Using Neo4j in a Rails Application: One Programmer's Experience

Joe Morgan
9mmedia

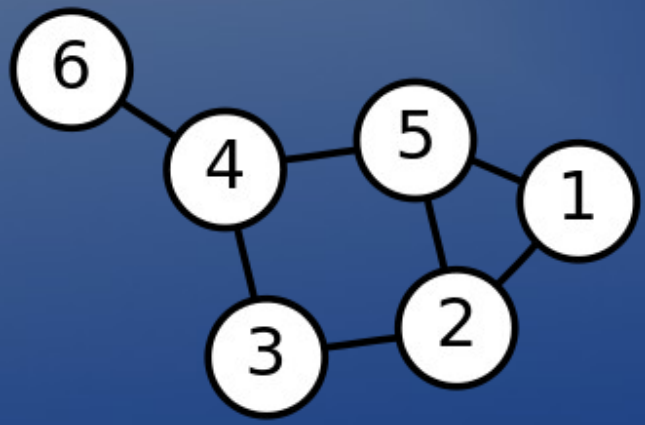
What is Neo4j?

- Neo4j is an open source graph database written in Java.
- Neo4j's data model is a property graph: Nodes, relationships and properties.

What is a graph?

“In mathematics, a graph is a representation of a set of objects where some pairs of the objects are connected by links.”

“Graph (mathematics)” http://en.wikipedia.org/wiki/Graph_(mathematics)

