Who We Are: The Graph Platform

Neo4j is an enterprise-grade native graph platform that enables you to:

• **Store, reveal and query** data relationships
• **Traverse and analyze** any levels of depth in real-time
• **Add context and connect** new data on the fly

Designed, built and tested natively for graphs from the start for:

• Performance
• ACID Transactions
• Schema-free Agility
• Graph Algorithms

• Developer Productivity
• Hardware Efficiency
• Global Scale
• Graph Adoption
Neo4j - The Graph Company
Industry’s Largest Dedicated Investment in Graphs

- Creator of the Neo4j Graph Platform
- ~200 employees
- HQ in Silicon Valley, other offices include London, Munich, Paris, and Malmö Sweden
- $80M new funding led by Morgan Stanley & One Peak. Total $160M from Fidelity, Sunstone, Conor, Creandum, and Greenbridge Capital
- Over 15M+ downloads & container pulls
- 300+ enterprise subscription customers with over half with >$1B in revenue

Adoption
- 7/10 Top Retail Firms
- 12/25 Top Financial Firms
- 8/10 Top Software Vendors

Ecosystem
- 720+ Startups in program
- 300+ Enterprise customers
- 100+ Partners
- 53K+ Meet up members
- 450+ Events per year

Customers
- LinkedIn
- UBS
- Comcast
- Cisco
- Lufthansa
- Walmart
- eBay

Partners
- Google Cloud Platform
- Amazon Web Services
- IBM
- Microsoft Azure
- Tata
- Pitney Bowes
- GraphAware
- struktur
Native Graph Platform: Tools for Many Users

Drivers & APIs

Discovery & Visualization

Data Integration

Graph Transactions

Graph Analytics

Applications

Business Users

Developers

Admins

Data Analysts

Data Scientists

Enterprise Data Hub
Neo4j Invented the Labeled Property Graph Model

Nodes
• Can have *Labels* to classify nodes
• Labels have *native indexes*

Relationships
• Relate nodes by type and direction

Properties
• Attributes of Nodes & Relationships
• Stored as Name/Value pairs
• Can have indexes and composite indexes
• Visibility security by user/role

name: “Dan”
born: May 29, 1970
twitter: “@dan”

name: “Ann”
born: Dec 5, 1975

[Diagram showing relationships: PERSON (Dan) MARRIED TO PERSON (Ann) LIVES WITH PERSON (Ann) OWNS CAR (brand: “Volvo” model: “V70” Latitude: 37.562990° Longitude: -122.325530°)]
Neo4j Bloom

- High fidelity
- Scene navigation
- Property views
- Search suggestions
- Saved phrase history
- Property editor
- Schema perspectives
- Bloom chart type

- Visualize
- Communicate
- Discover
- Navigate
- Isolate
- Edit
- Share
Neo4j is the #1 Graph Platform

- Financial Services: 20 of top 25 use
- Airlines: 3 of top 5 use
- Hotels: 3 of top 5 use
- Airplane Makers: 3 of top 5 use
- Retailers: 7 of top 10 use
- Software Companies: 7 of top 10 use
- Global Telcos: 4 of top 5 use
Largest Pool of Graph Technologists

Downloads: 13M+
- 4M+ from Neo4j Distribution
- 9M+ 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
How Neo4j Fits — Common Architecture Patterns

From Tabular Data To Connected Data
From Disparate Silos To Cross-Silo Connections
From Data Lake Analytics to Real-Time Operations
Graph Transformation Maturity

1st Graph
Connects similar objects: Networks of people, computers, etc.

Desire for more context to follow connections.

Context Paths

Cross-Connect
Connect dissimilar objects: person to product or location. Mobile app explosion. Recommendation engines. Fraud detectors.

Graph Layers

Auto-Graphs
Cross-tech applications. Internet of Things operations. Transparent Neural Networks. Blockchain-managed systems.

Networks of people, computers, etc.
Data Science Algorithms

Pathfinding

- Single Source Short Path
- All-Nodes SSP
- Parallel paths

Determines the importance of distinct nodes in the network
- PageRank
- Betweenness
- Closeness
- Degree

Centrality

Community Detection

Evaluates how a group is clustered or partitioned
- Label Propagation
- Union Find
- Strongly Connected Components
- Louvain
- Triangle-Count

Finds the optimal path or evaluates route availability and quality
Graph Use Cases
Common Use Cases for Graph Technology

- Real-Time Recommendations
- Fraud Detection
- Network & IT Operations
- Master Data Management
- Knowledge Graph
- Identity & Access Management

Logos: Walmart, Itaú, HP, Cisco, AirBnb, UBS
Background

- International Consortium of Investigative Journalists (ICIJ), a team of data journalists.
- International investigative team specializing in cross-border crime, corruption, and accountability of power.
- Works regularly with leaks and large datasets.

