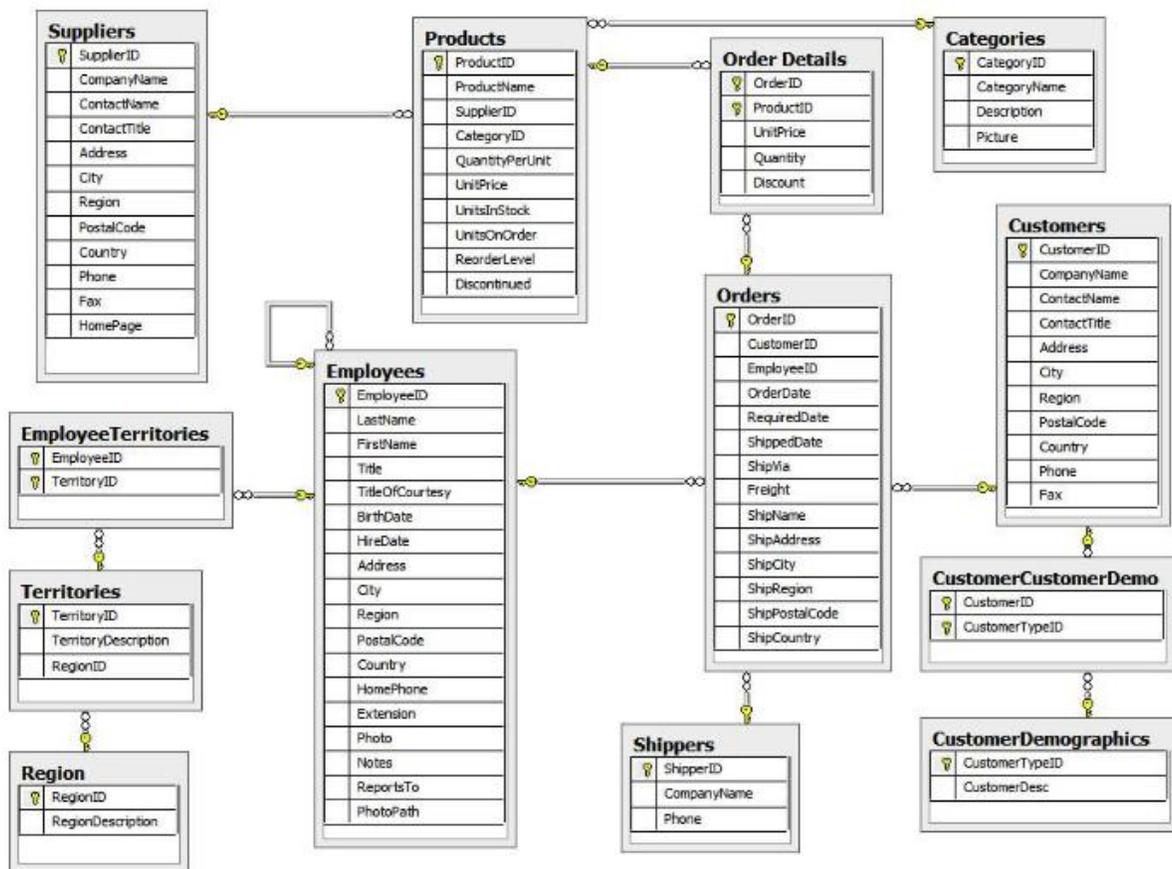
图数据库在处理关联关系上具有完全的优势，特别是在我们这个社交网络得到极大发展的互联网时代。例如我们希望知道谁LIKES（喜欢）谁（喜欢可以是单向或双向），也想知道谁是谁的FRIEND_OF（朋友），谁是所有人的LEADER_OF（领导）。除了在关联查询中尤为明显的优越性，图数据库还有如下优势：

- a) 用户可以面向对象的思考，用户使用的每个查询都有显式语义；
- b) 用户可以实时更新和查询图数据库；
- c) 图数据库可以灵活应对海量的关系变化，如增加删除关系、实体等；
- d) 图数据库有利于实时的大数据挖掘结果可视化。