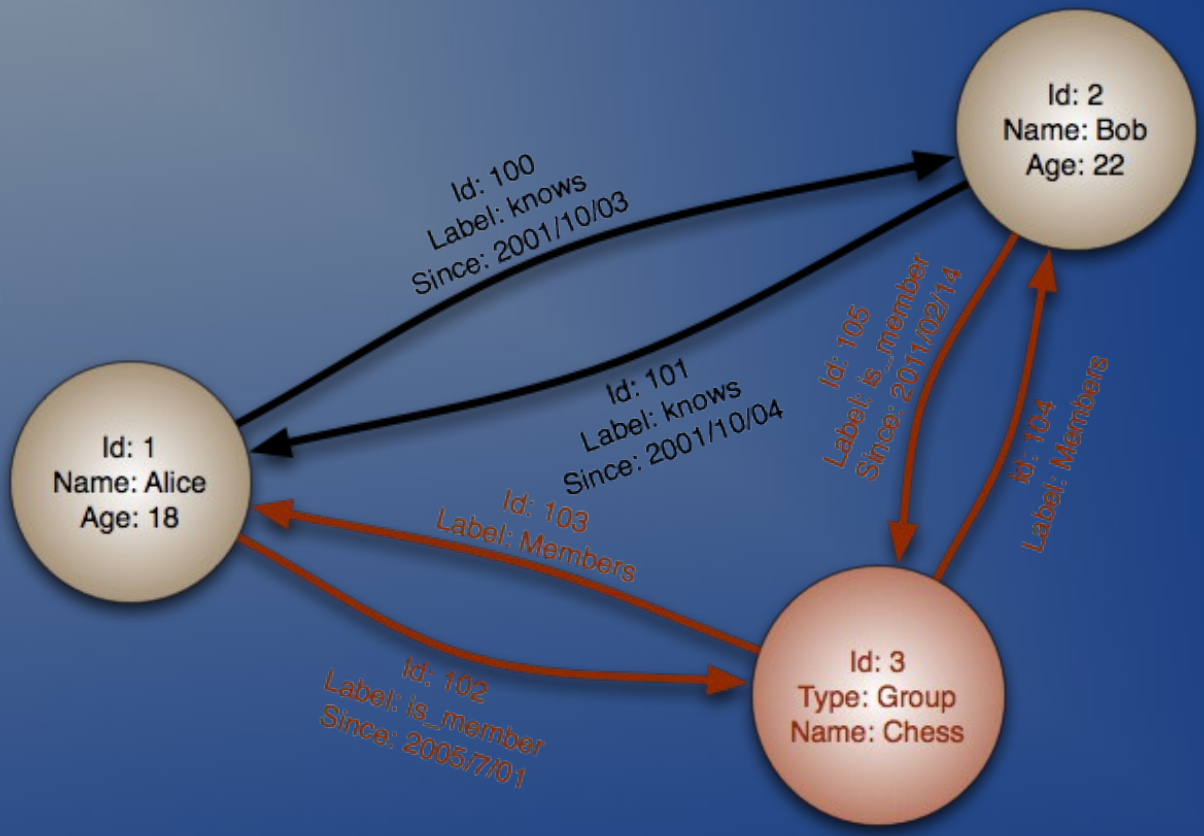


What is a graph database?

“A graph database uses graph structures with nodes, edges, and properties to represent and store data. By definition, a graph database is any storage system that provides index-free adjacency.”

“Graph Database” http://en.wikipedia.org/wiki/Graph_database

What is a graph database?



Graph database vs. RDBMS

Graph database

- Data is relational.
- Data structure is determined on-the-fly.
- Less ACID-ic. Tradeoffs typically made to optimize for speed and availability (ACID vs BASE).

RDBMS

- Data is tabular.
- Data structure is determined in advance.
- Overhead for consistency checks (key contraints, triggers, etc.)

Graph database vs. key-value store

Graph database

- Better for more structured data.
- Better for queries that require “joining” objects.

Key-value store

- Simple R/W operations.
- Ideal for “schema-less” data models.
- Less ideal for mapping relationships between objects and performing traversals.

Graph database vs. document database

Graph database

- Less emphasis on internal node structure.
- Support for querying on node and relationship properties varies among implementations.
- Support for arbitrary and ad-hoc relationships.

Document database

- Ideal for semi-structured objects (documents).
- Support for querying objects beyond simple key value look-ups.
- Relationships are usually supported but highly structured.

When might a graph database be appropriate?

Often, when you're using an RDBMS or key-value store to recreate a graph:

- Tables with lots of columns, each of which is only used by a few rows.
- Attribute tables.
- Lots of many-to-many relationships.
- Data model has tree-like characteristics.
- Data model requires frequent schema changes.

Vicknair, et. al. “A Comparison of a Graph Database and a Relational Database” http://www.cs.olemiss.edu/~ychen/publications/conference/vicknair_acmse10.pdf

Common graph databases use cases

- Social Networking/Recommendations
- Network and Cloud Management
- Master Data Management
- Geospatial
- Bioinformatics
- Content Management and Security and Access Control

Eifram, Emil (September 26, 2012) “Graph Databases: The New Way to Access Super Fast Social Data” http://mashable.com/2012/09/26/graph-databases/

What are some of the leading graph databases?

- Neo4j (Java, open-source)
- AllegroGraph (Lisp, RDF-oriented)
- Sones GraphDB (.NET, available as SaaS)
- DEX (C++)
- InfiniteGraph (Java)
- FlockDB (Scala, developed by Twitter)

Neo4j Overview

- Disk-based storage
- Embedded and Standalone Server options
- ACID transactional
- Uses Apache Lucene as default indexing engine.

Neo4j Data Storage Overview

- Several disk stores: node store; relationship store; property store; additional stores for things like large strings, etc.
- Several layers of in-memory caching, including:
  - Memory-mapped files
  - Heap-based object graph
- Memory-map buffering is configurable

Developing with Neo4j: What are my options?

- Embedded in a Java application
- Embedded in a Python application
- Remote client libraries
- REST API

REST API

- Includes a `cypher` method to a allow running arbitrary Cypher queries against the database.
- Includes CRUD methods for nodes, relationships, properties and indexes.

Cypher

- Query language for Neo4j.
- SQL-esque syntax. Combinations of clauses:

```
START x=node:users('muzik_id:18')
MATCH x<-[:FOLLOWS]-follower
RETURN follower
ORDER BY ID(follower)
SKIP 0
LIMIT 5000;
```

- Evolving. Very much a work-in-progress.

Cypher Query Examples

Cypher Query Example Through REST API

Wrapping Up