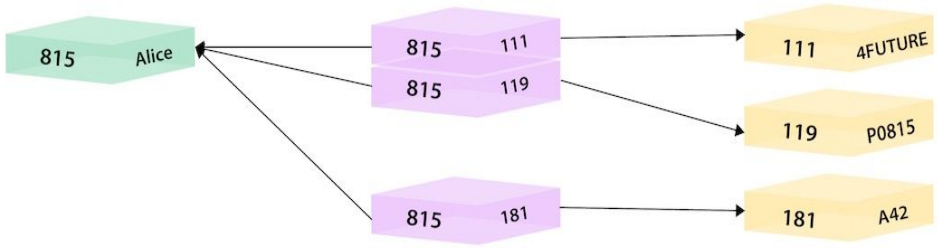
netherlands eScience center

Data SIG - Graph databases

Faruk

25/10/2018

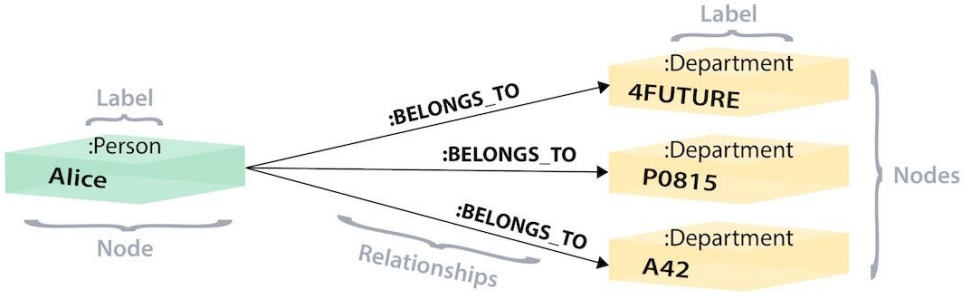
Graph databases vs Relational



Person

Person-Department

- Find user ‘Alice’ → ID of 815
- Search departments Department IDs that reference Alice
- Search Department IDs (111, 119, 181) to get more info



- Search which departments Alice belongs to

A single hop!

Single node for Alice with a **label** of Person. Alice belongs to 3 different departments, so we create a node for each one and with a label of Department.

https://neo4j.com/developer/graph-db-vs-rdbms/

List of Graph databases

https://en.wikipedia.org/wiki/Graph_database#List_of_graph_databases

Neo4j

Neo4j is one of the most popular open source graph databases.

Main Features

- Highly scalable
- Web-based administration tool
- Built-in REST web API interface
- Bolt protocol with official drivers
- Doesn't support native date & time as field type
- Languages: JAVA, .NET, JavaScript, Python, Ruby

What most users not about this graph database is its native and elegant solutions and ease of data remodeling. In Neo4j all data is stored either as a node, edge or attribute.

OrientDB

OrientDB is a Java-written NoSQL database management system and is also incredibly popular among developers.

Main Features

- Fast to install and run
- Support for SQL queries
- Native support for HTTP, RESTful protocol, JSON (libraries or components)
- Runs anywhere: Linux, Windows, OS X
- Language: Java

OrientDB has a multi-model graph engine and it supports different models: graph, document, object and key/value. This database has numerous applications, including fraud prevention and banking.

ArangoDB

This database has an open-source license and is considered one of the most popular NoSQL databases. It stands out due to its high performance and at the same time low consumption of resources.

Main Features

- Multi-model
- Has AQL query language
- Stores key/value, documents, graph data
- Works in distributed cluster
- Language: C++, Javascript

ArangoDB is often called a universal database and has a wide range of applications due to its easy solutions.

MarkLogic

MarkLogic is another Java-written and multi-model database that is known by its powerful search and flexible application services.

Main Features

- NoSQL database
- Has a built-in search engine
- Very scalable and elastic
- Distributed architecture
- Language: Java

MarkLogic is used to store documents and semantic graph data and is mainly applied in the fields with a lot of large-scale systems.

AllegroGraph

Unlike previous graph databases, AllegroGraph is a closed-source database, used for storing RDF triples.

