

Neo4j图数据库概述

大数据时代的数据库平台

俞方桦 博士

Joshua.yu@neo4j.com

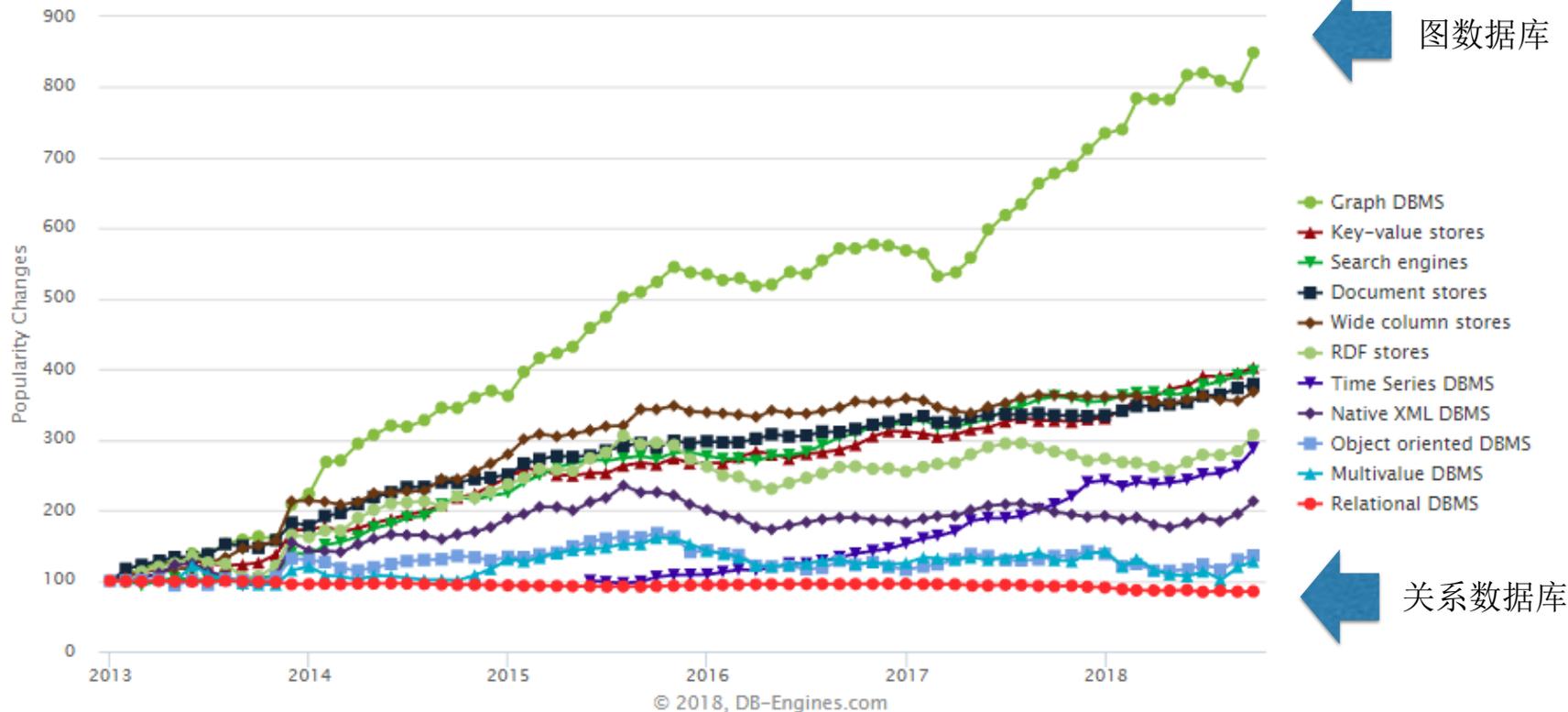
Neo4j APAC, 2019.

