

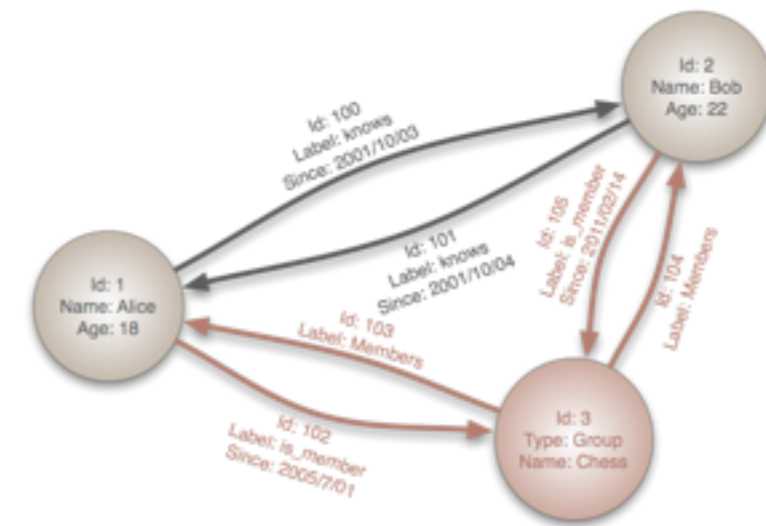
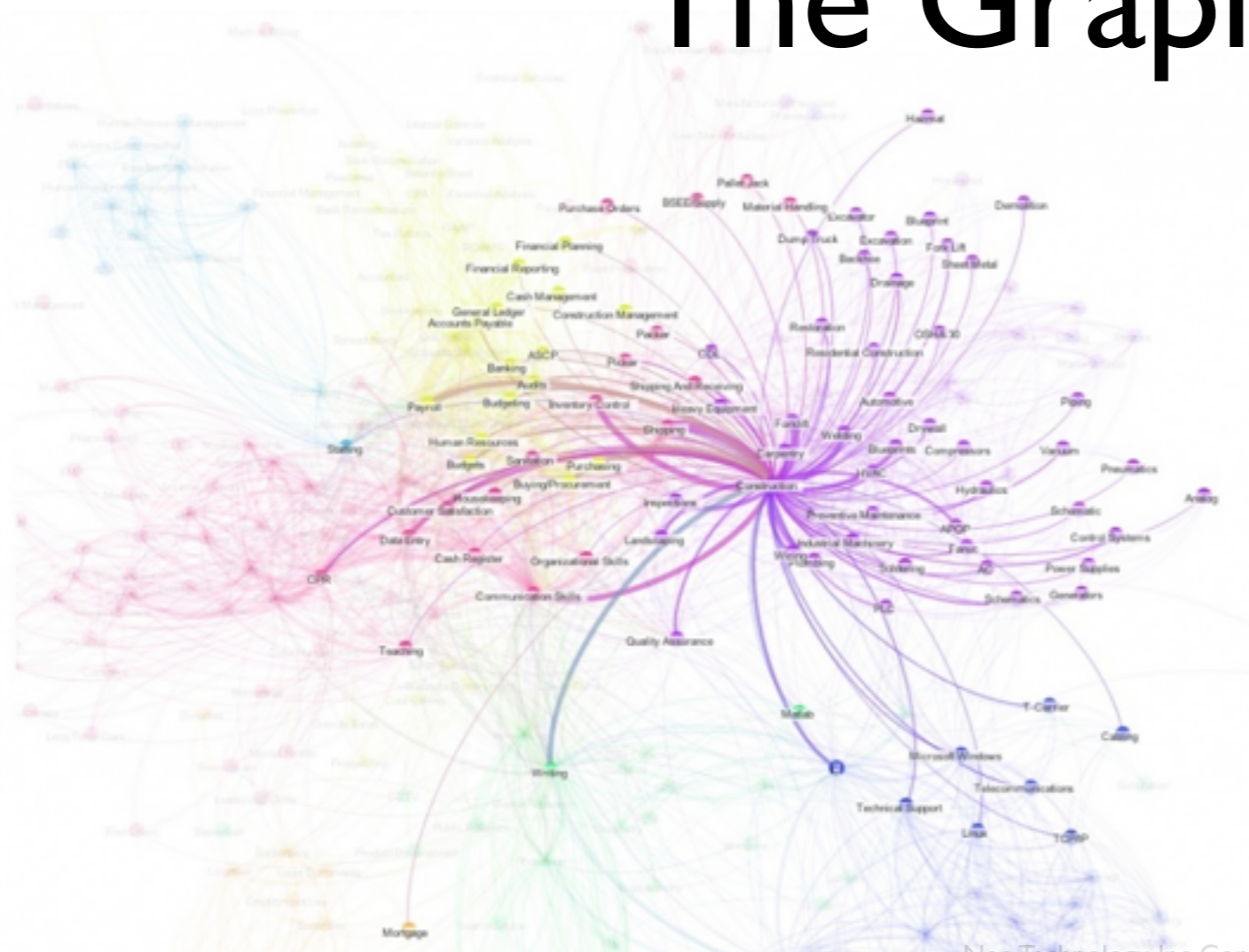
# Why you must use a graph database or fail

*As presented at NOSQL Roadshow SF on 2013-06-06*

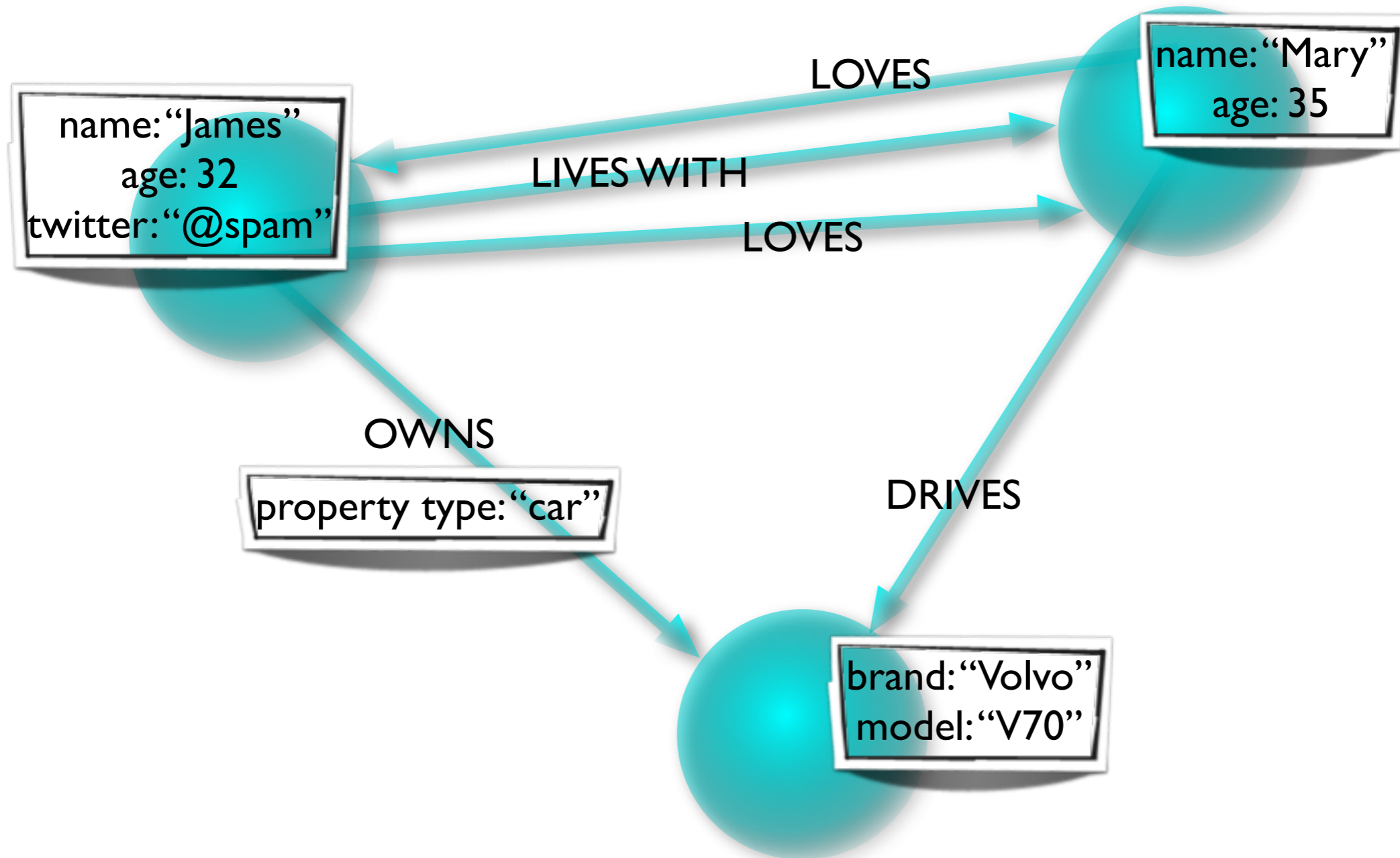
David Montag  
<[david@neotechnology.com](mailto:david@neotechnology.com)>  
@dmontag