Business Problem

- Find relationships between people, shell companies, and offshore accounts.
- Journalists are non-technical users.
- Biggest “Snowden-style” document leak ever: 11.5 million documents, 2.6 terrabytes of data.

Solution and Benefits

- Pulitzer Prize winning investigation resulted in robust coverage of fraud and corruption.
- PM of Iceland & Pakistan resigned, exposed Putin, Prime Ministers, gangsters, celebrities.
- Led to assassination of journalist in Malta.
ICIJ Paradise Papers

**Business Problem**
- Find relationships between people, corporations, accounts, shell companies and offshore accounts
- Journalists are non-technical
- 2017 Leak from Appleby tax sheltering law firm matched 13.4 million account records with public business registrations data from across Caribbean

**Solution and Benefits**
- Exposed tax sheltering practices of Apple, Nike
- Revealed hidden connections among politicians and nations, like Wilbur Ross & Putin’s son in law
- Triggered government tax evasion investigations in US, UK, Europe, India, Australia, Bermuda, Canada and Cayman Islands within 2 days.
- $1M endowment from Golden Globes’ HFPA

**Background**
- International Consortium of Investigative Journalists (ICIJ), Pulitzer Prize winning journalists
- Fourth blockbuster investigation using Neo4j to reveal connections in text-based, and account-based data leaked from offshore law firms and government records about the “1% Elite”
- Appends Neo4j-based, “Offshore Leaks Database”
11.5 million documents

Emails, Scanned Documents, Bank Statements etc...

2.6 TB
11.5 million documents
Emails, Scanned Documents, Bank Statements etc...

Bank US
Bank Bahamas
Company
Person A
Person B
Address

NODE
RELATIONSHIP

REGISTERED
IS_OFFICER_OF
LIVES_AT
HAS
WITH
For a distinguished example of explanatory reporting that illuminates a significant and complex subject, demonstrating mastery of the subject, lucid writing and clear presentation, using any available journalistic tool, Fifteen thousand dollars ($15,000).

International Consortium of Investigative Journalists, McClatchy and Miami Herald

For the Panama Papers, a series of stories using a collaboration of more than 300 reporters on six continents to expose the hidden infrastructure and global scale of offshore tax havens. (Moved by the Board from the International Reporting category, where it was entered.)
Visualize the ICIJ Dataset
Density Drives Value In Graphs

Metcalfes Law of the Network, Value = \( n^2 \)

5 hops: Limited Value

100's of hops: IMMENSE VALUE
Paradise Papers – Metadata Model
3 Types of Knowledge Graphs

**Context Rich Search**
- Internal knowledge documents & files, with meta data tagging
  - Examples:
    - MDM, Search
    - Customer support
    - Document classification

**External Insight Sensing**
- External data source aggregation mapped to entities of interest
  - Examples:
    - Supply chain/compliance risk
    - Market activity aggregation
    - Sales opportunities

**Enterprise NLP**
- Graph technical terms, acronyms, abbreviations, misspellings, etc.
  - Examples:
    - Improved search
    - Chatbot implementation
    - Improved classification

---

**Logical Warehouse**

**Real-Time Warehouse**

**Context Independent Warehouse**
Handling Large Graph Work Loads for Enterprises

Real-time promotion recommendations
- Record “Cyber Monday” sales
- About 35M daily transactions
- Each transaction is 3-22 hops
- Queries executed in 4ms or less
- Replaced IBM Websphere commerce

Marriott’s Real-time Pricing Engine
- 300M pricing operations per day
- 10x transaction throughput on half the hardware compared to Oracle
- Replaced Oracle database

Handling Package Routing in Real-Time
- Large postal service with over 500k employees
- Neo4j routes 7M+ packages daily at peak, with peaks of 5,000+ routing operations per second.
Background
- Large Public University – “U-Dub”
- IT staff for 80K+ students and employees
- Transforming IT systems from mainframe to cloud
- Providing IT & data warehousing services to 3 campuses, 6 hospitals, and 6,300 EDW users

Business Problem
• Old Sharepoint metadata was too complicated for users, not flexible and not transparent
• $1B project to migrate HR system from mainframe to Workday needed to be smooth
• Future projects needed repeatable predictability
• Needed new glossary, impact analysis, analytics