All rights reserved. 2018 - 2019



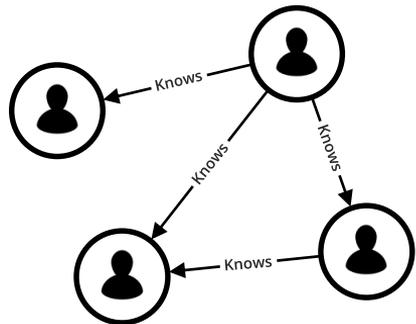
图数据库是近年来增长最快的数据库技术

Complete trend, starting with January 2013



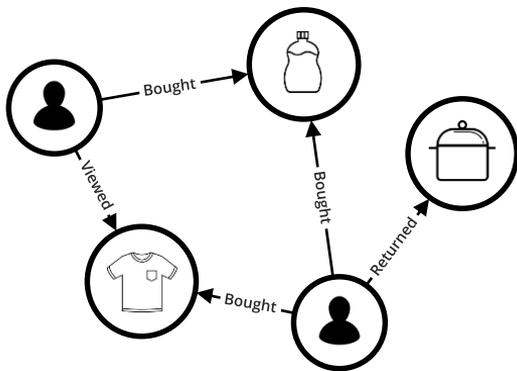
大数据是内在关联的数据

数据之间的内在关联与数据的规模同步增长



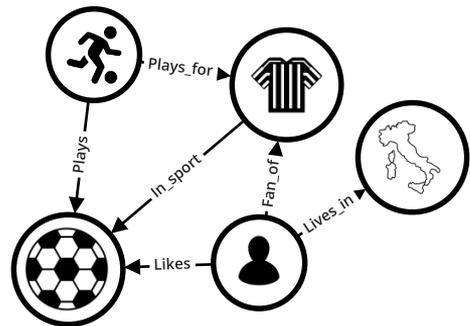
社交网络：人群关联

员工、顾客、供应商、合作伙伴、具有影响力的人群等等。



业务流程：事件关联

风险管理、供应链管理、支付处理等等。

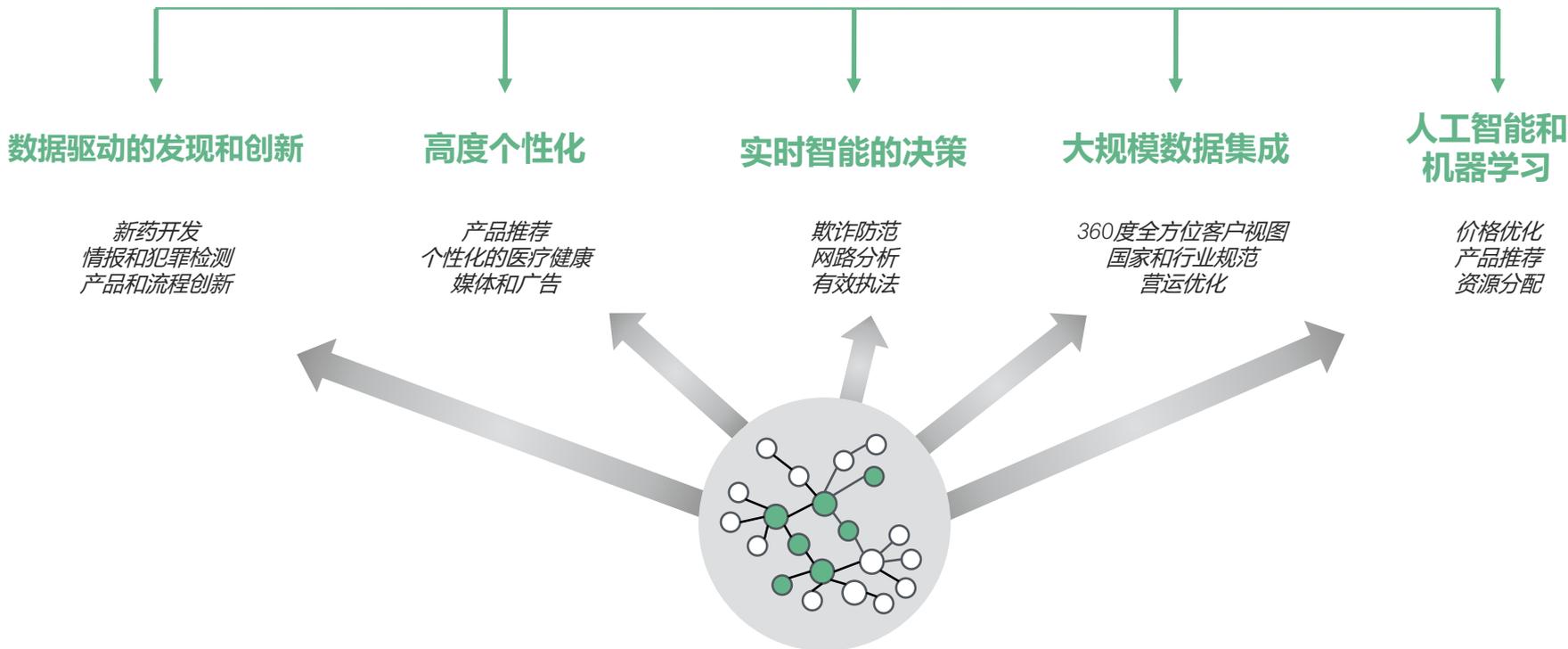


内容网络：知识关联

企业内容管理、领域相关内容、电子商务等等。

创造新的价值需要创新性地挖掘数据之间的关联

企业数字化转型的主流趋势



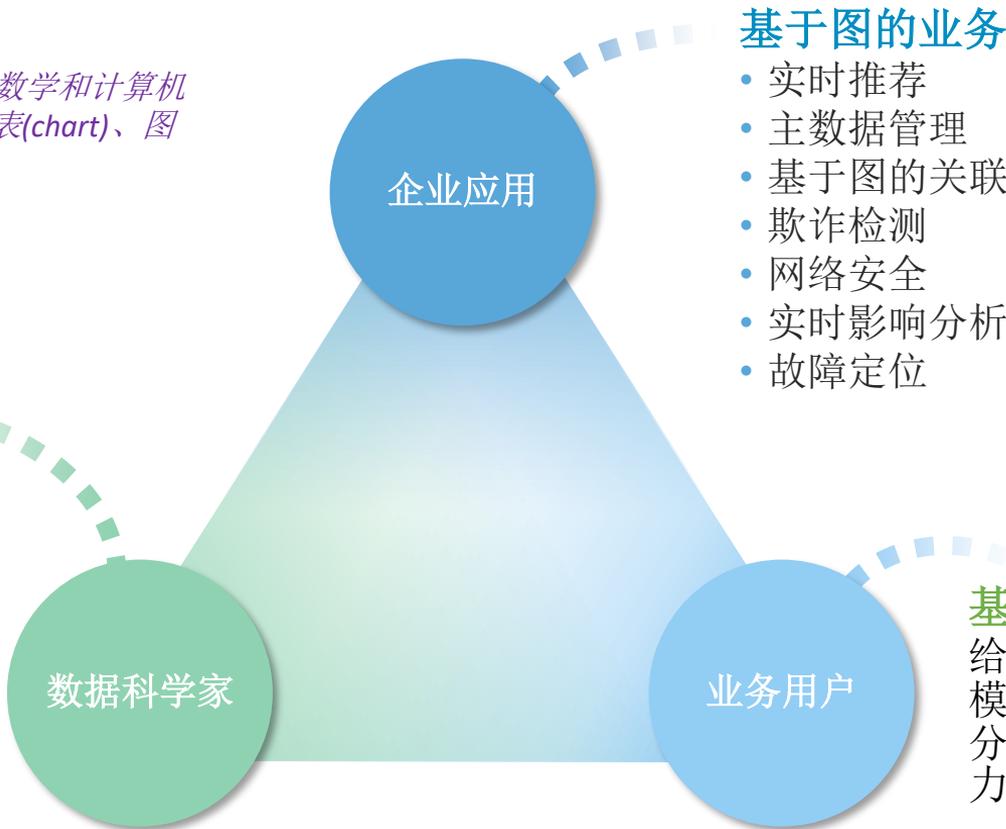
数据和数据之间的关联是成功转型的重要基础

图数据库是分析数据间关联的最佳技术

注：这里的图(Graph)指的是数学和计算机科学中的“图论”，而非图表(chart)、图片(picture)。

基于图的分析

- 舆情分析
- 客户细分
- 机器学习
- 识别型计算
- 社区检测



基于图的业务应用

- 实时推荐
- 主数据管理
- 基于图的关联搜索
- 欺诈检测
- 网络安全
- 实时影响分析
- 故障定位

基于图的可视化

给予业务用户针对大规模关系网络图的可视化分析界面，提高决策能力、缩短决策时间。

Neo4j图数据库平台架构

Neo4j图数据库平台(版本3.4)



应用



业务用户

驱动和API
(Python, Java, JavaScript, R ...)

发现和可视化

Neo4j
Bloom

Neo4j
Desktop



开发、部署和
管理



交易型图数
据处理



- 原生的图数据库
- Cypher查询语言
- 集群和高可用性

Neo4j
Browser



分析型图数
据处理



分析工具

Neo4j
ETL



流数据 /
实时更新



ETL / 批量更新

数据集成



开发人员



管理员

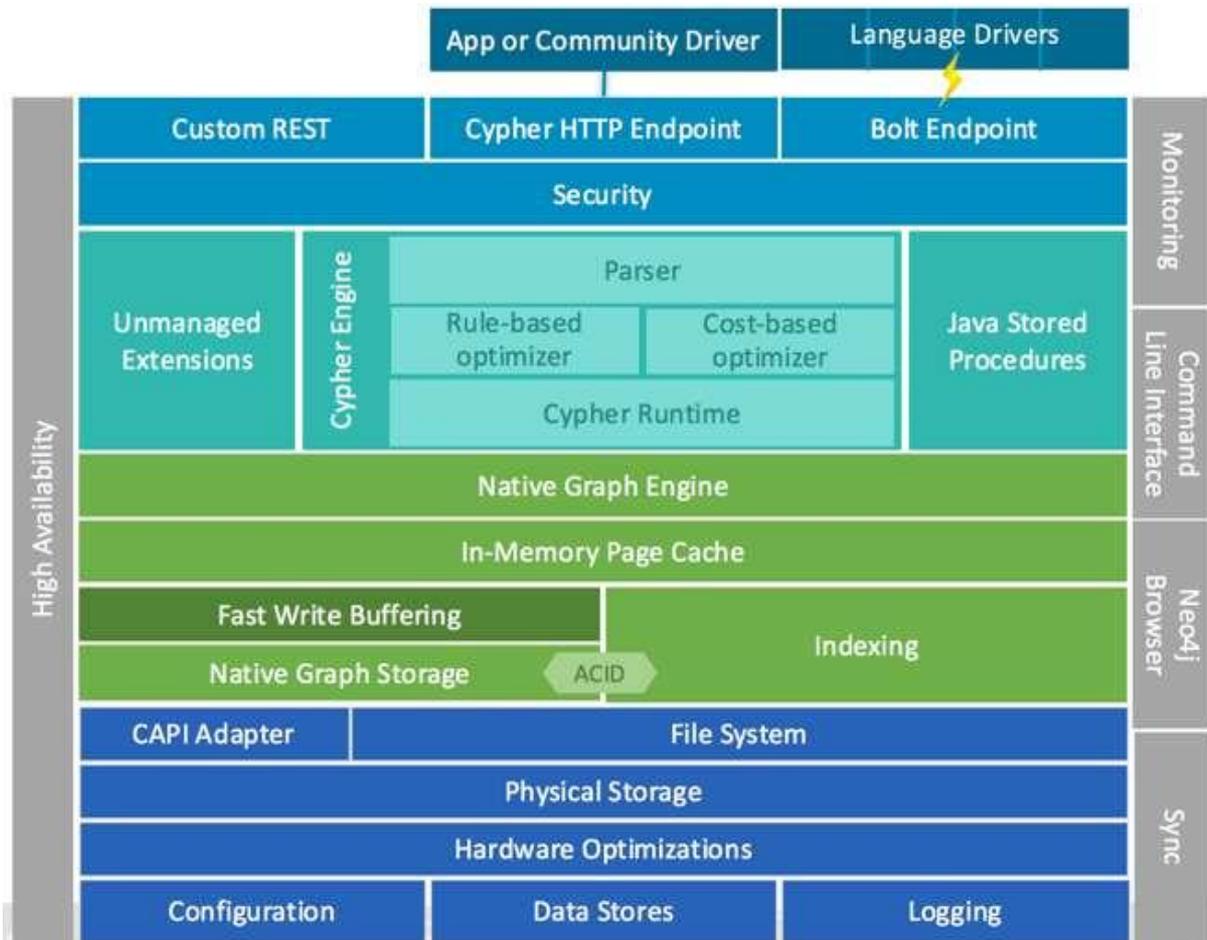


数据分析师

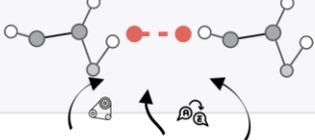


数据科学家

Neo4j图数据库平台应用架构 (版本3.4)