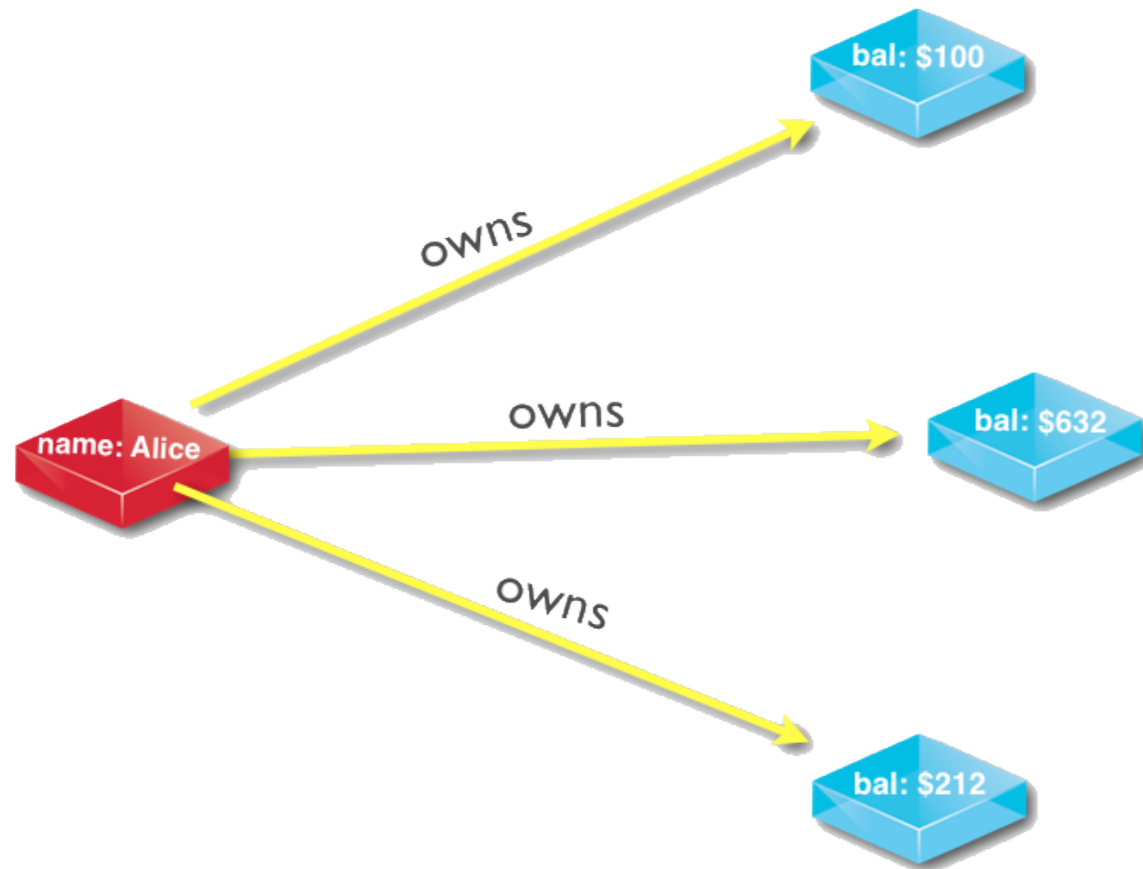
# The Graph Trend



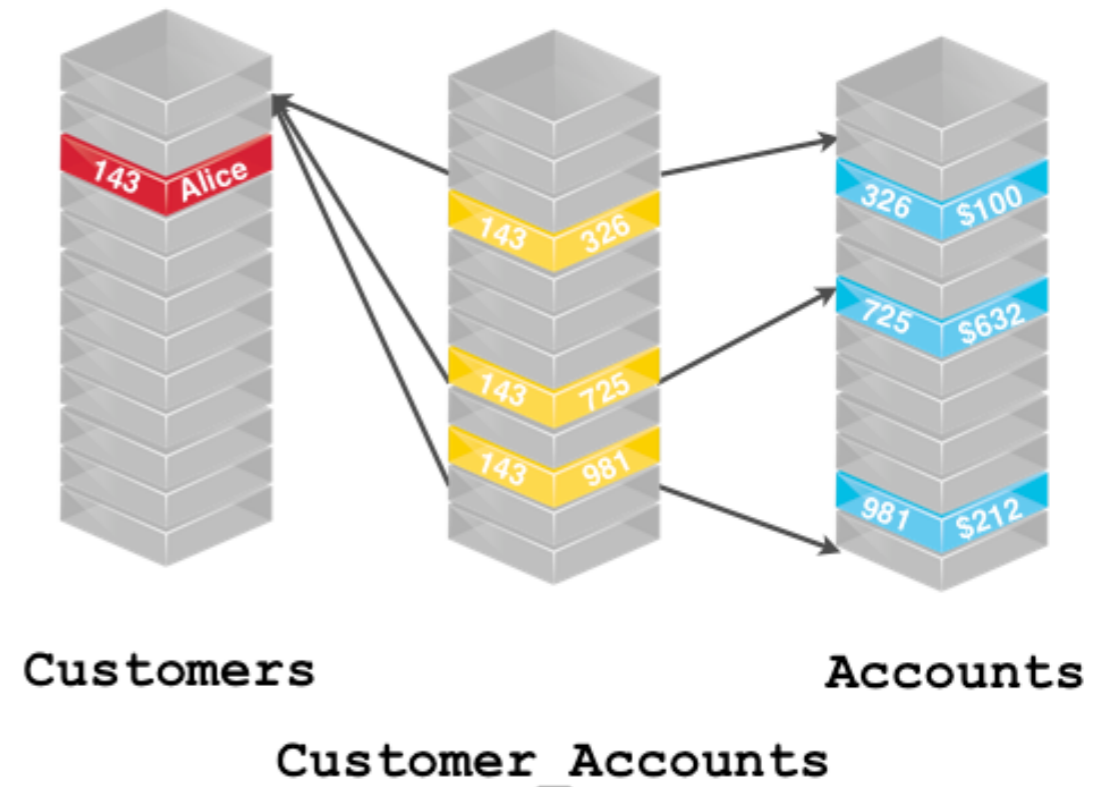
# What is a Graph?



# Contrasted with Relational



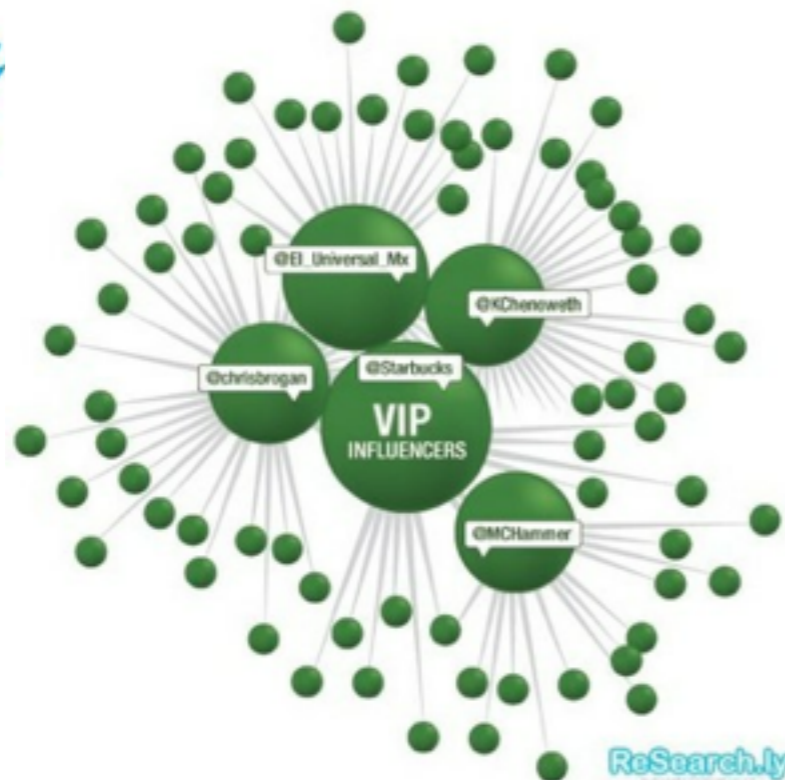
VS.



- Graphs are often referred to as “Whiteboard Friendly”. The data model reflects the way a domain expert would naturally draw their data on a whiteboard
- “The schema is the data”. Schema flexibility allows the system to change in response to a changing environment

# Early Adopters of Graph Tech

facebook.