Solution and Benefits
• Consulted with NDU peers, built simple model
• Built Visualizer with Elasticsearch, Neo4j & D3.js
• Improved predictability, lineage, and impact understanding for over 6,300 users
Customer/360 and Deduplication go hand-in-hand

Deduplication success benefits from connections between multiple sources of data
eBay ShopBot

Background
• Personal shopping assistant
• Converses with buyer via text, picture and voice to provide real-time recommendations
• Combines AI and natural language understanding (NLU) in Neo4j Knowledge Graph
• First of many apps in eBay's AI Platform

Business Problem
• Improve personal context in online shopping
• Transform buyer-provided context into ideal purchase recommendations over social platforms
• "Feels like talking to a friend"

Solution and Benefits
• 3 developers, 8M nodes, 20M relationships
• Needed high-performance traversals to respond to live customer requests
• Easy to train new algorithms and grow model
• Generating revenue since launch
Thomson Reuters Graph

- Data Fusion for Portfolio Managers
- Graph layers

GRAPH ANALYTICS

A one week snapshot:
- 6,778 news articles with company news where at least one organization has 80% relevance to the article
- 135,267 companies are 2 steps away
- 217,387 strategic relationships
- Typical analyst portfolio is 200 companies
- Each customer creates their own relative weights for each type of relationship
- Requires around 800,000 shortest path calculations to deliver the ranked news feed. Each calculation optimised to take 10ms.
In-Q-Tel’s Mission Economy

- Venture Capital sponsored by National Intelligence
- Decomposes and reassembles technology stacks into common “genome” vocabulary
- Matches mission problems to technology assemblies and vendors
- Evaluates tech across communications, Bio tech, robotics, software, hardware, IoT
- Faster evaluations, better innovations
Background

- Largest Cable TV & Internet Provider in US
- 3rd Largest network on the planet
- xFi is consumer experience in 3M houses
- Internet, router, devices, security, voice & telephony
- Transformational customer experience

Business Problem

- Integrate all experience in a smart home
- Create innovative ideas based on cross-platform and household member preferences
- Add integrated value of xFinity triple play & quad-play services (internet, VoIP, cable TV & home security)

Solution and Benefits

- Custom content per household member
- Security reminders (kids are home, garage left open)
- Serves millions of households
- Makes content recommendations based on occupant, time of day, permissions and preferences
- Siri-like voice commands.
<table>
<thead>
<tr>
<th>Software</th>
<th>Financial Services</th>
<th>Telecom</th>
<th>Retail &amp; Consumer Goods</th>
<th>Media &amp; Entertainment</th>
<th>Other Industries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitney Bowes</td>
<td>UBS</td>
<td>COMCAST</td>
<td>eBay</td>
<td>Gamesys</td>
<td>Airbus</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>ADP</td>
<td></td>
<td>Walmart</td>
<td>IEEE</td>
<td></td>
</tr>
<tr>
<td>Digitate</td>
<td>RBS</td>
<td></td>
<td>B&amp;H</td>
<td>Nexeven</td>
<td>Marriott</td>
</tr>
<tr>
<td>Gainsight</td>
<td>Unibanco</td>
<td></td>
<td>Hästens</td>
<td>Quantone</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>Ria</td>
<td></td>
<td>Telenor</td>
<td>Onefinestay</td>
<td></td>
</tr>
<tr>
<td>Lufthansa</td>
<td>Societe Generale</td>
<td></td>
<td>Telia</td>
<td>ScrippsNetworks</td>
<td>Monsanto</td>
</tr>
<tr>
<td>Godfrance</td>
<td>Nomura</td>
<td></td>
<td>Schleich</td>
<td>Beamly</td>
<td></td>
</tr>
<tr>
<td>Virtual</td>
<td>Zurich</td>
<td></td>
<td>CenturyLink</td>
<td>FT Financial Times</td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2017 Neo4j, Inc.  Company Confidential
Graph Query Language: Cypher and ISO Standardization

Jeff Morris
Cypher: Powerful & Expressive Query Language

MATCH (:Person { name: "Dan"}) -[:MARRIED_TO]-> (spouse)
Nodes
• Can have *Labels* to classify nodes
• Labels have *native indexes*

Relationships
• Relate nodes by type and direction

Properties
• Attributes of Nodes & Relationships
• Stored as Name/Value pairs
• Can have indexes and composite indexes
• Visibility security by user/role

Neo4j Invented the Labeled Property Graph Model

- PERSON
  - name: “Dan”
  - born: May 29, 1970
  - twitter: “@dan”

- PERSON
  - name: “Ann”
  - born: Dec 5, 1975
  - MARRIED TO

- CAR
  - brand: “Volvo”
  - model: “V70”
  - Latitude: 37.562990°
  - Longitude: -122.325530°

