Graph Engines: Database + Compute

CNSV
September 12, 2017
ABOUT ME

• Developed web apps for 5 years

• Worked at Google for 8 years on Google Apps, Cloud Platform

• Technologies: Python, Java, BigQuery, Oracle, MySQL, OAuth

ryan@neo4j.com
@ryguyrg
ACID
The Impact of Connected Data

“Graph analysis is possibly the single most effective competitive differentiator for organizations pursuing data-driven operations and decisions”

Gartner®
Connected Data is Transforming Industries

Social Graph
LinkedIn

People & Products
Walmart

People & Content
adidas
Introducing Neo4j
Neo4j is the world’s leading database for connections in data. We power the next generation applications and analytics, in areas like machine learning, personalized recommendations, fraud detection, and data governance.
Neo4j is the world’s leading database for connections in data. We power the next generation applications and analytics, in areas like machine learning, personalized recommendations, fraud detection, and data governance.
Neo4j is the world’s leading database for connections in data. We power the next generation applications and analytics, in areas like machine learning, personalized recommendations, fraud detection, and data governance.
This is why digital native companies like Medium, Ebay, and LinkedIn, as well as companies in transformation like Walmart, Adidas and Airbus, have chosen to adopt Neo4j.
Hundreds of successful deployments ranging from both Fortune 500 companies as well as exciting startups

Examples of enterprise adoption:

- UBS
- eBay
- orange
- LendingClub
- IBM
- ICIJ
- FT
- MITRE
- LinkedIn
- Cerved
- Medium
- Microsoft
- Cisco
- Monsanto
- Pitney Bowes
- Walmart
Adoption Highlights
(As per 2017)

Retail
7 out of top 10 retailers in the world

Finance
12 out of 25 top financial services firms

Software
8 out of top 10 software vendors
“Graph Thinking” is all about considering connections in data as important as the data itself — and how this is reflected in your business.

Neo4j is an internet-scale, native graph database which executes connected workloads faster than any other database management system.

3.5M downloads
500+ partners
47,000 group members
61,000 trained engineers
Mindset

“Graph Thinking”
Neo4j — Native Graph Platform

Data Modelled as a Graph

RELATIONSHIP

NODE

Cypher
Graph API
Native Graph Engine
Relationship Queries on non-native Graph Architectures

Using Other NoSQL to Join Data

Slow queries due to index lookups & network hops

Using Neo4j

Lightning-fast queries due to replicated in-memory architecture and index-free adjacency
How Data Structures Affects your Organisation
Hierarchies | Linear Supply Chain | Information

**Business Processes**

On Stage

Behind the Scene

**Data Structure**
Business Processes

On Stage

Behind the Scene

Data Structure
How Graphs Work to Connect your data
Store / Retrieve
Store / Retrieve
Store / Retrieve
Store / Retrieve
Store / Retrieve
Load Data

Store / Retrieve

Actionable Insights
Leveraging Cross-Silo Connections
Case Studies
Common Graph DB Use Cases

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

Companies: Walmart, Fortune 100, HP, Cisco, Lufthansa, UBS
An engineer working on the Orion mission found information from the Apollo project, using Neo4j, which prevented an issue and saved “well over two years of work and one million dollars of taxpayer funds.”

— David Meza, Chief Knowledge Architect, NASA

Built a Knowledge Graph of lessons learned from all of NASA’s prior missions.

David Meza, Chief Knowledge Architect, says “There are many different ways to look and search for information rather than just a keyword search. And I think utilizing new types of graph databases and other types of NoSQL databases really showcases this – often there are better ways than a traditional relational database management system.”
How to Get Started with Neo4j
CREATE MODEL  LOAD DATA  QUERY DATA
Property Graph Model

Ann Loves Dan
Property Graph Model

CREATE (:Person { name: "Ann"}) - [:LOVES]-> (:Person { name: "Dan"})

NODE

CREATE (:Person { name: "Ann"}) - [:LOVES]-> (:Person { name: "Dan"})

NODE

LABEL

PROPERTY

LABEL

PROPERTY
Querying in the Browser
Querying with Code
Official Language Drivers

JS
Java
Node.js
Python
Community Language Drivers
Extending Neo4j
User-Defined Procedures & Functions

Java Stored Procedures let you write custom code that is:

- Written in any JVM language
- Stored on the server
- Accessed by applications over Bolt
- Executed on the server

```
CALL apoc.create.uuid() YIELD uuid
CALL apoc.date.formatDefault(timestamp(),"ms") YIELD value as date
CREATE (:Document {id:uuid, created:date})
```
APOC Library
I can't believe that @neo4j is actually real. Seems like a dream come true
Neo4j Sandbox

neo4j.com/sandbox
Architectural Options

End User

Bulk Analytic Infrastructure
Hadoop, EDW...