Google™



The Knowledge Graph

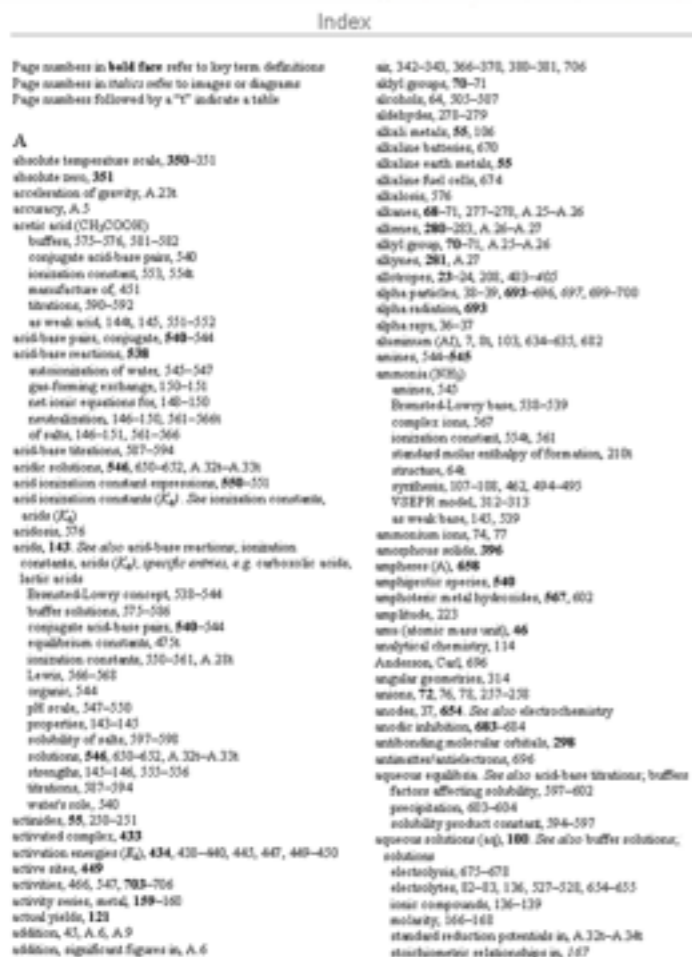
Learn more about one of the key breakthroughs behind the future of search.

# Survival of the Fittest

## Evolution of Web Search

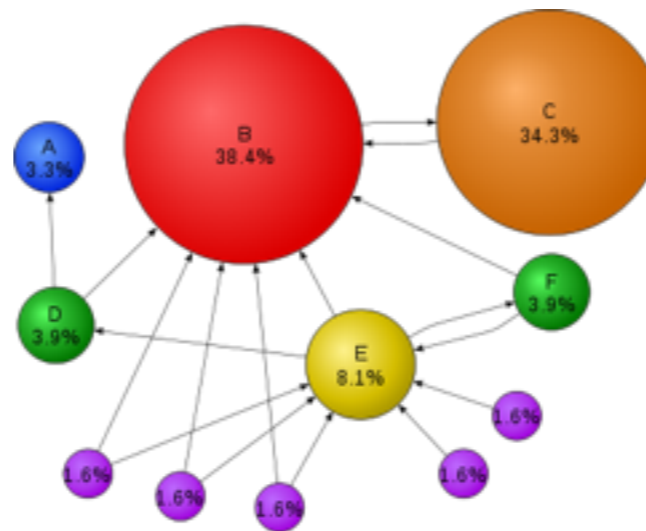
### Pre-1999

WWW Indexing



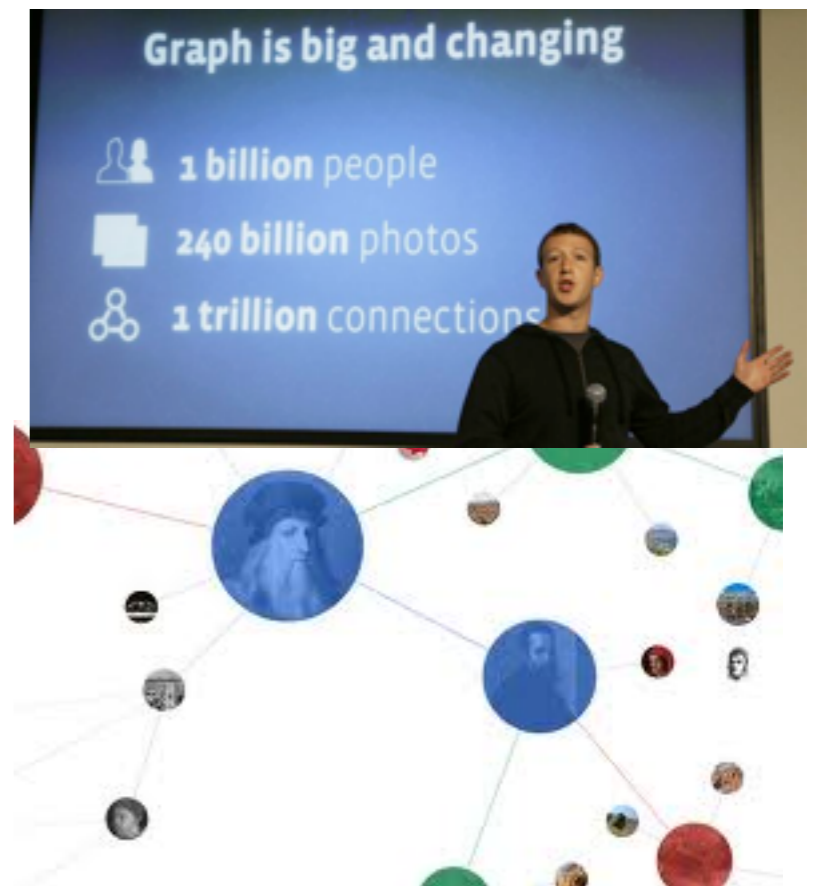
### 1999 - 2012

Google Invents PageRank



### 2012-?

Google Knowledge Graph, Facebook Graph Search



Discrete Data +

Connected Data (Simple)

+

Connected Data (Rich)

# Survival of the Fittest

## Evolution of Online Job Search



### 2010-11

Resume Searching & Scoring



95 score Experient UI/UX De  
50 score UI Designer Intern

#### Get Your Bright Scores

The Bright Score instantly scores jobs relative to your experience. The higher the score, the better the fit.

JobFocus™ Score *i*  
ReCalculate 6.2

Discrete Data



### 2011-12

Social Job Search



**Most jobs are found through an inside connection**

Each friend that joins Glassdoor allows you to see more connections at more companies