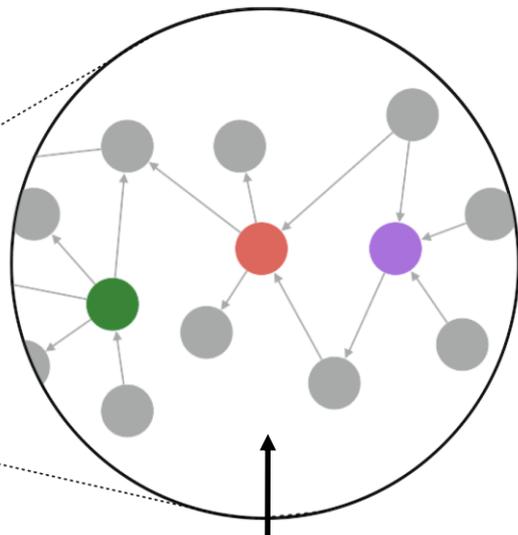
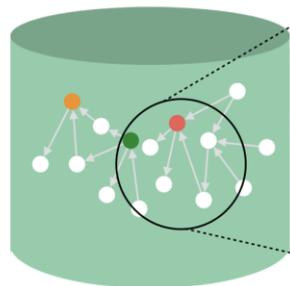


Neo4j与其他数据库的主要区别 – 原生图数据库

	原生图数据库	非原生图数据库	关系型数据库
可视化			
查询语言	Cypher <code>(graphs)-[are]->(everywhere)</code>	Cypher/Gremlin /Proprietary	SQL
处理			
存储			

专门针对图数据处理而优化

Neo4j的原生化图存储提供最高效的关联查询



无需麻烦的索引，关系遍历查询也可以在极短的时间内完成。

无需索引的关系遍历

与其他任何数据库所不同，Neo4j在存储数据的同时存储数据之间的关联。

“Neo4j的查询性能比我们原先使用的MySQL要快上千倍，而代码长度却只有1/10甚至1/100。”

– Volker Pacher, eBay

性能比较1: Neo4j和关系数据库

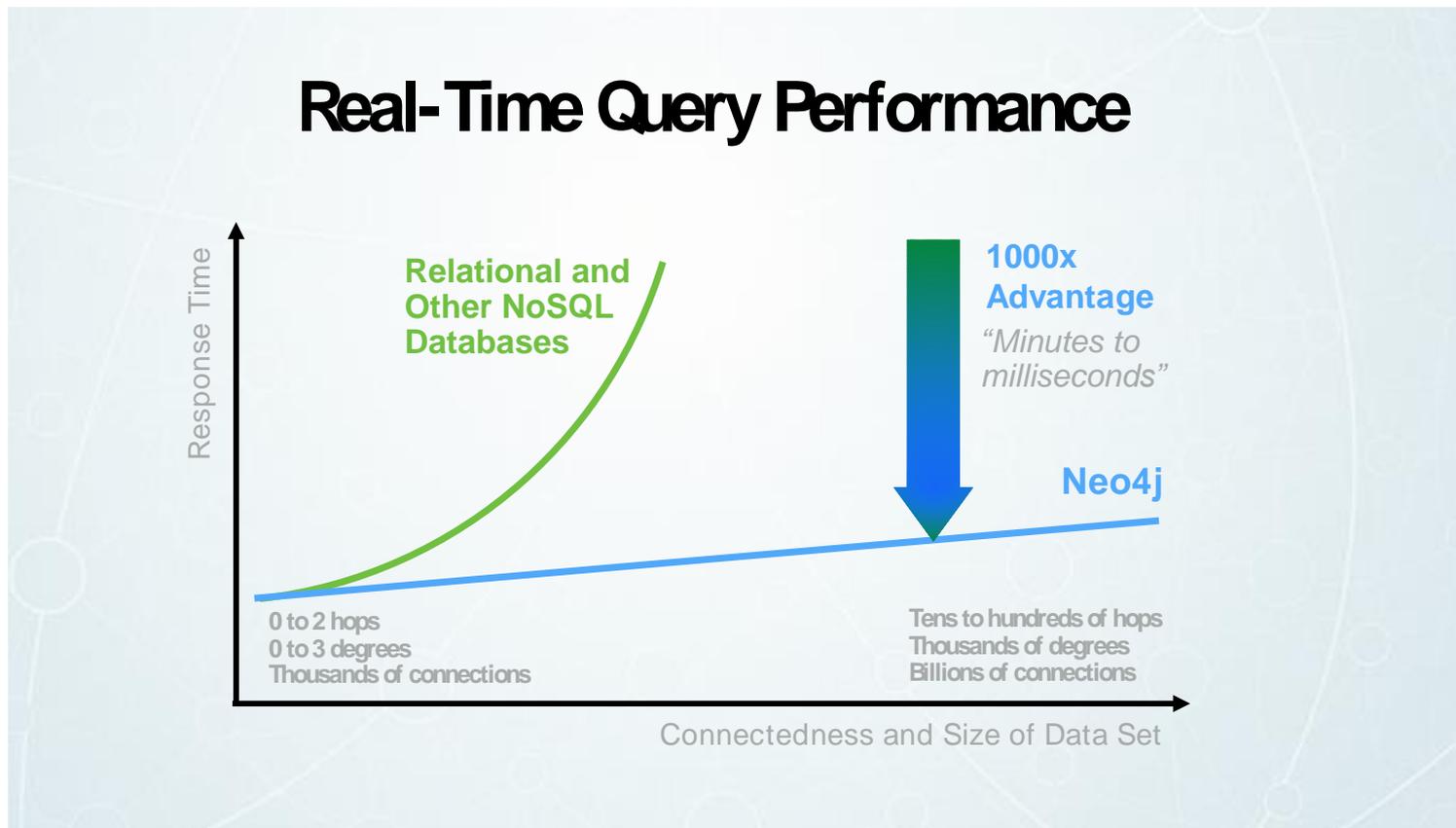
测试用例: 搜索朋友的朋友时的查询响应时间比较

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

测试说明:

- 1百万成员的社交网络
- 平均每个人有50个朋友

性能比较2：查询性能与数据库规模的关系



Neo4j图数据库平台特点

Neo4j图数据库的特点



实时交易和分析型数据处理

- 业务运行的交易型数据(OLTP)
- 分析型应用的数据负载
- 完全支持ACID



高性能

- 无需索引的关系遍历
- 每秒数百万次的关系节点跳转



灵活性

- 原生化的属性-图模型
- 动态的数据库模式极大地方便了异构数据的存储和处理



关系的发现和可视化

- 交互型的图/网络数据操作
- 将数据以图的方式进行可视化, 更易于发现在浏览彼此孤立的数据表时所无法发现的关联



易于开发、测试和部署

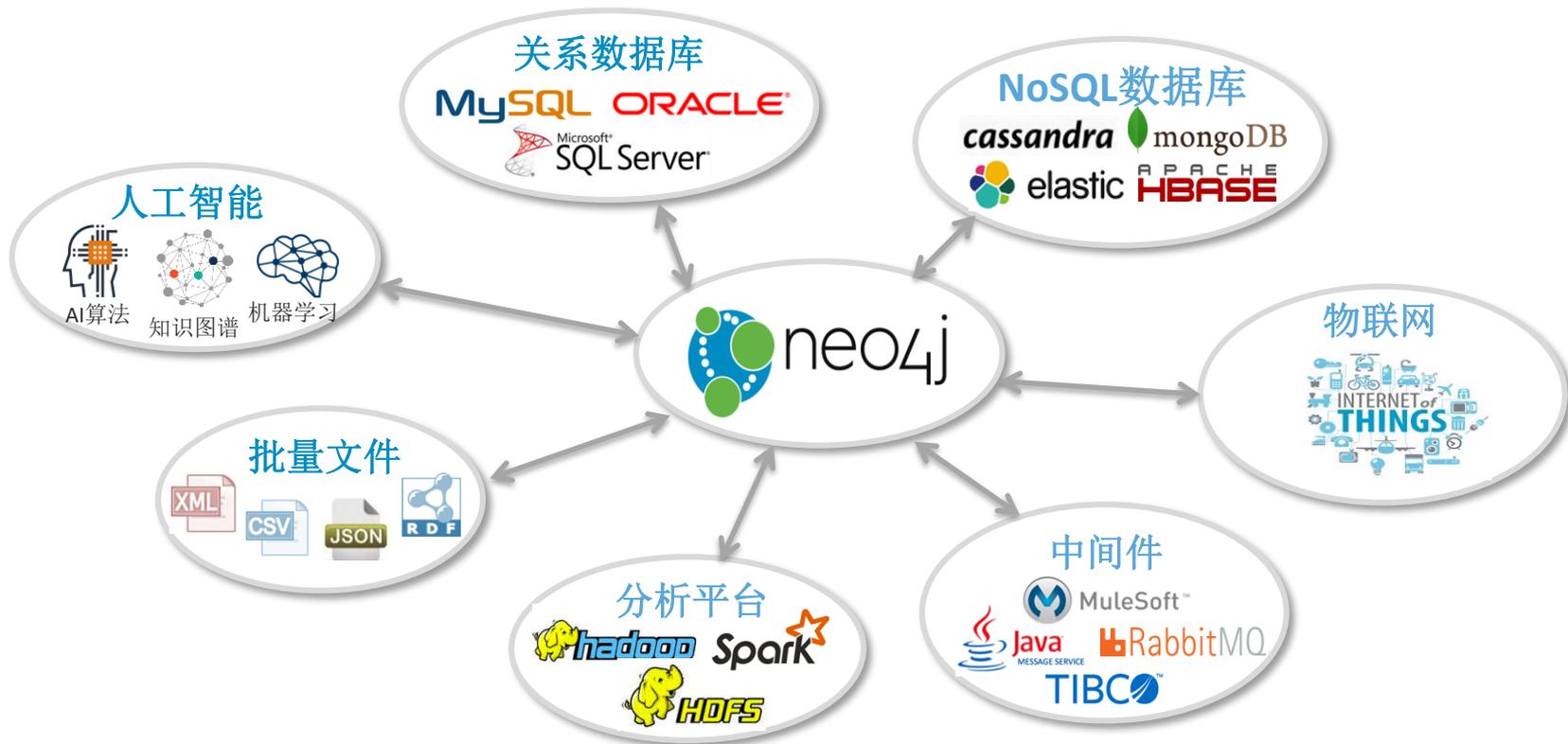
- 独创的Cypher查询语言: 面向数据专家的声明性查询语言
- 支持可扩展的存储过程(Java)
- 最活跃的全球开发和使用者社区



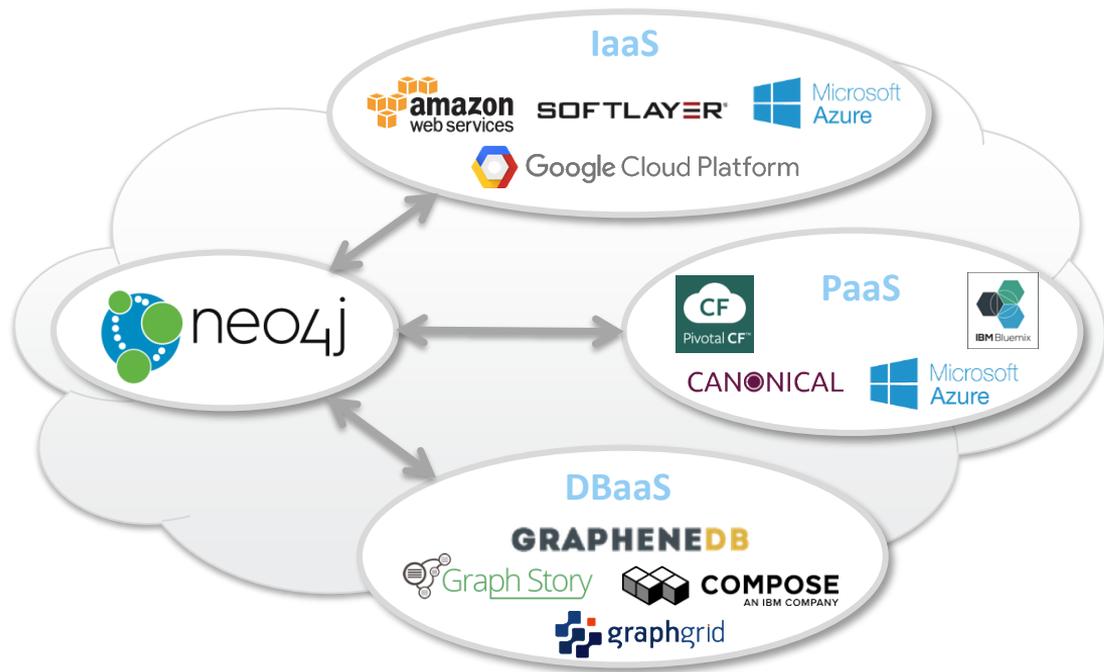
更高效的硬件使用率

- 无需索引的关系遍历能节省90%的CPU
- 相比其他数据库平台, 能节省90%的硬件资源

Neo4j 支持广泛的集成技术和平台



Neo4j 可以部署到任何云平台



Neo4j的下载、安装和配置

Neo4j的版本



Neo4j的图数据库产品有多个版本：

- 社区版 (Community)

开源、免费、单机数据库、非商用环境、社区支持

- 企业版 (Enterprise)

闭源、基于CPU内核的许可证、集群、商业环境、专业支持，丰富的应用支持(例如Bloom)

- 桌面版 (Desktop)

包含限期使用的企业版数据库，以及开发管理工具

Neo4j的安装



- 社区版和企业版

绿色软件、解压ZIP文档到特定目录下即可。

- 桌面版 (Desktop)