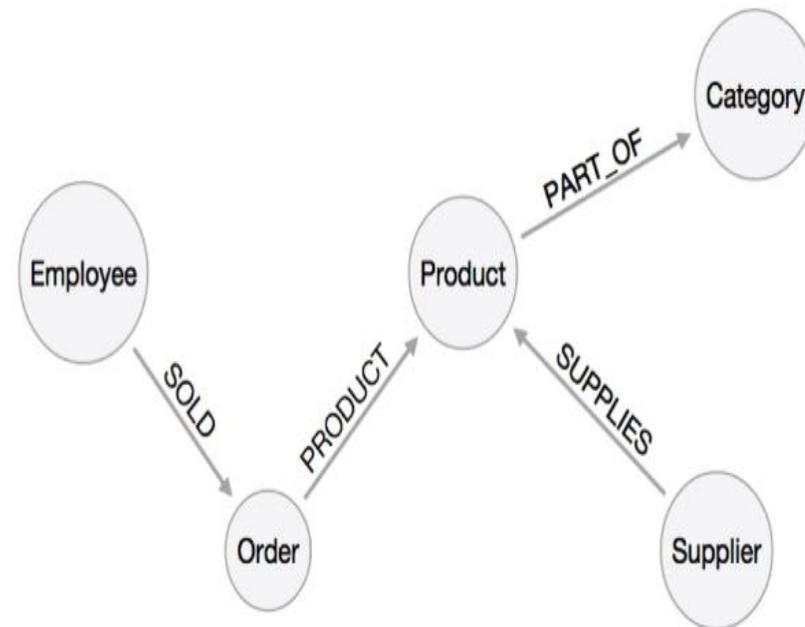
图数据库虽然弥补了很多关系型数据库的缺陷，但还有一些不足地方，如：

- a) 不适合记录大量基于事件的数据（例如日志条目）；
- b) 二进制数据存储。
- c) 并发性能要求高的项目。
- d) 目前相关图查询语言比较多，尚未有很好统一。
- e) 图数据库相关的一些书籍文档偏少。
- f) 相关生态还在不断完善。

Relational model



Graph Model



<https://neo4j.com/developer/guide-sql-to-cypher/>



```
SELECT p.*FROM products as p;
```

```
MATCH(p:Product)RETURNp;
```

```
SELECT p.ProductName,  
p.UnitPriceFROM products as pORDER  
BY p.UnitPrice DESC LIMIT 10;
```

```
MATCH(p:Product)  
RETURNp.productName,p.unitPrice ORDERB  
Yp.unitPrice DESC LIMIT10;
```

```
SELECT p.ProductName, p.UnitPriceFROM  
products AS pWHERE p.ProductName  
LIKE 'C%' AND p.UnitPrice>100;
```

```
MATCH(p:Product)WHEREp.productNameST  
ARTS  
WITH"C"ANDp.unitPrice>100RETURNp.prod  
uctName,p.unitPrice;
```



```
SELECT DISTINCT c.CompanyNameFROM
customers AS cJOIN orders AS o ON
(c.CustomerID = o.CustomerID)JOIN order_details
AS od ON (o.OrderID = od.OrderID)JOIN products
AS p ON (od.ProductID = p.ProductID)WHERE
p.ProductName = 'Chocolate';
```

```
SELECT p.ProductNameFROM Product AS pJOIN
ProductCategory pc ON (p.CategoryID = pc.CategoryID
AND pc.CategoryName = "Dairy Products")JOIN
ProductCategory pc1 ON (p.CategoryID =
pc1.CategoryIDJOIN ProductCategory pc2 ON
(pc2.ParentID = pc2.CategoryID AND pc2.CategoryName
= "Dairy Products")JOIN ProductCategory pc3 ON
(p.CategoryID = pc3.CategoryIDJOIN ProductCategory
pc4 ON (pc3.ParentID = pc4.CategoryID)JOIN
ProductCategory pc5 ON (pc4.ParentID = pc5.CategoryID
AND pc5.CategoryName = "Dairy Products");
```

```
MATCH(p:Product{productName:"Chocolate
"})<-[[:PRODUCT]]-([:Order])<-[[:PURCHASED]]-
(c:Customer)RETURNdistinctc.companynam
e;
```