9 friends on glassdoor | 3,905 inside connections | 3,882 companies

Invite more friends — ask them to share their connections

#### Jobs with Connections

**Sr. Statistical Analyst, Product Innovation**

Netflix — Los Gatos, CA

From: Job.com — 1 days ago

**Creative Director**

frog design — San Francisco, CA

From: Experteer — 8 days ago

**Java Server Software Engineer**

Electronic Arts — Redwood City, CA

From: Experteer — 3 days ago

**EMERGENCY MEDICAL TECH**

U. S. NAVY — Fremont, CA

From: Monster — 14 days ago

Want better jobs? Tell us your current job title

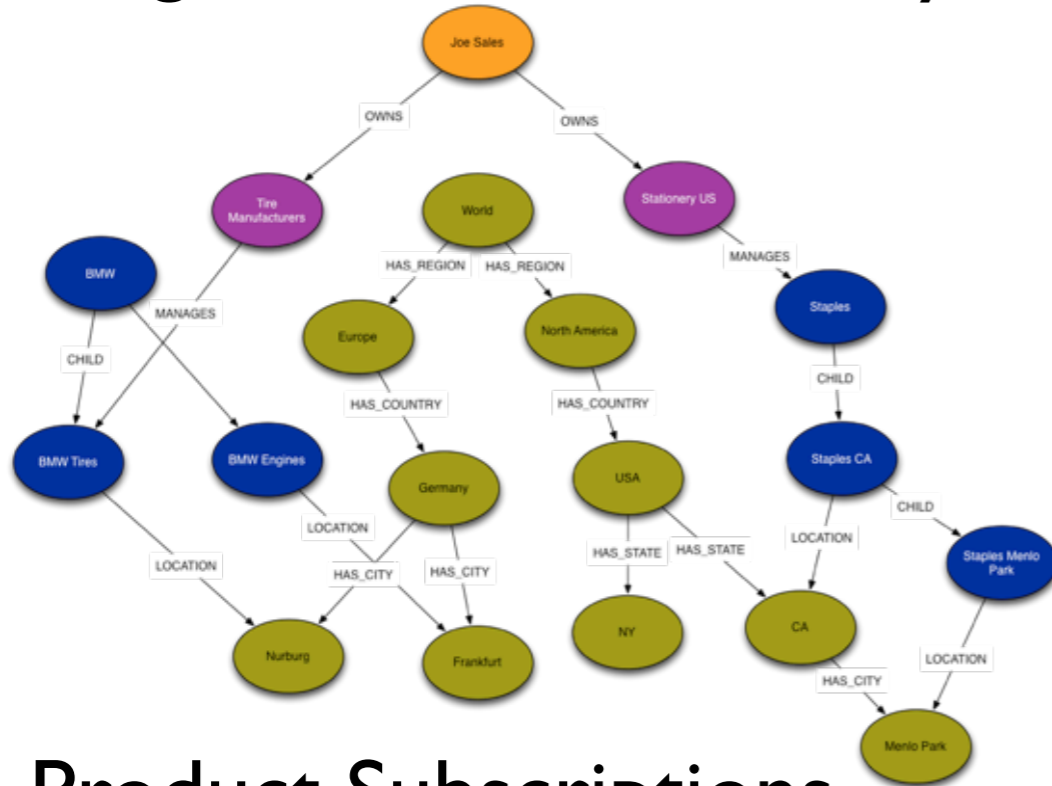
Connected Data

# A Few Uses of Graphs in Industry

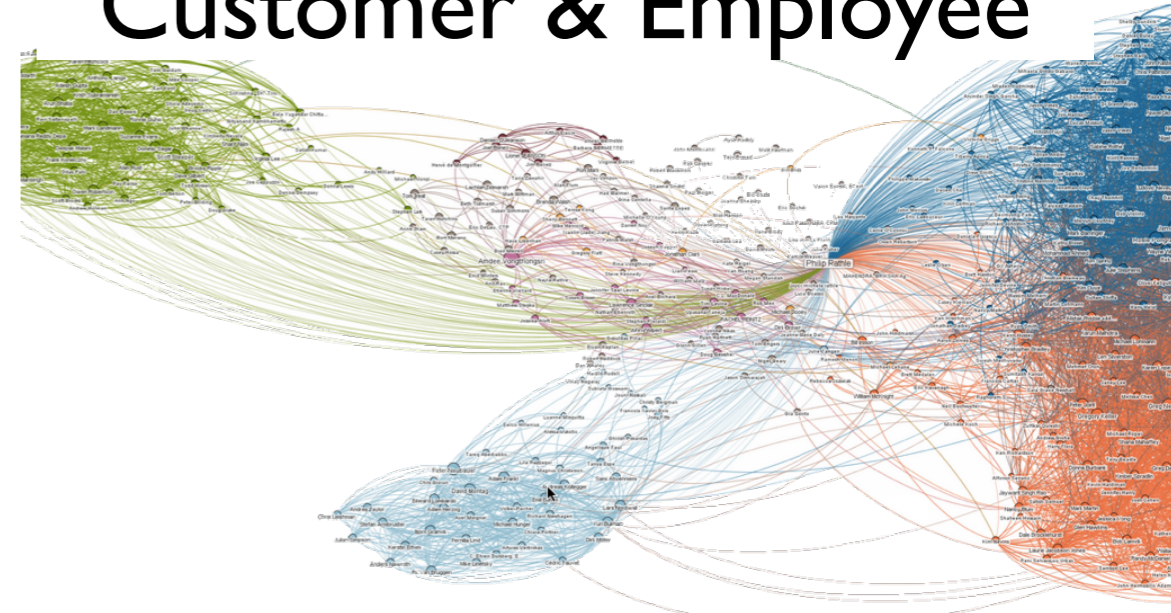


(Actual Neo4j Graphs)

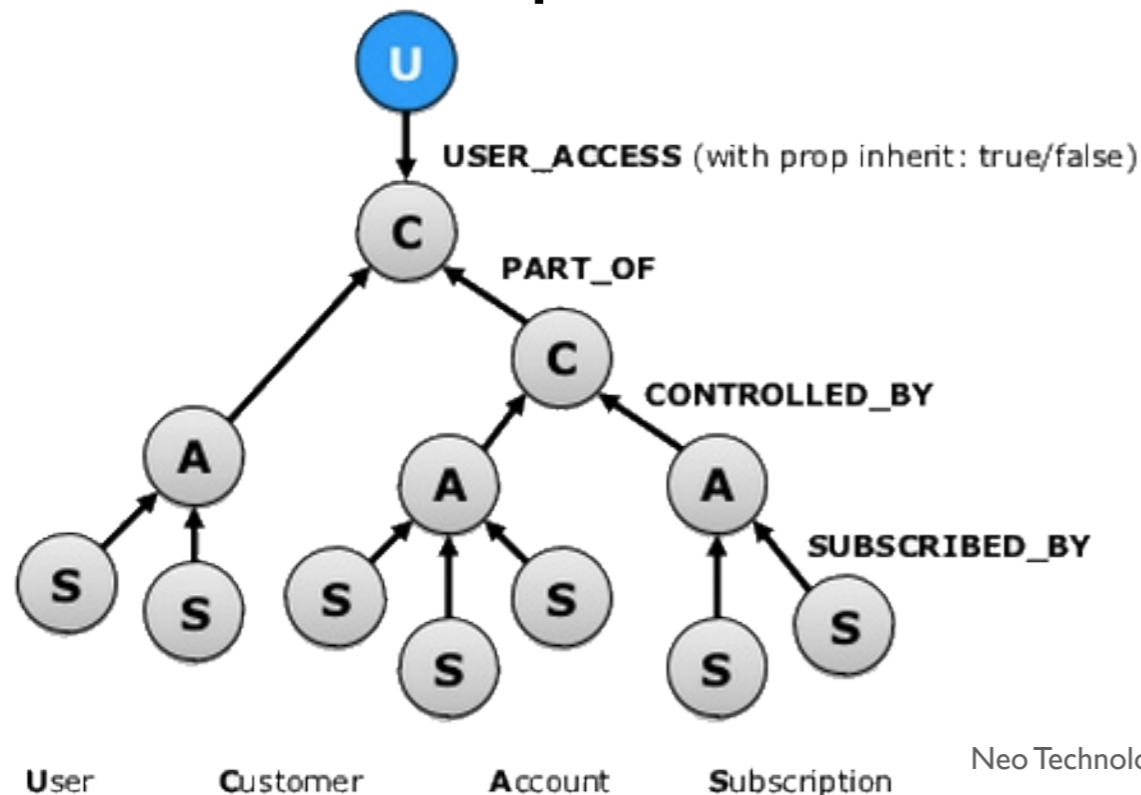
## Organizational Hierarchy



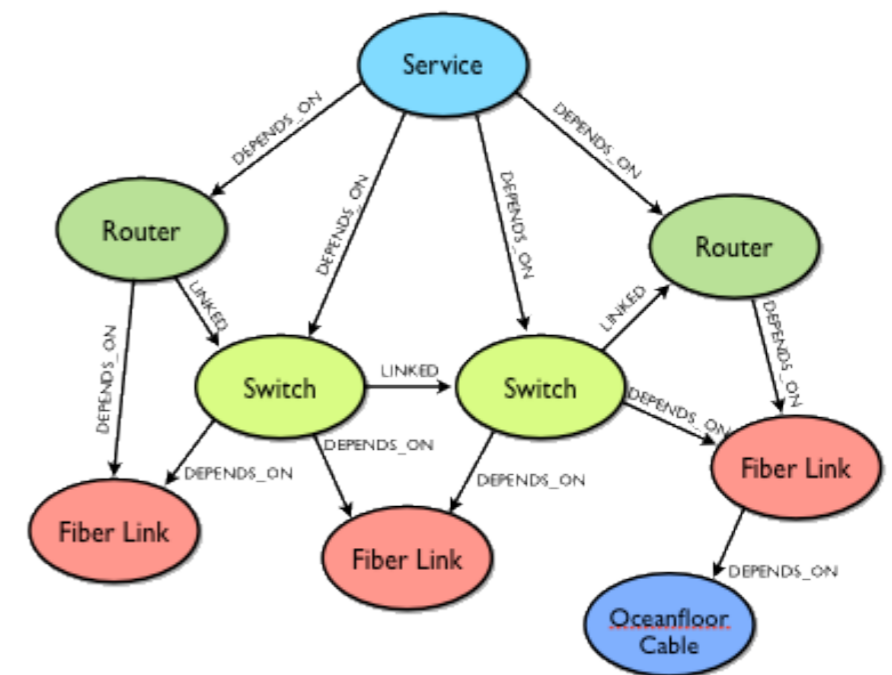
## Social Networks Customer & Employee



## Product Subscriptions



## CMDB (Network Inventory)





# A Few Uses of Graphs in Industry



(Actual Neo4j Graphs)