- PERSON
  - born: Jan 10, 2011
  - LIVES WITH

- PERSON
  - born: Dec 5, 1975
  - OWNS

- PERSON
  - born: May 29, 1970
  - DRIVES

- CAR
  - brand: “Volvo”
  - model: “V70”
  - Latitude: 37.562990°
  - Longitude: -122.325530°
The GQL Manifesto: https://gql.today/

- Introduced this week: https://gql.today/
- An initiative to immediately rally support for a unified *Graph Query Language*
- Standards meetings are ongoing
- All community members are encouraged to Vote their support at https://gql.today/#vote
Graph Analytics with Neo4j

Jeff Morris
Four Pillars of Graph-Enhanced AI

1. Knowledge Graphs
   Context for Decisions

2. Connected Feature Extraction
   Context for Accuracy

3. Graph-Accelerated AI
   Context for Efficiency

4. AI Explainability
   Context for Credibility
Data Science Algorithms

Pathfinding

Finds the optimal path or evaluates route availability and quality
- Single Source Short Path
- All-Nodes SSP
- Parallel paths

Centrality

Determines the importance of distinct nodes in the network
- PageRank
- Betweenness
- Closeness
- Degree

Community Detection

Evaluates how a group is clustered or partitioned
- Label Propagation
- Union Find
- Strongly Connected Components
- Louvain
- Triangle-Count
Machine Learning & Graph Algorithms in Neo4j

**GraphConnect 2017**
- Parallel Breadth First Search & DFS
- Shortest Path
- Single-Source Shortest Path
- All Pairs Shortest Path
- Minimum Spanning Tree

**GraphConnect 2018**
- A* Shortest Path
- Yen’s K Shortest Path
- K-Spanning Tree (MST)

**Pathfinding & Search**
- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- PageRank

**Centrality / Importance**
- Harmonic Closeness Centrality
- Dangalchev Closeness Centrality
- Wasserman & Faust Closeness Centrality
- Approximate Betweenness Centrality
- Personalized PageRank

**Community Detection**
- Triangle Count
- Clustering Coefficients
- Connected Components (Union Find)
- Strongly Connected Components
- Label Propagation
- Louvain Modularity – 1 Step

**Similarity & ML Workflow**
- Euclidean Distance
- Cosine Similarity
- Jaccard Similarity
- Random Walk
- One Hot Encoding

Reference Implementations for Graph Embeddings (Node to Vector)
- DeepGL
- DeepWalk

2019 Q1
Graph Algorithms: Practical Examples in Apache Spark and Neo4j, 1st Edition

by Mark Needham (Author), Amy E. Hodler (Author)

Paperback

$59.99

Pre-order

This title will be released on June 4, 2019.
Ship from and sold by Amazon.com. Gift-wrap available.

FREE Shipping.

More Buying Choices

1 New from $59.99
Different Data Types Morph Tables into Graphs, Graphs into Tables

Morpheus for Apache Spark: June 2018 GA

Future: Any Kettle Source RDBMS & JSON

Future: Other Graph Data Sources
APPENDIX: Neo4j Competitive Differentiation
Roles & Projects around Enterprise Data Hub

- **Developers & Prod Managers**
- **Data Scientists**
- **Analysts and Business Users**
- **Chief Officers of …**
  - Compliance, Data, Digital, Information, Innovation, Marketing, Operations, Risk & Security…
- **Enterprise Data Hub/Lake Architects**
- **ID, Auth & Security**
- **Real-time Graph traversal Applications**
- **Real-Time Cyber Security**
- **Network & IT Ops**
- **Metadata & Risk Management**
- **Marketing Customer 360**
- **Digital Transformation Initiatives**
- **Knowledge Graphs**

The image describes various roles and projects around an enterprise data hub, including developers, data scientists, analysts, business users, chief officers, enterprise data hub/lake architects, ID, auth & security, real-time graph traversal applications, real-time cyber security, network & IT ops, metadata & risk management, marketing customer 360, and digital transformation initiatives.
Key Neo4j Database Components

1. Index-Free Adjacency
   - In memory and on flash/disk
   - At Write Time: data is connected as it is stored

2. ACID Foundation
   - Required for safe writes

3. Full-Stack Clustering
   - Causal consistency

4. Language, Drivers, Tooling
   - Developer Experience, Graph Efficiency, Type Safety

5. Graph Engine
   - Cost-Based Optimizer, Graph Statistics, Cypher Runtime

6. Hardware Optimizations
   - For next-gen infrastructure
## Connectedness Differentiates Neo4j

<table>
<thead>
<tr>
<th></th>
<th>Native Graph DB</th>
<th>Non-Native Graph DB</th>
<th>RDBMS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conceive</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td><strong>Code</strong></td>
<td>Cypher</td>
<td>Cypher/Gremlin/Proprietary</td>
<td>SQL</td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td></td>
<td><img src="image4.png" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td><strong>Compute</strong></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td><strong>Store</strong></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>

