Fontys UNIVERSITY OF APPLIED SCIENCES

FOR SOCIETY

Neo4j

M. Gorlas, N. Adamavicius

FHTenL

April 19, 2024

Roadmap

1. Data model and schema
2. Consistency and Replication
3. Security and Performance
4. Specific use cases and bit of history
5. Demo

Fontys
UNIVERSITY OF
APPLIED SCIENCES

Data model

Neo4j is a **graph database**.
There are two main types of graph databases:

- Property graph model
- RDF graph model

Fontys
UNIVERSITY OF
APPLIED SCIENCES

Property graph model

In Neo4j information is organized as nodes, relationship and properties.

Property graph model

In Neo4j information is organized as nodes, relationship and properties.

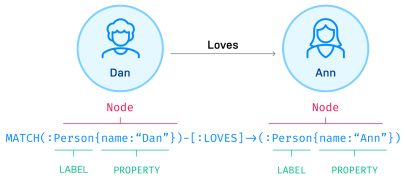

Figure 1: Example of the property graph model[1]

In the property graph model, **nodes** are the **entities** in the graph. **Relationship** provide directed, named connections between two nodes entities.

[1] Neo4j Inc. *What is a graph database?* URL: https://neo4j.com/docs/getting-started/get-started-with-neo4j/graph-database/.

Database schema 1/2

- Neo4j is **schema-less**
- Limitations can be added manually (e.g. constraints for uniqueness)
- Neo4j is just dealing with instance model directly without node predefinition

Database schema - example 2/2

You don’t need to tell that **Person** node consists of specific properties, you start creating a node with label **Person** for Bob and give it all properties with values directly.

- Validation is left up to you (constraints or by application).

Fontys
UNIVERSITY OF
APPLIED SCIENCES

Consistency

- Neo4j provides strong consistency.
- Neo4j employs causal consistency model.
- Isn’t recommended DBMS when working with WAN.

Replication

.

- Raft protocol.[2]
- Leader/Follower Model. [2]
- High Availability.[3]

[2] Neo4j Inc. *Leadership, routing and load balancing*. URL: https://neo4j.com/docs/operations-manual/current/clustering/setup/routing/.

[3] Sonal Raj. *Neo4j High Performance*. URL: https://www.oreilly.com/library/view/neo4j-high-performance/9781783555154/ch06s07.html.

Fontys UNIVERSITY OF APPLIED SCIENCES

Security

- Schema-based Security[4]
- Role-based access control[5]

[4] Neo4j Inc. *Fine-Grained Access Control for Better Security and Privacy*. URL: https://neo4j.com/product/neo4j-graph-database/security/.

[5] Neo4j Inc. *Built-in roles and privileges*. URL: https://neo4j.com/docs/operations-manual/current/authentication-authorization/built-in-roles/#auth-built-in-roles.

Fontys UNIVERSITY OF APPLIED SCIENCES

Schema-based Security

- Protect the nodes and relationships by controling users' ability to traverse and read from different parts of the graph.
- Ensures that only authorized users have access to the data they need to protect sensitive data.

Role-based access control

- An approach, where you can apply restrictions to roles assigned to users at any level of granularity throughout the graph.
- Simplifies the task of assigning permissions and helps ensure that your data is secure.

Performance

- Compared to relational DBMS (MySQL in this case)
- Compared to other NoSQL DBMS

Compared to MySQL

Based on the benchmark using real-world data from Career Village, the experiment done by Rodrigues et. al, showed that Neo4j was faster than MySQL in most cases, particularly in pattern matching and recursive queries. However, MySQL has advantages in terms of data consistency and transactional support.

Category	Query	Neo4j	MySQL
Selection	Q1	2ms	31ms
	Q2	8ms	323ms
	Q3	32ms	438ms
Recursive	Q4	2ms	757ms
	Q5	2ms	290ms
	Q6	3ms	305ms
Aggregation	Q7	43ms	146ms
	Q8	18ms	40ms
	Q9	62ms	290ms
Pattern Matching	Q10	5ms	360ms
	Q11	10ms	455ms
	Q12	1ms	68ms

Figure 2: Performance comparison between Neo4j and MySQL[6]

[6] Cajetan Rodrigues, Mit Ramesh Jain, and Ashish Khanchandani. *Performance Comparison of Graph Database and Relational Database*. May 2023. DOI: 10.13140/RG.2.2.27380.32641.

Fontys UNIVERSITY OF APPLIED SCIENCES

Compared to MySQL

In the benchmark the following types of queries were used:

- selection/search
- recursion
- aggregation
- pattern matching

Fontys
UNIVERSITY OF
APPLIED SCIENCES

Compared to other NoSQL DBMS (1/2)

Based on the WDBench, a benchmark for graph databases focused on querying the Wikidata, Neo4j was the slowest of all tested graph databases, on all types of queries.[7]

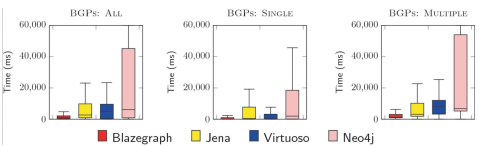

Figure 3: Basic Graph Patterns queries

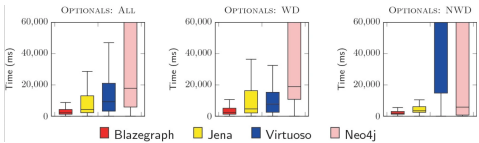

Figure 4: Optional Graph Patterns queries

[7] Renzo Angles et al. *WDBench: A Wikidata Graph Query Benchmark*. Oct. 2022. DOI: 10.1007/978-3-031-19433-7_41.

Compared to other NoSQL DBMS (2/2)

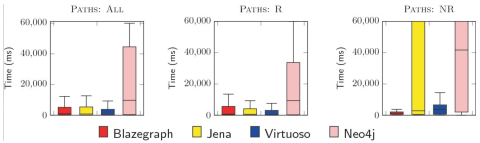

Figure 5: Path Patterns queries

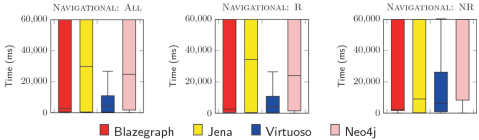

Figure 6: Navigational Graph Patterns queries

Specific use cases

- Knowledge Graphs.
- Recommendation Systems.
- Fraud Detection.

Fontys
UNIVERSITY OF
APPLIED SCIENCES

Famous cases

- NASA - uses the Neo4j knowledge graph to enhance its Lessons Learned Database, allowing engineers to identify trends and correlations between past projects and apply these insights to prevent future failures and improve decision-making.[8]
- eBay - utilizes Neo4j to power their chat bot recommendation system, enhancing user interactions with fast, context-aware responses and scalable graph database technology.[9]
- Fortune 500 Financial Services - uses Neo4j to visualize complex transaction relationships in real-time, enabling analysts to quickly detect and stop fraudulent activities, thereby saving thousands of dollars daily.[10]

[8] Neo4j Inc. *Getting to Mars Faster with a Knowledge Graph*. URL: https://neo4j.com/case-studies/nasa/.

[9] Neo4j Inc. *Neo4j Powers Intelligent Commerce for eBay App on Google Assistant*. URL: https://neo4j.com/case-studies/ebay/.

[10] Neo4j Inc. *Real-Time Graph Analysis Creates Potential for Millions in Fraud Detection Savings*. URL:

Fontys UNIVERSITY OF APPLIED SCIENCES

Demo

Fontys
UNIVERSITY OF
APPLIED SCIENCES