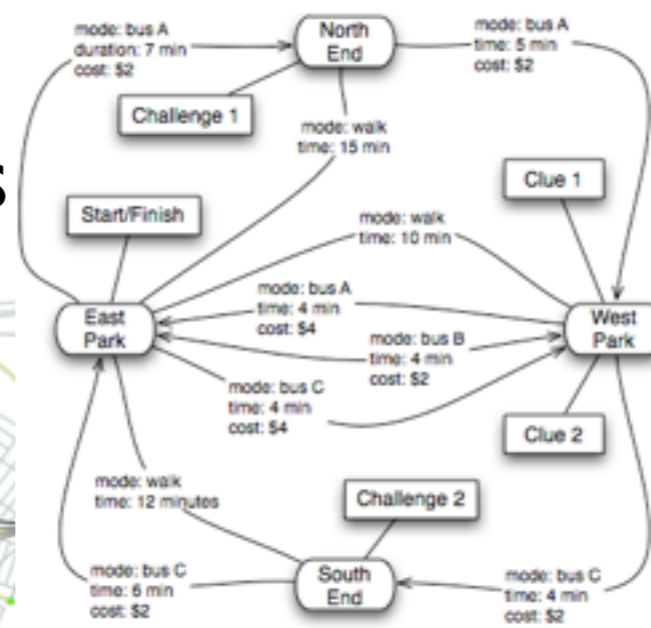
## Entitlements & Identity Management



## Insurance Risk Analysis



## Geo Routing (Public Transport)



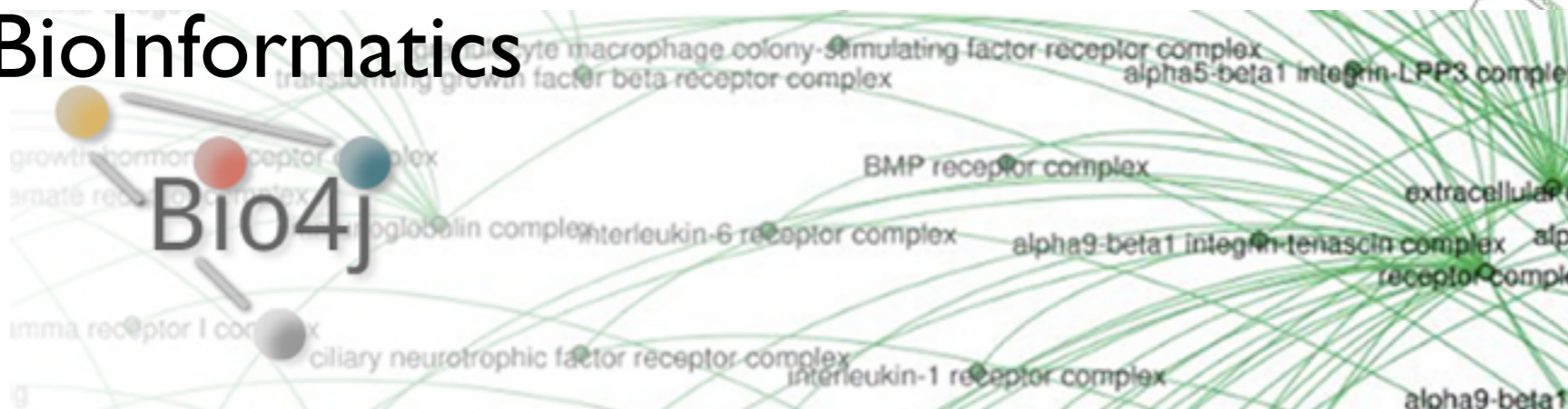
## Network Asset Management



## Network Cell Analysis



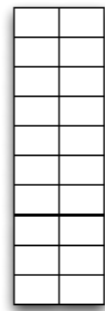
## BioInformatics



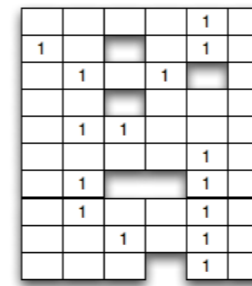
# About Neo Technology

# Four Categories of NoSQL

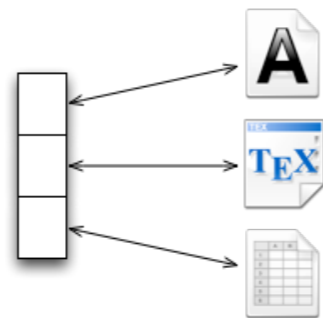
Key-Value



Column-family /  
BigTable



Document



Graph



# Our Mission:

Help the world to  
make sense of data



## neo technology **The Company**

- Commercial sponsor of Neo4j
- 50 people / Ten countries / Four continents
- \$25M in venture funding from Fidelity, Sunstone & Conor
- 100+ Customers, incl. 30+ of the Global 2000.

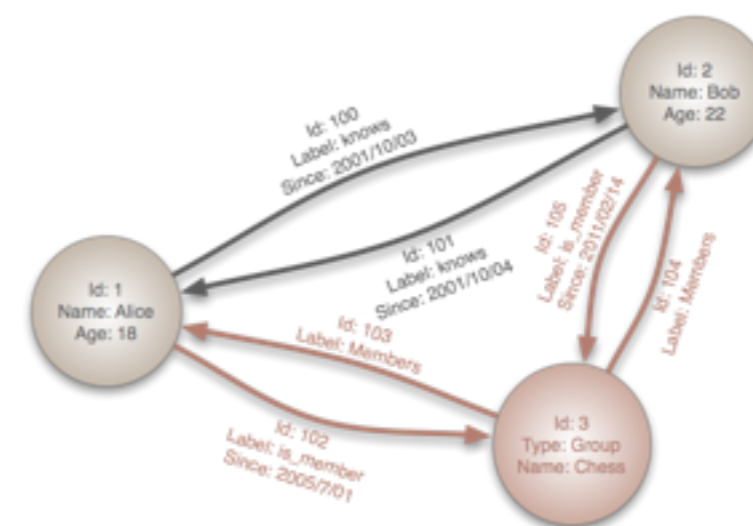


## Neo4j **The Product**

- Development started in 2000 in Sweden
- 10+ years of mission-critical 24x7 deployments (since 2003)
- Open sourced in 2007
- **40,000+** downloads per month. Over half a million downloads.
- Today the leading graph database



# Use Case & Industry Matrix



# Early Adopter Segments





















**(What we expected to happen - view from several years ago)**

<b>Core Industries &amp; Use Cases:</b>	<b>Web / ISV</b>	<b>Finance &amp; Insurance</b>	<b>Datacom / Telecom</b>
<b>Network &amp; Data Center Management</b>			
<b>MDM</b>			
<b>Social</b>			
<b>Geo</b>			

# Neo4j Adoption Snapshot



Select Commercial Customers\* Across Anticipated Segments

Core Industries & Use Cases:	Web / ISV	Finance & Insurance	Telecommunications
Network & Data Center Management	    		 
MDM			
Social	     		
Geo			 <p>India's No.1 local search engine</p>

\*Community Users Not Included

Neo Technology, Inc Confidential



# Neo4j Adoption Snapshot

neo technology  
graphs are everywhere

## Select Commercial Customers (Community Users Not Included)

Core Industries & Use Cases:	Software	Financial Services	Telecommunications	Web Social, HR & Recruiting	Health Care & Life Sciences	Media & Publishing	Energy, Services, Automotive, Gov't, Logistics, Education, Gaming, Other
Network & Data Center Management	Zenoss, Juniper, NetApp, SERENA, gen		hp, SFR				
MDM / System of Record			CISCO		Janssen, ZEPHYR HEALTH INC, HealthUnlocked		teachscape
Social	Glowbl	ICE	Deutsche Telekom, maaii	viadeo, glassdoor		LIFECHURCH.TV, SQUIDOO	LAUREATE INTERNATIONAL UNIVERSITIES
Geo	Dinglicom		Justdial	classmates.com		indiatimes	gamesys, Accenture
Identity & Access Mgmt	aikux.com, entropy		telenor				shuti
Content Management	springcm, Adobe			Dshini, SRM	SevenBridges	zeebox, LifeWay	DOSB NEW MEDIA GMBH, DEUTSCHER OLYMPISCHER SPORTBUND
Recommendations	AXON ACTIVE			hinge, careerbuilder, InfoJobs		<fuseworks/>, Perigee, CHIP	research now, compete
BI, CRM, Impact Analysis, Fraud Detection, Resource Optimization, etc.	kitedesk, SODIFRANCE, idMISSION, Humanvest.co	DRW TRADING GROUP, NexLP		moviepilot, t		DRAKER	LOCKHEED MARTIN, Juice PLUS+

# Graph Buzz!

## MacArthur 'Genius Grant' Winner Maria Chudnovsky on Graph Theory

Wednesday, October 03, 2012



## Bright Launches Bright Packed With Jobs Data Seeking Tips



InfoWorld Home / InfoWorld Tech Watch / Buzz grows around graph databases



The First Word on Tech  
INFOWORLD TECH WATCH

AUGUST 29, 2012

## Buzz grows around graph databases

Interest in graph databases will continue to grow, given its ability to analyze data delivered in a non-relational format, such as social networking data

By Paul Krill | InfoWorld

Follow @pjkrill

## Graph Databases: The New Way to Access Super Fast Social Data

September 26, 2012 by Emil Elfrem



259

Like

969

Tweet

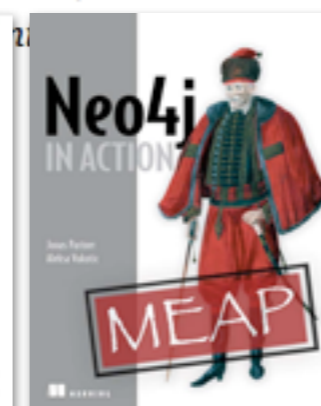
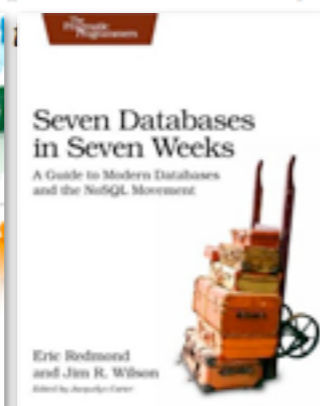
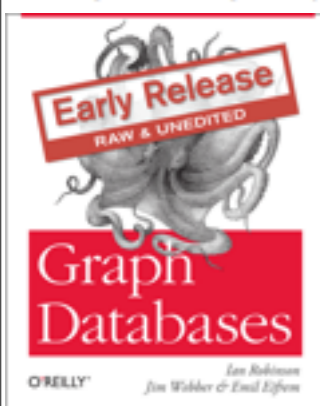
neo4j leaves me speechless. Good job at building the best graph database in the world!