*Optimized for graph workloads*
Neo4j Technology Differentiation & Value

Performance
- Index-free adjacency delivers millions of hops per second
- In-memory pointer chasing

<table>
<thead>
<tr>
<th>Workload</th>
<th>Non-native graph DB: 6 machines, each with 48 VCPUs, 256 GB disk and 256 GB of RAM</th>
<th>Neo4j: single thread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count nodes</td>
<td>201s</td>
<td>&lt; 1ms</td>
</tr>
<tr>
<td>Count outgoing rels</td>
<td>203s</td>
<td>&lt; 1ms</td>
</tr>
<tr>
<td>Count outgoing rels at depth 2</td>
<td>276s</td>
<td>23s</td>
</tr>
<tr>
<td>Count outgoing rels at depth 3</td>
<td>511s</td>
<td>423s*</td>
</tr>
<tr>
<td>Count nodes by property val</td>
<td>212s</td>
<td>8s</td>
</tr>
<tr>
<td>Group rels by type</td>
<td>198s</td>
<td>54s</td>
</tr>
<tr>
<td>Count depth 2 know-likes</td>
<td>324s</td>
<td>140s*</td>
</tr>
<tr>
<td>Page Rank</td>
<td>2571s</td>
<td>27s*</td>
</tr>
</tbody>
</table>

Agility
- Native property graph model
- Modify schema as business changes without disrupting existing data

Developer Productivity
- Easy to learn, declarative openCypher graph query language
- Procedural language extensions & open library of extensions
- Worldwide developer community

Hardware Efficiency
- Native graph query processing and storage often requires 10x less hardware.
- Index-free adjacency requires 10x less CPU.
Different Paradigms

Collections-Focused
Multi-Model, Documents, Columns & Simple Tables, Joins

Connections-Focused
Focused on Data Relationships

Neo4j Graph Platform

NoSQL
Relational DBMS

Neo4j is designed for data relationships

Development Benefits
Easy model maintenance
Easy query

Deployment Benefits
Ultra high performance
Minimal resource usage
Competitive Entrants

50% increase in new Graph Technologies in last 2 years
# Emerging Competitive Landscape

**NEO4J OUTPACES COMPETITION IN GRAPH LEADERSHIP & INVESTMENT, TECHNOLOGY CAPABILITY, COMMUNITY BREADTH AND PRODUCT MATURITY**

<table>
<thead>
<tr>
<th>Existing databases</th>
<th>Cloud vendors</th>
<th>Early stage Graph databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>DataStax</td>
<td>AWS Amazon Neptune</td>
<td>ArangoDB</td>
</tr>
<tr>
<td>SAP HANA</td>
<td>Azure Cosmos DB</td>
<td>Dgraph</td>
</tr>
<tr>
<td>SQL Server</td>
<td></td>
<td>Memgraph</td>
</tr>
<tr>
<td>MongoDB</td>
<td></td>
<td>OrientDB</td>
</tr>
<tr>
<td>Oracle</td>
<td></td>
<td>TigerGraph</td>
</tr>
</tbody>
</table>

**Strong with ‘non-graph’ workloads but perform poorly for graphs**
- Slow, corrupt data in production, or both
- Lacking in graph affordances
- No graph community here
- But: market advantage

**Nascent product offering significantly behind Graph**
- Products designed for cloud operator, not end user
- Arbitrary technical limits
- Some end user convenience (e.g. billing)
- But: market & relationship advantage

**Playing ‘catch-up’ aggressively**
- Mostly irrelevant, lacking any realistic community or production following
- But Monash’s rule applies, TG is potentially competitive. Dgraph may be at some point too
## Connectedness Differentiates Neo4j

<table>
<thead>
<tr>
<th>Concept</th>
<th>Native Graph DB</th>
<th>Non-Native Graph DB</th>
<th>RDBMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceive</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Table" /></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Cypher</td>
<td>Cypher/Gremlin</td>
<td>SQL</td>
</tr>
<tr>
<td>Compute</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td>Table</td>
</tr>
<tr>
<td>Store</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td>Table</td>
</tr>
</tbody>
</table>

Optimized for graph workloads
Thank You

Jeff Morris