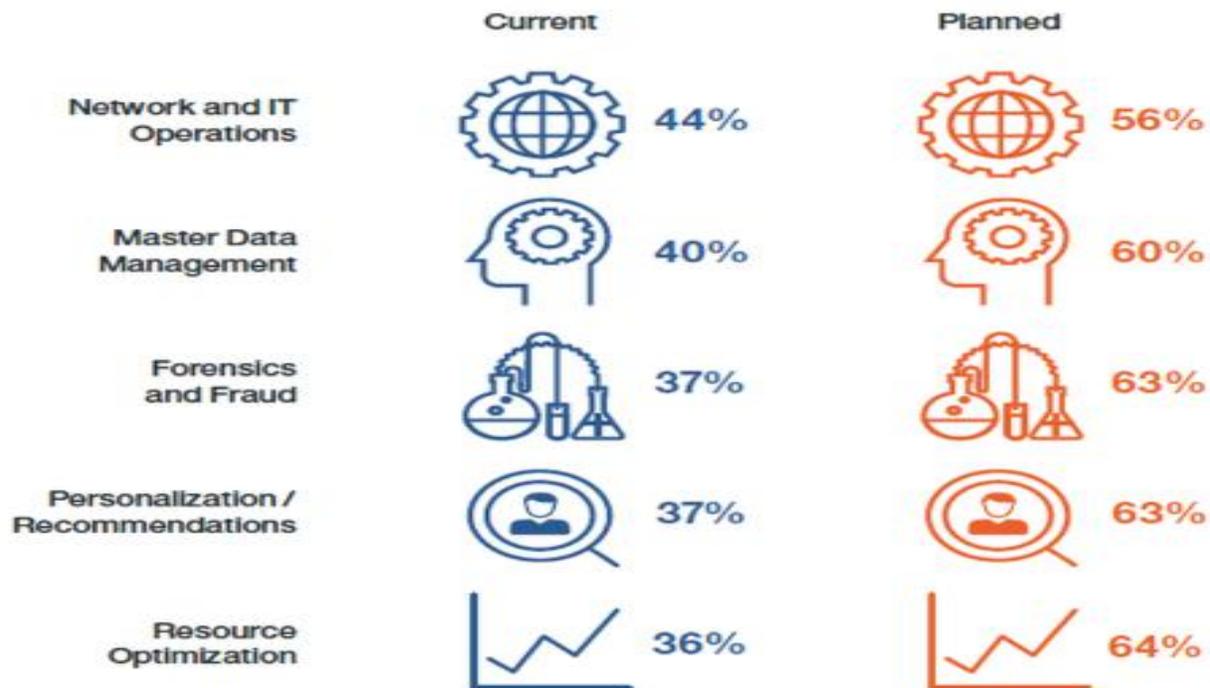
```
MATCH(p:Product)-[:CATEGORY]-
>(l:ProductCategory)-[:PARENT*0..]-
(:ProductCategory{name:"Dairy
Products"})RETURNp.name
```



4. 行业展望

Current and Planned Uses of Graph Databases

In what ways is your organization using (or planning to use) graph technology?



Source: TechValidate survey of 1125 users of Graph Database TVID: 965-47B-F88

<https://www.zdnet.com/article/the-rise-and-rise-of-graph-databases/>



76%

FORTUNE 100

have adopted or
piloted Neo4j



Finance

20 of top 25



Software

7 of top 10



Logistics

3 of top 5



Retail

7 of top 10



Airlines

3 of top 5



Telco

4 of top 5



Hospitality

3 of top 5



我们的公众号：腾讯云数据库CDB

Thanks