2:41 AM Mar 24th via Tweetie  
Retweeted by you and 1 other

fokussiertnet  
Daniel

## Facebook's Social Graph, Neo4j show rising use of graph databases

**Summary:** Facebook's Social Graph -- the database underlying its Graph Search engine unveiled yesterday-- is just one of many graph databases being employed for complex, connected data. Neo4j



I saw my own Interest Graph and it's scary-accurate. We'd certainly pay for the ability to use the Gravity personalization technology I saw today at TechCrunch to help target content to users.

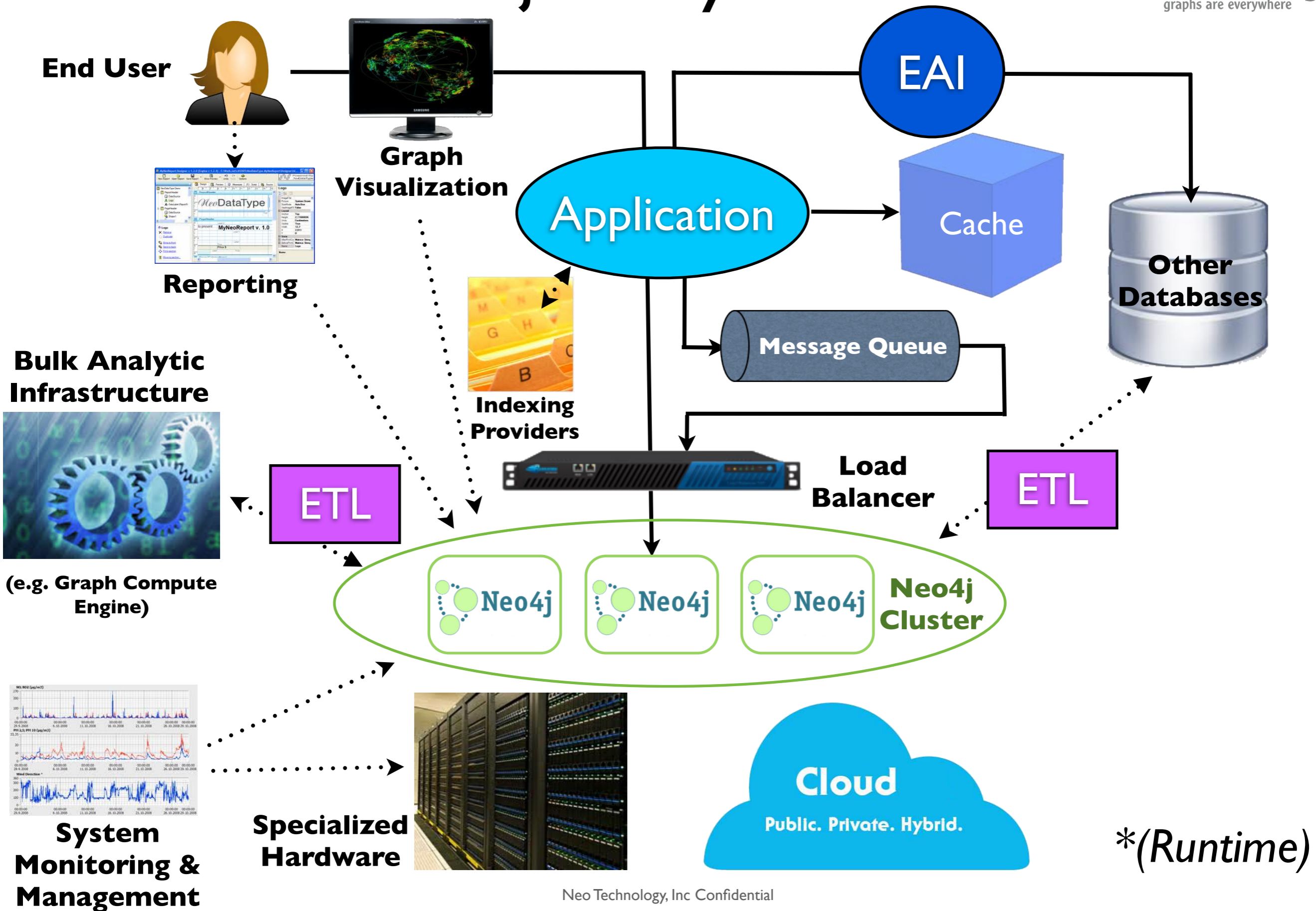
TC Michael Arrington, TechCrunch

# Top 3 Reasons for Choosing Neo4j



1. **Problems with JOIN performance**
2. **Domain Fit** for graph
3. **Open-ended business requirements**  
necessitating fast, iterative development

# Neo4j Ecosystem\*



# Selected Case Studies



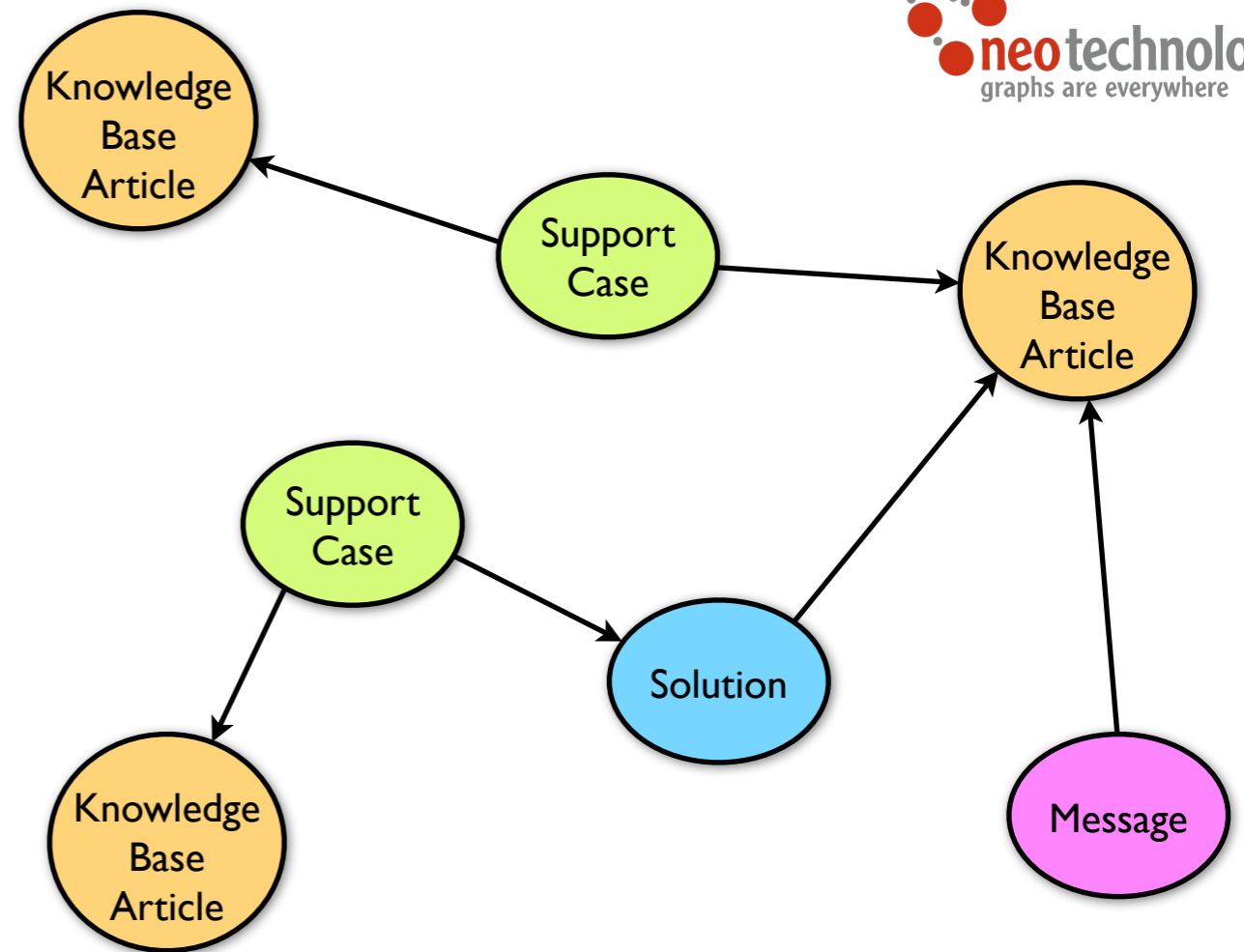
Industry: Communications  
Use case: Recommendations  
San Jose, CA