Windows和MacOS的安装程序。需要实时Internet连接下载更新。

Neo4j的系统需求



Neo4j的图数据库产品有多个版本：

- 社区版和企业版

JDK 1.8; Windows或Linux。

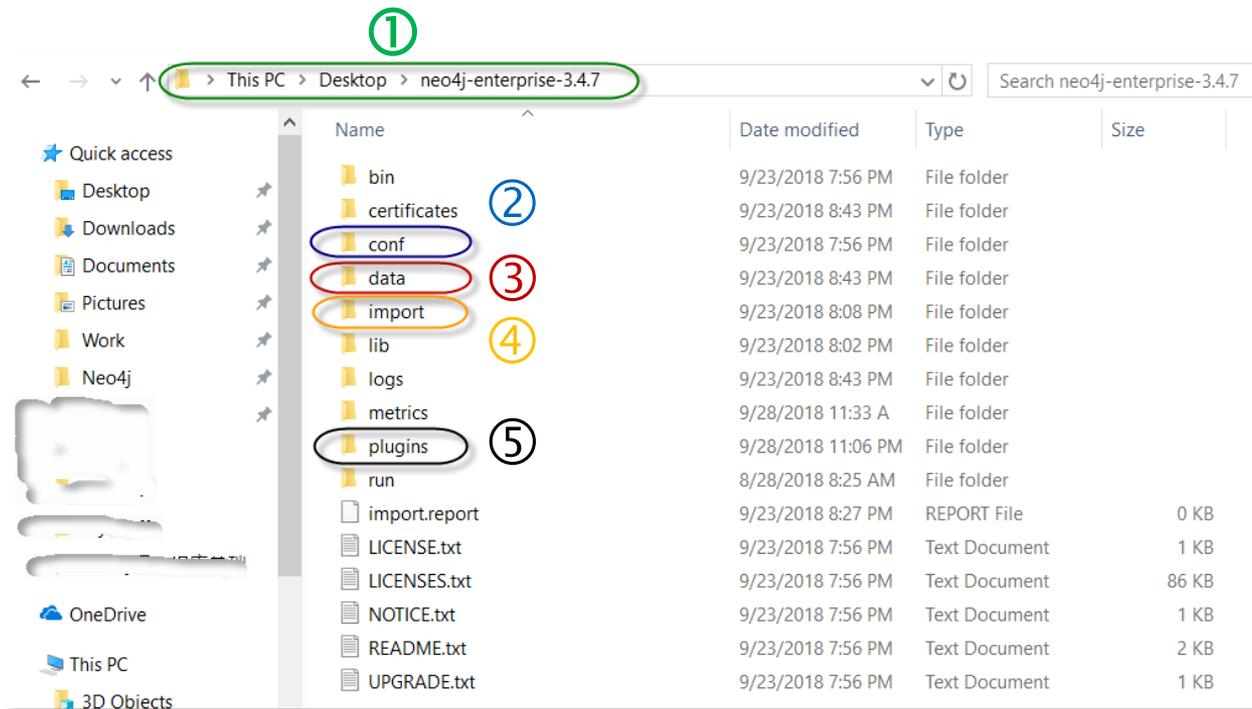
- 桌面版

自带JDK。

- 系统需求(最低):

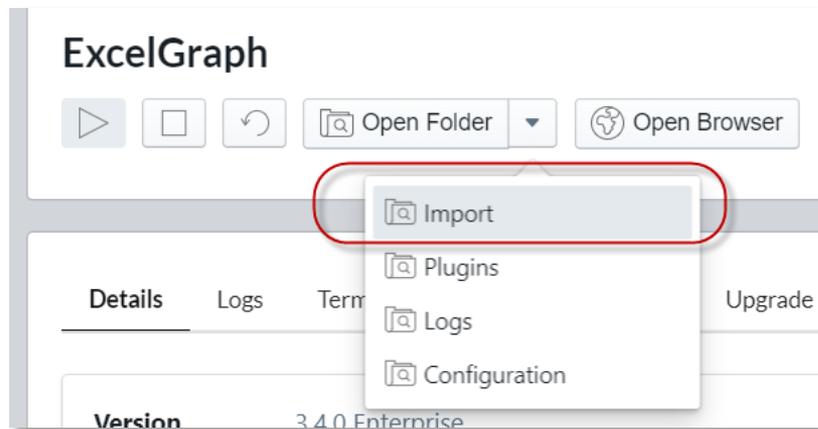
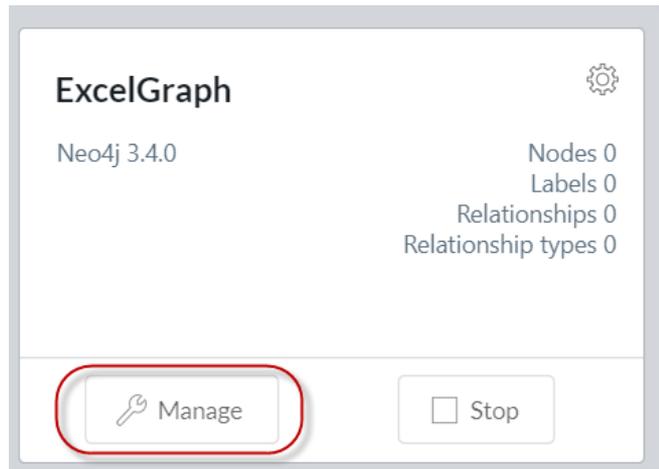
2 CPU内核, 4GB内存, 10GB硬盘空间

Neo4j的目录结构



- ① 安装目录
- ② 配置文件目录
- ③ 数据目录
- ④ 导入文件目录
- ⑤ 服务器端扩展目录

Neo4j的目录结构 – 桌面版



Neo4j的基本配置(单服务器)



<neo4jserver>\conf\neo4j.conf是唯一的配置文件。

缺省数据库名

```
dbms.active_database=graph.db
```

缺省路径

```
dbms.directories.data=C:/Users/xxxx/yyyy/Neo4j
```

Java堆内存配置：数据库更新事务

```
dbms.memory.heap.initial_size=2g
```

```
dbms.memory.heap.max_size=8g
```

Java页缓存配置：查询缓存

```
dbms.memory.pagecache.size=4g
```

客户端连接地址：允许远程访问

```
dbms.connectors.default_listen_address=0.0.0.0
```

Neo4j的启动和停止



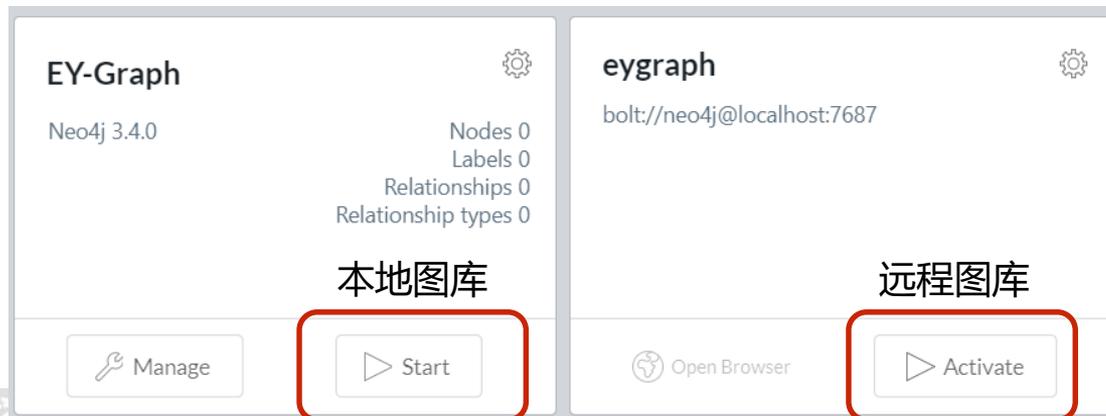
社区版和企业版：

```
<neo4jserver>\bin\neo4j.bat console
```

```
<neo4jserver>\bin\neo4j.bat start
```

```
<neo4jserver>\bin\neo4j.bat stop
```

桌面版：



Neo4j的标签属性图模型

Label Property Graph (LPG)

Neo4j 标签属性(LPG)图模型



- Nodes – 节点。在其他图模型中称作“点”、“顶点”、“对象”。
- Relationships – 关系。在其他图模型中也称作“边”、“弧”、“线”。关系拥有类型。
- Properties – 属性，可以定义在节点和关系上。
- Labels – 标签，代表节点的类别。

标签属性图模型 – 示例



节点

- 代表图中的对象
- 节点可以有一个或多个标签，也可以没有标签



标签属性图模型 – 示例(续)