Main Features

- Available for different platforms: Linux,. Windows, OS X
- Disk-based storage
- Supports SPARQL, RDFS++
- Includes implementation of Prolog
- Languages: C#, C, Java, Common Lisp, Python

AllegroGraph is used in commercial and open-source projects and also serves as storage component for TwitLogic.

https://dzone.com/articles/top-5-graph-databases

Multi-Model Database Use Cases

By combining the power of graphs with the flexibility of documents, multi-model databases such as OrientDB are suitable for virtually any use case. Some of the uses include:]

- **Recommendation engines:** Graphs naturally lend themselves to quickly form links between data sets and **analyze large amounts of data**. Recommendation engines are one of the main examples of how graphs can find relationships between distinct datasets to provide the best matches.
- **Banking and financial applications:** RDBM systems are simply not capable of quickly exploring relationships to uncover crimes like fraud rings or identity theft or protect sensitive banking data. Graph databases, on the other hand, **can navigate connections in real time in order to discover patterns**, match data and stop fraud before it happens.
- **Biotech and pharmaceutical applications:** Universities, governments and pharmaceutical companies are turning to graph databases to create innovative applications, **study DNA sequencing** and discover new treatments for diseases.
- **Online retail applications:** In today's world of rapidly expanding data, staying ahead of the curve means providing the fastest and most efficient method to handle online purchases. Multi-model databases exploit the advantages of graphs to **quickly find links between data** but also maximize the efficient of the application itself by handling financial, product, user session and search engine data.

https://orientdb.com/graph-database/

Technologies @NLeSC: RDF - SPARQL - grlc

Projects:

- candYgene
- DIVE+
- Data quality
- ODEX4ALL
- EOSC Pilot for LOFAR

Case law may benefit

https://grandstack.io

- **G**raphQL – a query language for APIs and a runtime for fulfilling those queries with your existing data
- **R**eact – a JavaScript library for building user interfaces
- **A**pollo Client – a fully-featured, production-ready caching GraphQL client for every server or UI framework
- **N**eo4j **D**atabase – a graph database that is ACID-compliant and built to store and retrieve connected data

```
git clone https://github.com/grand-stack/grand-stack-starter.git
cd grand-stack-starter
docker-compose up
```

Open the links in your browser:

GraphQL Playground: http://0.0.0.0:4000/

Neo4j Browser: http://localhost:7474/browser/ (user=neo4j password=letmein)

React App: http://localhost:3000/

In Neo4j Browser:

- Click on Star Icon
- Click on Example Graphs → Movie Graph
- Click on Play button on right top corner to execute the command
- Now you can follow a short tutorial to learn how to add, query data.
- In each slide, example instructions are given. Just click on Play button (on left top corner) in each code block and execute with the main Play button (right top corner of the page)
- Slides also have different options to display the result (buttons on the left of the slides): Graph, Table, Text

Extra: Northwind Graph Example → Shows how you can import CSV files and create nodes, relationships for the data

Optional - GraphQL

The default schema → api/src/graphql-schema.js

Create:

```
mutation {
  CreateBusiness(name: "nlesc" state: "growing") {
    state
    name
  }
}
```

Query:

```
{
        businesses(name: “nlesc”) {
                State
        }
}
```

Thank you.

netherlands eScience center

GRAPH DATABASE LANDSCAPE

	Real-Time Big Graph	Operational Graph	Multi-Modal Graph	Analytic Graph	RDF Graph
Vendor Examples	TigerGraph	Neo4j, Titan	DataStax Graph, Microsoft Azure Cosmos, ArangoDB	Apache Giraph, Turi	AllegroGraph, Virtuoso, Blazegraph, Stardog, GraphDB
Key Strengths & Focus	Real-time, General purpose, Scalability	General purpose	Supports multiple NoSQL data models	Analytics which can traverse the full graph	Triplet model, Semantic queries
Potential Drawbacks	No open-source vendors yet	Performance doesn't scale to Big Graphs	Compromise, not a leader in performance	Not strong on exploration or transactions	Not strong on analytics or transactions
Exploration (neighbor-hood search)	★★★	★★	★½	★	★★★
Analytics (full dataset read)	★★★	★★	★½	★★★	★
Transactions (real-time updates, concurrency)	★★★	★★	★½	★	★
Speed at Scale	★★★	★★	★	★	★