**Cisco.com**



## Background

- Cisco.com serves customer and business customers with Support Services
- Needed real-time recommendations, to encourage use of online knowledge base
- Cisco had been successfully using Neo4j for its internal master data management solution.
  - *Identified a strong fit for online recommendations*



## Business problem

- Call center volumes needed to be lowered by improving the efficacy of online self service
- Leverage large amounts of knowledge stored in service cases, solutions, articles, forums, etc.
- Problem resolution times, as well as support costs, needed to be lowered

## Solution & Benefits

- Cases, solutions, articles, etc. continuously scraped for cross-reference links, and represented in Neo4j
- Real-time reading recommendations via Neo4j
- Neo4j Enterprise with HA cluster
- The result: customers obtain help faster, with decreased reliance on customer support



## Background

- One of the world's largest logistics carriers
- Projected to outgrow capacity of old system
- New parcel routing system
  - *Single source of truth for entire network*
  - *B2C & B2B parcel tracking*
  - *Real-time routing: up to 5M parcels per day*



## Business problem

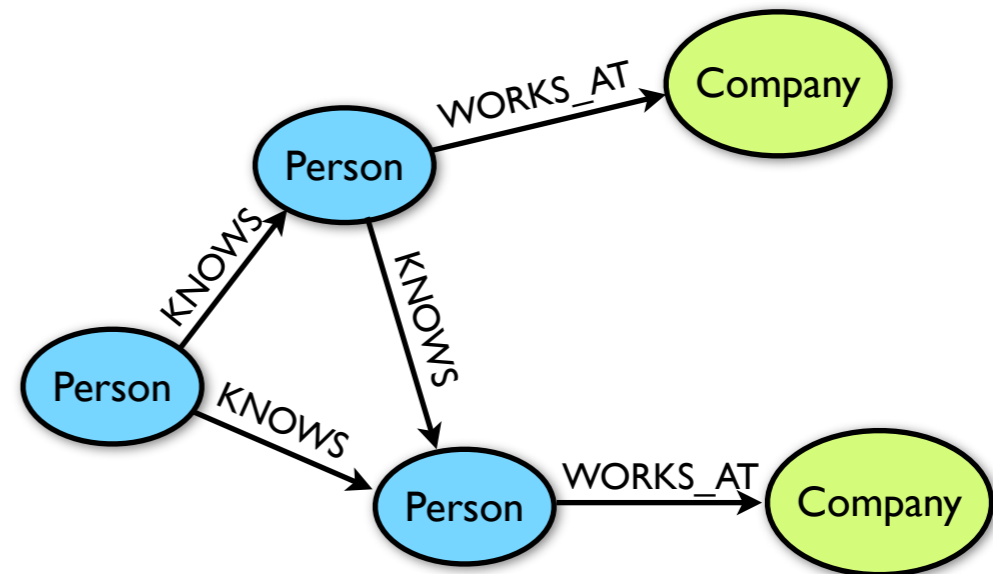
- 24x7 availability, year round
- Peak loads of 2500+ parcels per second
- Complex and diverse software stack
- Need predictable performance & linear scalability
- Daily changes to logistics network: route from any point, to any point

## Solution & Benefits

- Neo4j provides the ideal domain fit:
  - a logistics network is a graph
- Extreme availability & performance with Neo4j clustering
- Hugely simplified queries, vs. relational for complex routing
- Flexible data model can reflect real-world data variance much better than relational
- “Whiteboard friendly” model easy to understand

## Background

- Online jobs and career community, providing anonymized inside information to job seekers



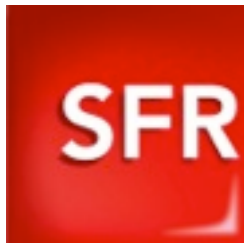
## Business problem

- Wanted to leverage known fact that most jobs are found through personal & professional connections
- Needed to rely on an existing source of social network data. Facebook was the ideal choice.
- End users needed to get instant gratification
- Aiming to have the best job search service, in a very competitive market

## Solution & Benefits

- First-to-market with a product that let users find jobs through their network of Facebook friends
- Job recommendations served real-time from Neo4j
- Individual Facebook graphs imported real-time into Neo4j
- Glassdoor now stores > 50% of the entire Facebook social graph
- Neo4j cluster has grown seamlessly, with new instances being brought online as graph size and load have increased





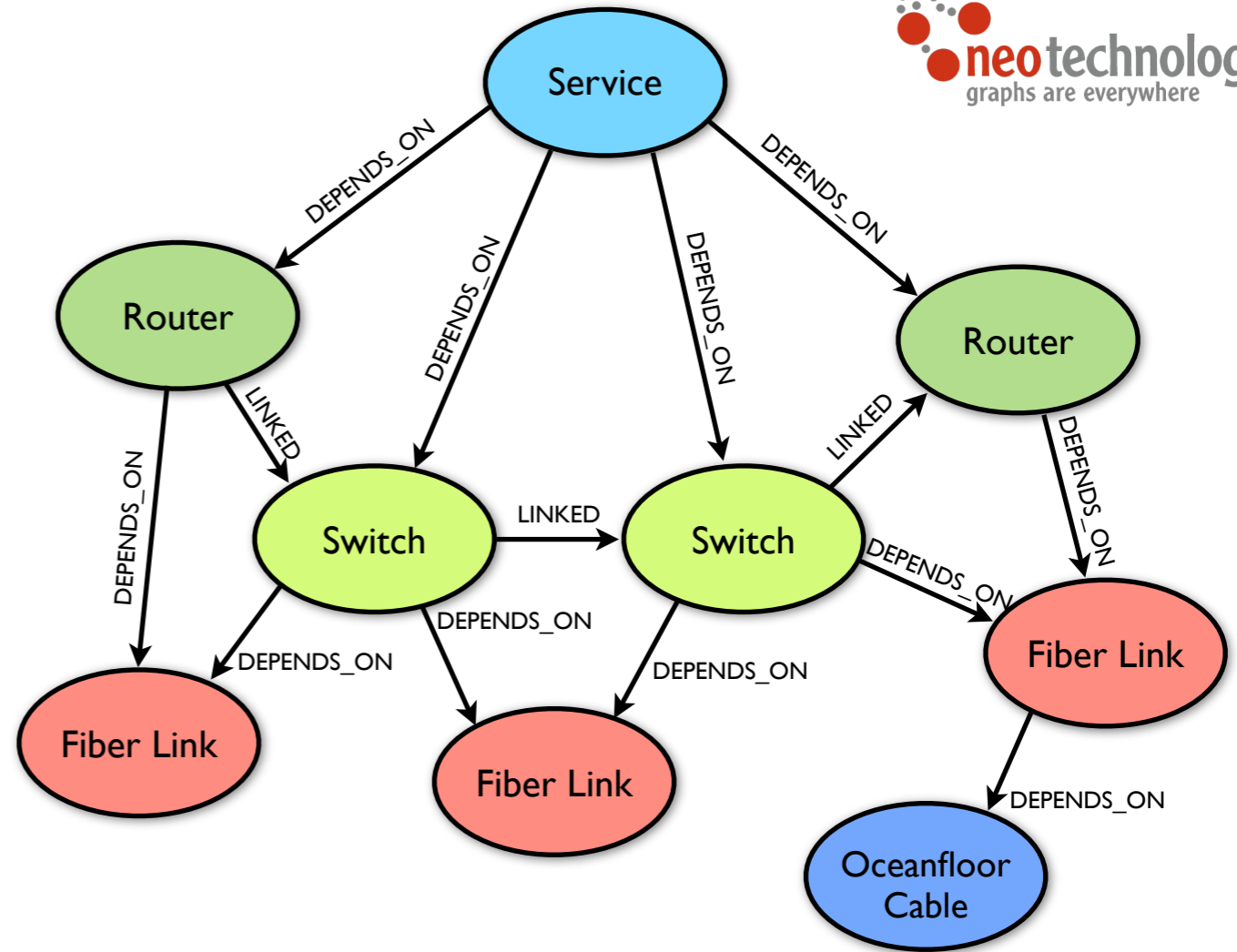
Industry: Communications  
Use case: Network Management

Paris, France



## Background

- Second largest communications company in France
- Part of Vivendi Group, partnering with Vodafone



## Business problem

- Infrastructure maintenance took one full week to plan, because of the need to model network impacts
- Needed rapid, automated “what if” analysis to ensure resilience during unplanned network outages
- Identify weaknesses in the network to uncover the need for additional redundancy
- Network information spread across > 30 systems, with daily changes to network infrastructure
- Business needs sometimes changed very rapidly

## Solution & Benefits

- Flexible network inventory management system, to support modeling, aggregation & troubleshooting
- Single source of truth (Neo4j) representing the entire network
- Dynamic system loads data from 30+ systems, and allows new applications to access network data
- Modeling efforts greatly reduced because of the near 1:1 mapping between the real world and the graph
- Flexible schema highly adaptable to changing business requirements