节点

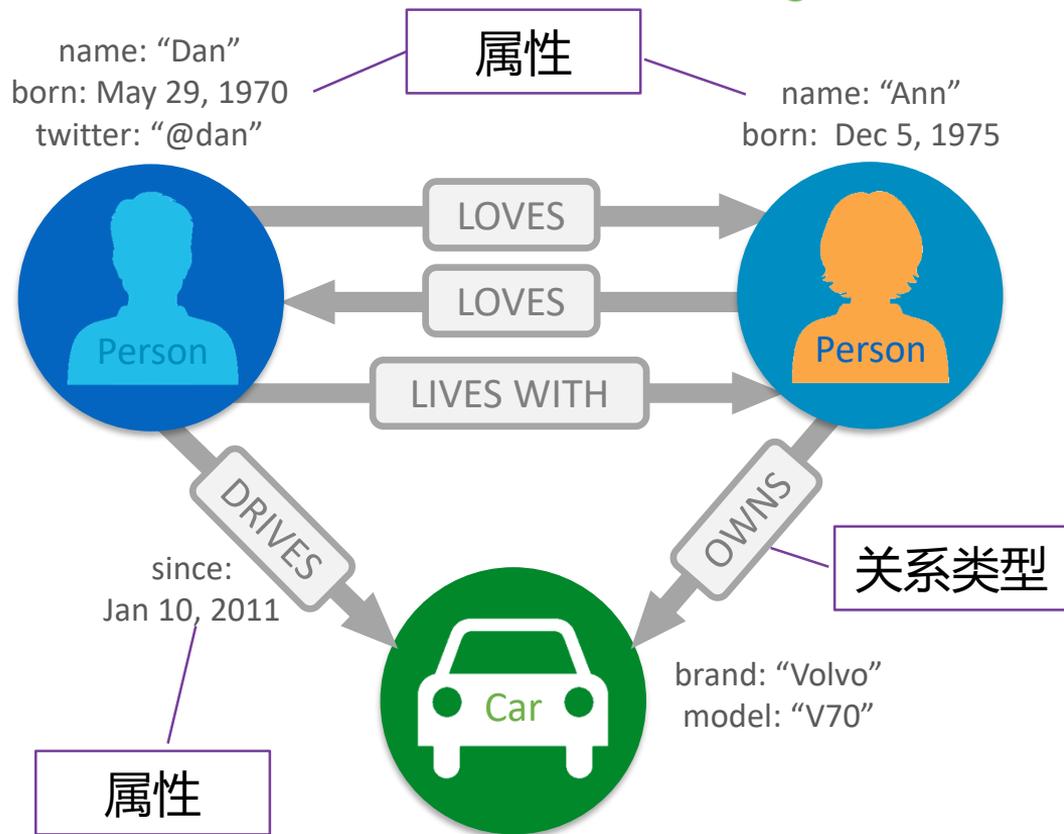
- 代表图中的对象
- 节点可以有一个或多个标签，也可以没有标签

关系

- 关系使用类型和方向将节点连结起来
- 在Neo4j中，关系必须是有向的
- 一个关系只连接两个节点
- 关系必须而且只能有一个类型

属性

- 键-值对，可以在节点和属性上定义。



小结



- **节点** – 实体或者复杂的数据类型
- **关系** – 连结实体、构造领域数据结构
- **属性** – 实体的属性、关系的量值、元数据
- **标签** – 按照角色、类型对节点分组



Cypher - 图数据库查询语言

Cypher – 灵活而且功能强大的图查询语言



- Declarative / 声明型：定义要找的数据，Cypher会决定怎样是最优的查找方法。
- Expressive / 丰富的表达力：易于被开发人员和业务人员理解。
- Pattern Matching / 模式匹配：人类的思维更加容易识别和接受模式。

Cypher: Express Graph Patterns



NODE/
节点



RELATIONSHIP
/关系



节点

```
(:Person { name:'刘备' }) -[:兄长]-> (:Person { name:'关羽' })
```

LABEL/
标签

PROPERTY/
属性

LABEL

PROPERTY

Cypher: CREATE Graph Patterns



节点



关系



节点

```
CREATE (:Person { name:'刘备' }) -[:兄长]-> (:Person { name:'关羽' })
```

LABEL

PROPERTY

LABEL

PROPERTY

Cypher: MATCH Graph Patterns



节点



关系



节点

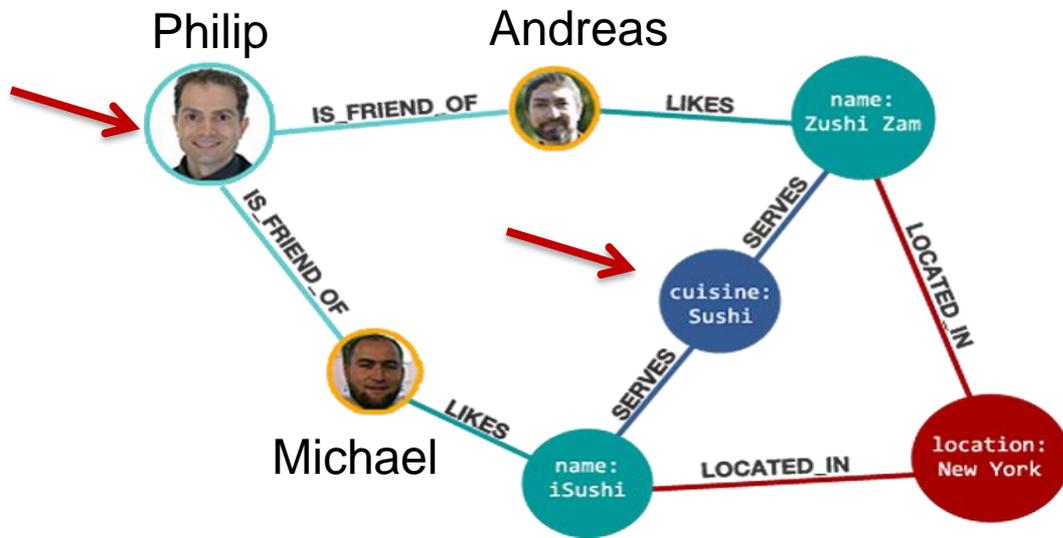
```
MATCH (:Person { name:'刘备' }) -[:兄长]-> ( whom ) RETURN whom
```

LABEL

PROPERTY

VARIABLE

更复杂的例子：根据社交关系提供的推荐



更复杂的例子：根据社交关系提供的推荐(续)



```
MATCH (person:Person)-[:IS_FRIEND_OF]->(friend),  
        (friend)-[:LIKES]->(restaurant),  
        (restaurant)-[:LOCATED_IN]->(loc:Location),  
        (restaurant)-[:SERVES]->(type:Cuisine)
```

```
WHERE person.name = 'Philip'
```

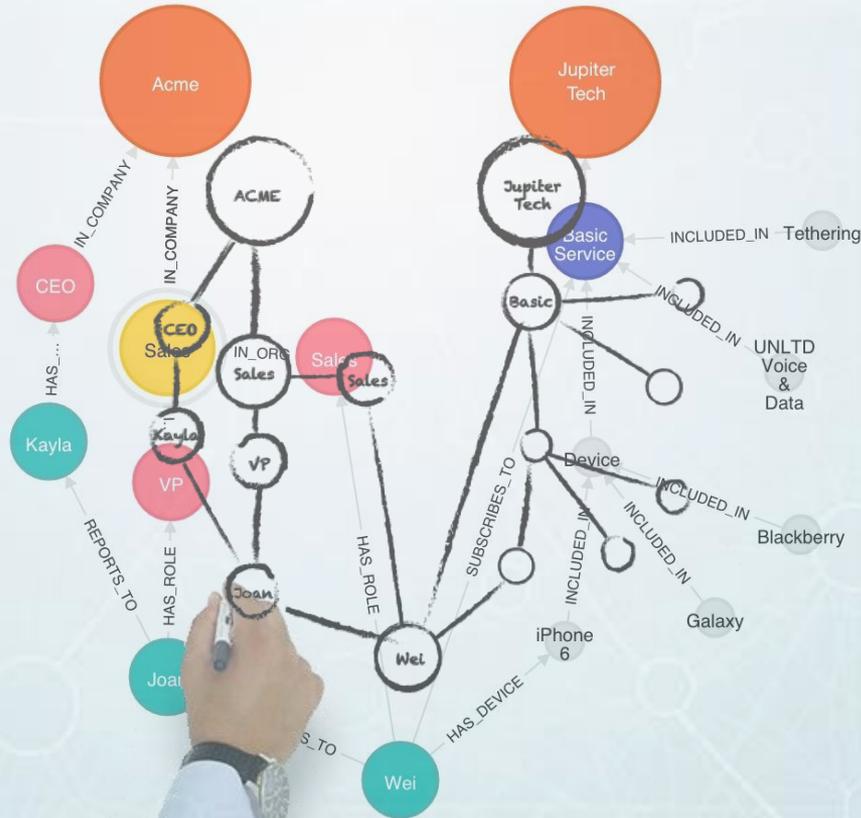
```
AND loc.location='New York'
```

```
AND type.cuisine='Sushi'
```

```
RETURN restaurant.name
```

Neo4j Property Graph

The Whiteboard Model is the Physical Model



A unified view for ultimate agility

- Easily understood
- Easily evolved
- Easy collaboration between business and IT

Neo4j Browser快速入门

Neo4j Browser



A GUI for power users and developers to access the graph database, and perform administrative tasks using Cypher.