Data Mining and Aggregation

Data Scientist

Graph Database Cluster
Neo4j
Neo4j
Neo4j

Ad Hoc Analysis

Data Storage and Business Rules Execution

Application

Databases
Relational NoSQL Hadoop
Two Editions: Community and Enterprise
Download Neo4j 3.2.2

Install Neo4j and Take it for a Spin

Experience powerful scalability, blazing speed and unparalleled flexibility – download and try Neo4j today.

For Business
The Neo4j Enterprise Edition offers incredible power and flexibility, with enterprise-grade availability, management and scale-up & scale-out capabilities.

For Individuals
Ideal for learning and smaller do-it-yourself projects that do not require high levels of scaling. Excludes professional services and support.

DOWNLOAD FREE ENTERPRISE TRIAL

DOWNLOAD COMMUNITY EDITION
OLTP vs OLAP
(aka Persistence vs Analytics)
Intuitiveness
Speed
Agility
igraph — The network analysis package

igraph is a collection of network analysis tools with the emphasis on efficiency, portability and ease of use. igraph is open source and free. igraph can be programmed in R, Python and C/C++.
Graph Analytics

GraphX is Apache Spark's API for graphs and graph parallel computation.

Flexibility
Seamlessly work with both graphs and collections.

GraphX unifies ETL, exploratory analysis, and iterative graph computation within a single system. You can view the same data as both graphs and collections, transform and join graphs with RDDs efficiently, and write custom iterative graph algorithms using the Pregel API.

Using GraphX in Scala

```
graph = Graph(vertices, edges)
messages = spark.textFile("hdfs://...")
graph2 = graph.leftVertices(messages)
{
  id, vertex, msg1 = ...
}
```

Download Spark
Speed

Simplicity
Moved from Twitter Search API to Streaming API

Replaced Python Twitter libraries (Tweepy) with raw API calls.

Queue streaming tweets in message queue
- Collect Tweets
- Collect User Data

Store full tweets & user information in MongoDB.

Build graph for analysis in Neo4j from tweets persisted in MongoDB

Analysis done in R
- Graph Statistics
- Community Detection

iGraph graph libraries for graph algorithms
- Also some text analysis e.g. LDA Topic analysis

Results pushed out to MySQL for Tableau Dashboard
- Graphml file for import into Gephi with all stats precalculated

Analysis done in R
- Graph Statistics
- Community Detection

Graphml file for import into Gephi with all stats precalculated
Moved from Twitter Search API to Streaming API

Replaced Python Twitter libraries (Tweepy) with raw API calls.

Queue streaming tweets in message queue
Store full tweets & user information in MongoDB.
Build graph for analysis in Neo4j from tweets persisted in MongoDB

Analysis done in R
iGraph graph libraries for graph algorithms
Also some text analysis e.g. LDA Topic analysis

Results pushed out to MySQL for Tableau Dashboard
Graphml file for import into Gephi with all stats precalculated
**Efficient Graph Algorithms for Neo4j**

http://neo4j-contrib.github.io/neo4j-

Add topics

- **213** commits
- **11** branches
- **3** releases
- **6** contributors
- **GPL-3.0**

<table>
<thead>
<tr>
<th>Branch: 3.1</th>
<th>New pull request</th>
<th>Clone or download</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Message</th>
<th>Commit ID</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>knutwalker</td>
<td>committed with jexp</td>
<td>commit message</td>
<td>a day ago</td>
</tr>
<tr>
<td></td>
<td>Preparatory Refactoring for Parallel Node Imports (#375)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>algo</td>
<td>Directed ShortestPath Dijkstra (#374)</td>
</tr>
<tr>
<td>benchmark</td>
<td>Preparatory Refactoring for Parallel Node Imports (#375)</td>
</tr>
<tr>
<td>core</td>
<td>Preparatory Refactoring for Parallel Node Imports (#376)</td>
</tr>
<tr>
<td>doc</td>
<td>Add triangles to index docs (#351)</td>
</tr>
<tr>
<td>tests</td>
<td>Directed ShortestPath Dijkstra (#374)</td>
</tr>
<tr>
<td>.gitignore</td>
<td>Ldbc benchmark for LabelPropagation (#143)</td>
</tr>
<tr>
<td>.travis.yml</td>
<td>Run benchmarks on travis with 4g memory</td>
</tr>
<tr>
<td>LICENSE</td>
<td>Adapt License to GPLv3 due to the use of Neo4j Java API</td>
</tr>
<tr>
<td>pom.xml</td>
<td>Adapt License to GPLv3 due to the use of Neo4j Java API</td>
</tr>
</tbody>
</table>
Simplicity

Speed

Simplicity
What’s Next?
GraphConnect
Powered by Neo4j

Connecting Data to Drive Innovation

REGISTER NOW

50% off for CNSV Members by 9/19
Code: “CNSV50”
Thank you!
@ryguyrg
devrel@neo4j.com