## Subtitle

### Background

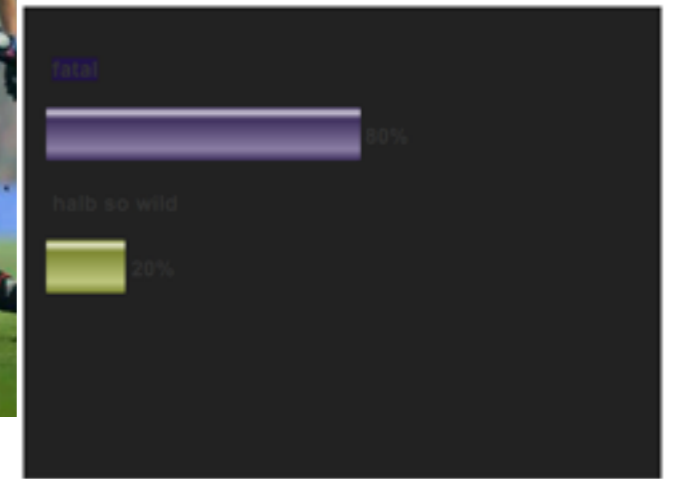
- Europe's largest communications company
- Provider of mobile & land telephone lines to consumers and businesses, as well as internet services, television, and other services

> 236,000  
Employees worldwide in 2011

50  
Countries

> 58 bn. €  
Revenue in 2011

### Interactive Television Programming



### Business problem

- The Fanorakel application allows fans to have an interactive experience while watching sports
- Fans can vote for referee decisions and interact with other fans watching the game
- Highly connected dataset with real-time updates
- Queries need to be served real-time on rapidly changing data
- One technical challenge is to handle the very high spikes of activity during popular games

### Solution & Benefits

- Interactive, social offering gives fans a way to experience the game more closely
- Increased customer stickiness for Deutsche Telekom
- A completely new channel for reaching customers with information, promotions, and ads
- Clear competitive advantage



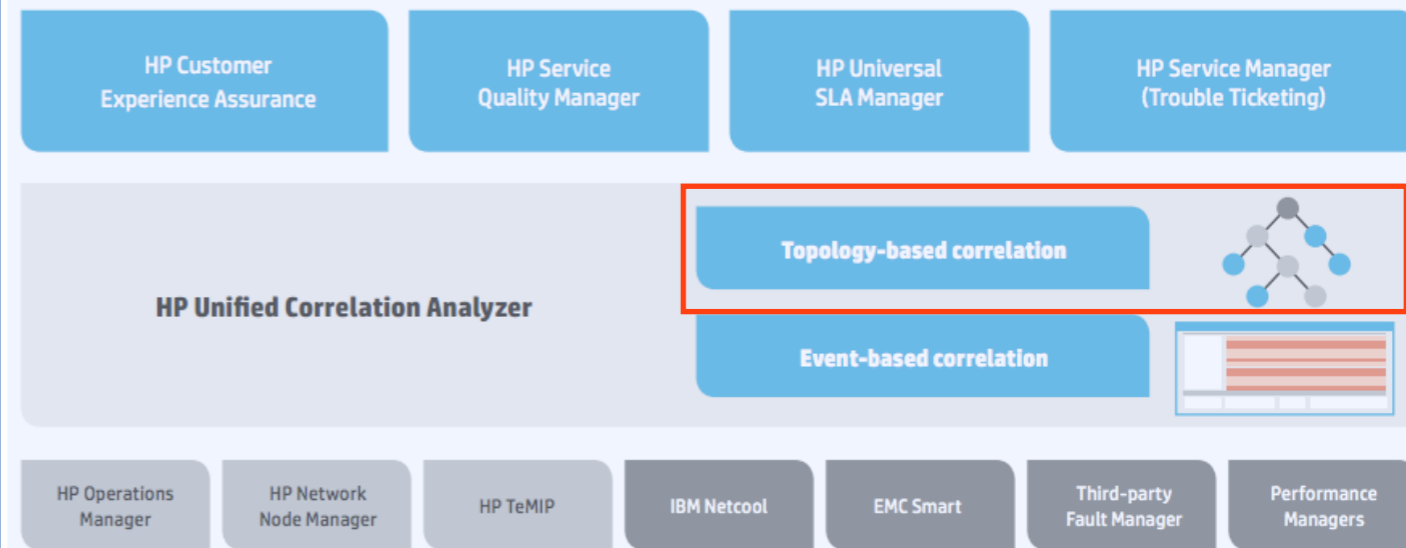
Industry: Web/ISV, Communications  
Use case: Network Management  
Global (U.S., France)



## Background

- World's largest provider of IT infrastructure, software & services
- HP's *Unified Correlation Analyzer* (UCA) application is a key application inside HP's OSS Assurance portfolio
- Carrier-class resource & service management, problem determination, root cause & service impact analysis
- Helps communications operators manage large, complex and fast changing networks

A cross-domain, umbrella correlation solution



## Business problem

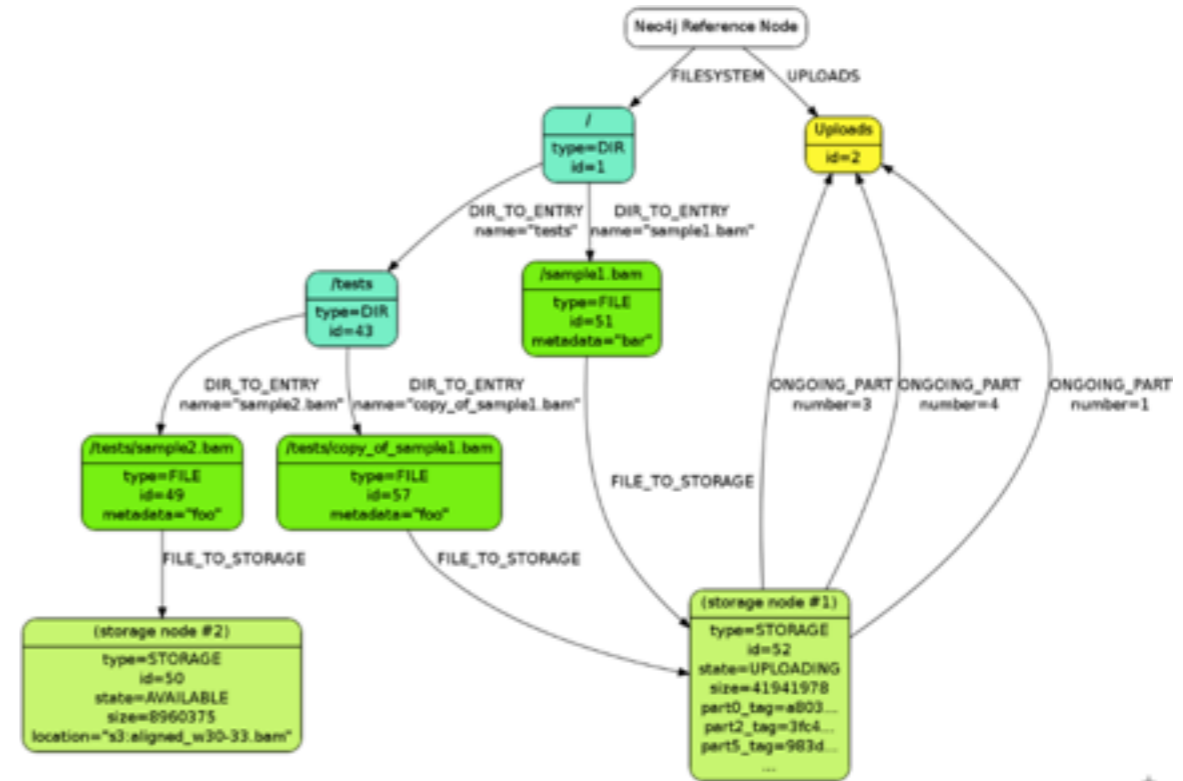
- Use network topology information to identify root problems causes on the network
- Simplify alarm handling by human operators
- Automate handling of certain types of alarms Help operators respond rapidly to network issues
- Filter/group/eliminate redundant Network Management System alarms by event correlation

## Solution & Benefits

- Accelerated product development time
- Extremely fast querying of network topology
- Graph representation a perfect domain fit
- 24x7 carrier-grade reliability with Neo4j HA clustering
- Met objective in under 6 months

## Background

- Bioinformatics company offering gene sequencing "as a service" (over the web)
- Provider of genomic information services
- Needed a new platform to support storage & retrieval of sequenced genomes in the cloud



## Business problem

- Neo4j is used to store metadata about each sequenced genome (including a pointer to the sequenced genome itself, which is a binary file stored on Amazon S3), and to support search and other forms of information processing against the genomic data.
- graph database was chosen because "Our specific domain maps naturally onto graph paradigm".

## Solution & Benefits

- **Domain fit**
  - Domain naturally lends itself to a graph representation.
  - Graph model determined to be a perfect fit.
- **Agility & Performance**
  - Saved time with Neo4j as compared to the alternatives.
  - Queries "practically write themselves."
- **Solution Completeness**
  - "Neo4j is incomparably better than other graph databases."