- Running in web browser
- Installed as part of Neo4j database
- Service launched together with database
- To launch, access: `http://<hostname>:7474`

Neo4j Browser 101



- Change node colors
- Change which node property is displayed
- Run queries with CMD (CTRL) + Enter
- Insert new line with SHIFT + Enter
- Expand the query bar with ESC
- Double-click a node and see what happens!
- Use `:clear` to clear past results
- CMD (CTRL) + Up Arrow to scroll through past queries



Neo4j Browser – Browse Database Schema



The screenshot shows the Neo4j Browser interface in a web browser window. The browser's address bar shows "localhost:7474/browser/". The interface is split into two main sections. On the left is a dark sidebar titled "Database Information" which contains a "Node Labels" section with various labels like "ArtWork", "Artist", "Associate", etc., and a "Relationship Types" section with labels like "ABOUT", "APPLIES_TO", etc. On the right is the main content area, which has a search bar at the top containing a "\$" symbol. Below the search bar is a code editor with the text "\$:play start". The main content area features three informational cards: "Learn about Neo4j" with a "Start Learning" button, "Jump into code" with a "Write Code" button, and "Monitor the system" with a "Monitor" button. At the bottom of the main content area, there is a copyright notice: "Copyright © Neo4j, Inc 2002-2017".

Neo4j Browser – Analyse Data



neo4j@bolt://localhost:7...
localhost:7474/browser/

```
$ MATCH (a:Artist{name:'Jackson POLLOCK'}) RETURN a;
```

Graph (39) Artist(1) Associate(1) Entity(3) Person(2) RightOwner(1) ArtWork(26) Right(1) Material(2) Technique(1) Organisation(1)

*(58) CREATED(25) INFLUENCED_BY(1) DEPICTS(2) OWNS(1) HAS_PARENT(12) APPLIES_TO(14) USES(2) MADE_BY(1)

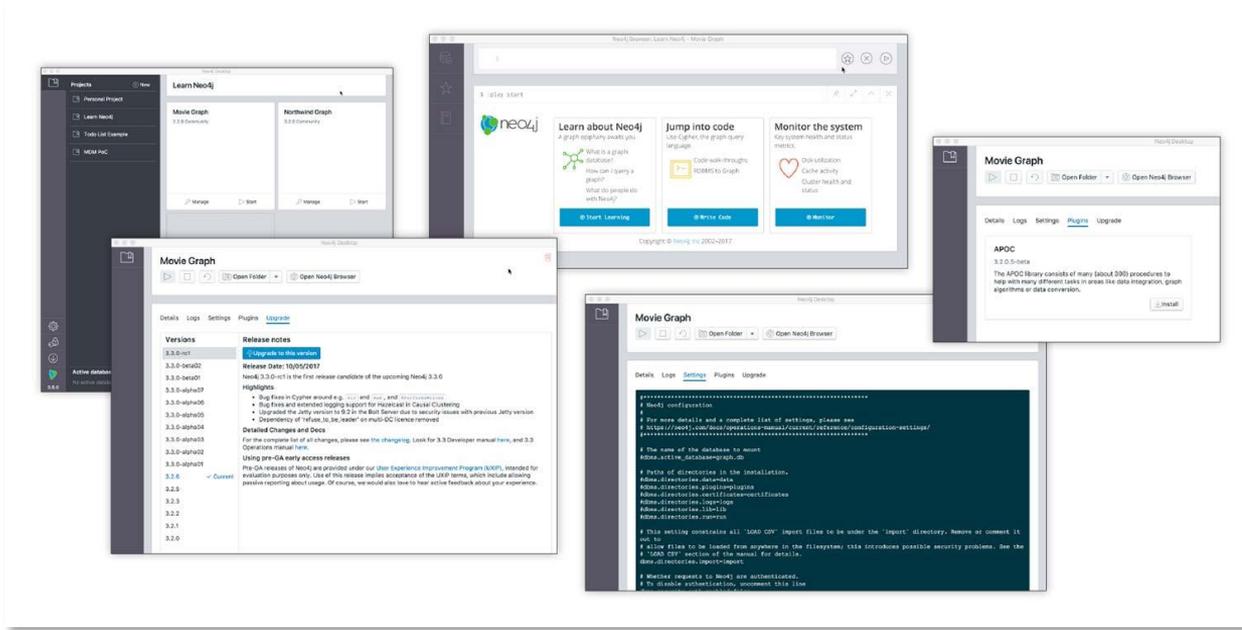
The graph visualization shows a central node for Jackson Pollock (purple circle) with numerous outgoing relationships. Key relationships include:

- Artists:** Created by Sainford MCCOY and BETTY FARGO.
- Materials:** Uses ink and paper.
- Artworks:** Created 25 artworks, including IDs 120467, 120466, 120462, 120458, 120455, 120454, 120450, 120449, 120452, 120451, 120446, 120445, 120444, 120443, 120442, 120441, 120440, 120439, 120438, 120437, 120436, 120435, 120434, 120433, 120432, 120431, 120430, 120429, 120428, 120427, 120426, 120425, 120424, 120423, 120422, 120421, 120420, 120419, 120418, 120417, 120416, 120415, 120414, 120413, 120412, 120411, 120410, 120409, 120408, 120407, 120406, 120405, 120404, 120403, 120402, 120401, 120400, 120399, 120398, 120397, 120396, 120395, 120394, 120393, 120392, 120391, 120390, 120389, 120388, 120387, 120386, 120385, 120384, 120383, 120382, 120381, 120380, 120379, 120378, 120377, 120376, 120375, 120374, 120373, 120372, 120371, 120370, 120369, 120368, 120367, 120366, 120365, 120364, 120363, 120362, 120361, 120360, 120359, 120358, 120357, 120356, 120355, 120354, 120353, 120352, 120351, 120350, 120349, 120348, 120347, 120346, 120345, 120344, 120343, 120342, 120341, 120340, 120339, 120338, 120337, 120336, 120335, 120334, 120333, 120332, 120331, 120330, 120329, 120328, 120327, 120326, 120325, 120324, 120323, 120322, 120321, 120320, 120319, 120318, 120317, 120316, 120315, 120314, 120313, 120312, 120311, 120310, 120309, 120308, 120307, 120306, 120305, 120304, 120303, 120302, 120301, 120300, 120299, 120298, 120297, 120296, 120295, 120294, 120293, 120292, 120291, 120290, 120289, 120288, 120287, 120286, 120285, 120284, 120283, 120282, 120281, 120280, 120279, 120278, 120277, 120276, 120275, 120274, 120273, 120272, 120271, 120270, 120269, 120268, 120267, 120266, 120265, 120264, 120263, 120262, 120261, 120260, 120259, 120258, 120257, 120256, 120255, 120254, 120253, 120252, 120251, 120250, 120249, 120248, 120247, 120246, 120245, 120244, 120243, 120242, 120241, 120240, 120239, 120238, 120237, 120236, 120235, 120234, 120233, 120232, 120231, 120230, 120229, 120228, 120227, 120226, 120225, 120224, 120223, 120222, 120221, 120220, 120219, 120218, 120217, 120216, 120215, 120214, 120213, 120212, 120211, 120210, 120209, 120208, 120207, 120206, 120205, 120204, 120203, 120202, 120201, 120200, 120199, 120198, 120197, 120196, 120195, 120194, 120193, 120192, 120191, 120190, 120189, 120188, 120187, 120186, 120185, 120184, 120183, 120182, 120181, 120180, 120179, 120178, 120177, 120176, 120175, 120174, 120173, 120172, 120171, 120170, 120169, 120168, 120167, 120166, 120165, 120164, 120163, 120162, 120161, 120160, 120159, 120158, 120157, 120156, 120155, 120154, 120153, 120152, 120151, 120150, 120149, 120148, 120147, 120146, 120145, 120144, 120143, 120142, 120141, 120140, 120139, 120138, 120137, 120136, 120135, 120134, 120133, 120132, 120131, 120130, 120129, 120128, 120127, 120126, 120125, 120124, 120123, 120122, 120121, 120120, 120119, 120118, 120117, 120116, 120115, 120114, 120113, 120112, 120111, 120110, 120109, 120108, 120107, 120106, 120105, 120104, 120103, 120102, 120101, 120100, 120099, 120098, 120097, 120096, 120095, 120094, 120093, 120092, 120091, 120090, 120089, 120088, 120087, 120086, 120085, 120084, 120083, 120082, 120081, 120080, 120079, 120078, 120077, 120076, 120075, 120074, 120073, 120072, 120071, 120070, 120069, 120068, 120067, 120066, 120065, 120064, 120063, 120062, 120061, 120060, 120059, 120058, 120057, 120056, 120055, 120054, 120053, 120052, 120051, 120050, 120049, 120048, 120047, 120046, 120045, 120044, 120043, 120042, 120041, 120040, 120039, 120038, 120037, 120036, 120035, 120034, 120033, 120032, 120031, 120030, 120029, 120028, 120027, 120026, 120025, 120024, 120023, 120022, 120021, 120020, 120019, 120018, 120017, 120016, 120015, 120014, 120013, 120012, 120011, 120010, 120009, 120008, 120007, 120006, 120005, 120004, 120003, 120002, 120001, 120000.
- Techniques:** Uses screen printing.
- Other:** Has parent 120467, made by 120466, and influenced by 120467.

APPLIES_TO <id>: 980469

Neo4j Desktop快速入门

Neo4j Desktop (Mac & Windows)



- The new mission control centre for developers
- Connecting to all your Neo4j databases within a single application
- Installing extensions like graph algorithms or ETL components
- Free to download, bundled with an Enterprise version of Neo4j for Developers

数据库扩展

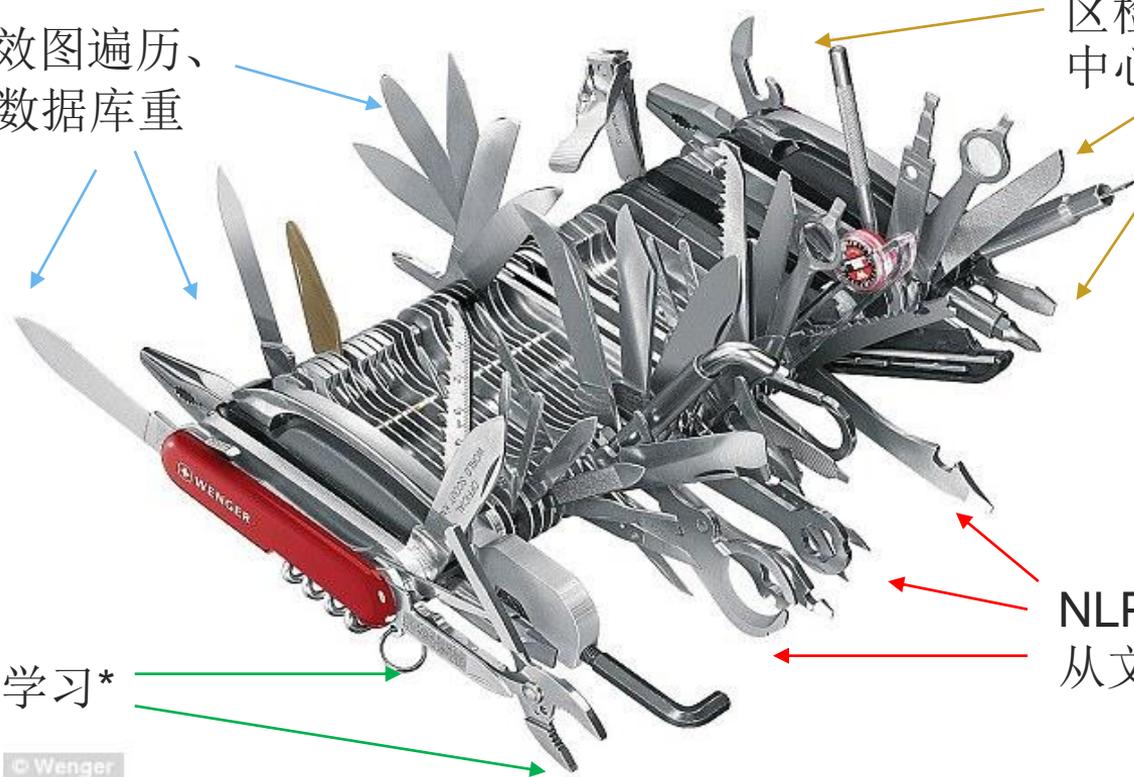
Neo4j数据库扩展(Database Extensions)

APOC – 高效图遍历、
数据集成、数据库重
构。。。

ALGO – 图算法：社
区检测、页面排行、
中心性、相似度。。。

ML – 机器学习*

NLP – 自然语言处理、
从文本到知识图谱



© Wenger

* Future development

Neo4j数据库扩展的下载和安装

对于社区版和企业版：

- 下载APOC(注意APOC和Neo4j的版本号的对应)

<https://github.com/neo4j-contrib/neo4j-apoc-procedures/releases/>

- 安装APOC

将下载的JAR文件复制到Neo4j根目录下的plugins文件夹中。

- 配置APOC权限

在neo4j.conf文件中加入下面的配置行：

```
dbms.security.procedures.unrestricted=apoc.*
apoc.import.file.enabled=true
apoc.export.file.enabled=true
```

- 重新启动Neo4j数据库服务
- 使用APOC，参见在线文档：

<https://neo4j-contrib.github.io/neo4j-apoc-procedures/>

在线教程

在Neo4j Browser中访问在线教程



- 入门教程 - 电影数据库
:play movies
- 更多在线教程
 - :play cypher (Cypher教程)
 - :play <https://guides.neo4j.com/import> (数据导入方法)
 - :play northwind-graph (从关系数据库到图数据库的转换)
 - :play query-template (常用查询)
 - :play http://guides.neo4j.com/modeling_airports/ (高级建模实例)
 - :play data_science (图数据分析和算法)

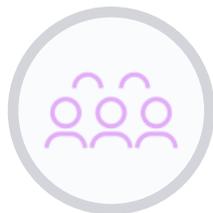
最大和最活跃的图数据库在线社区



Downloads

10M+

3M+ from Neo4j Distribution
7M+ from Docker



Events

400+

Approximate Number of
Neo4j Events per Year



Trained Developers

50k+

Trained/certified Neo4j
professionals



Meetups

50k+

Number of Meetup
Members Globally

- Neo4j中文社区: neo4j.com.cn
- 中国图数据库技术QQ和微信群
- ITPUB微课堂
- 中国用户和技术聚会

谢谢!

现在发送邮件提供反馈即可获得《写给关系型数据库开发人员的图数据库权威指南(白皮书)》电子版。