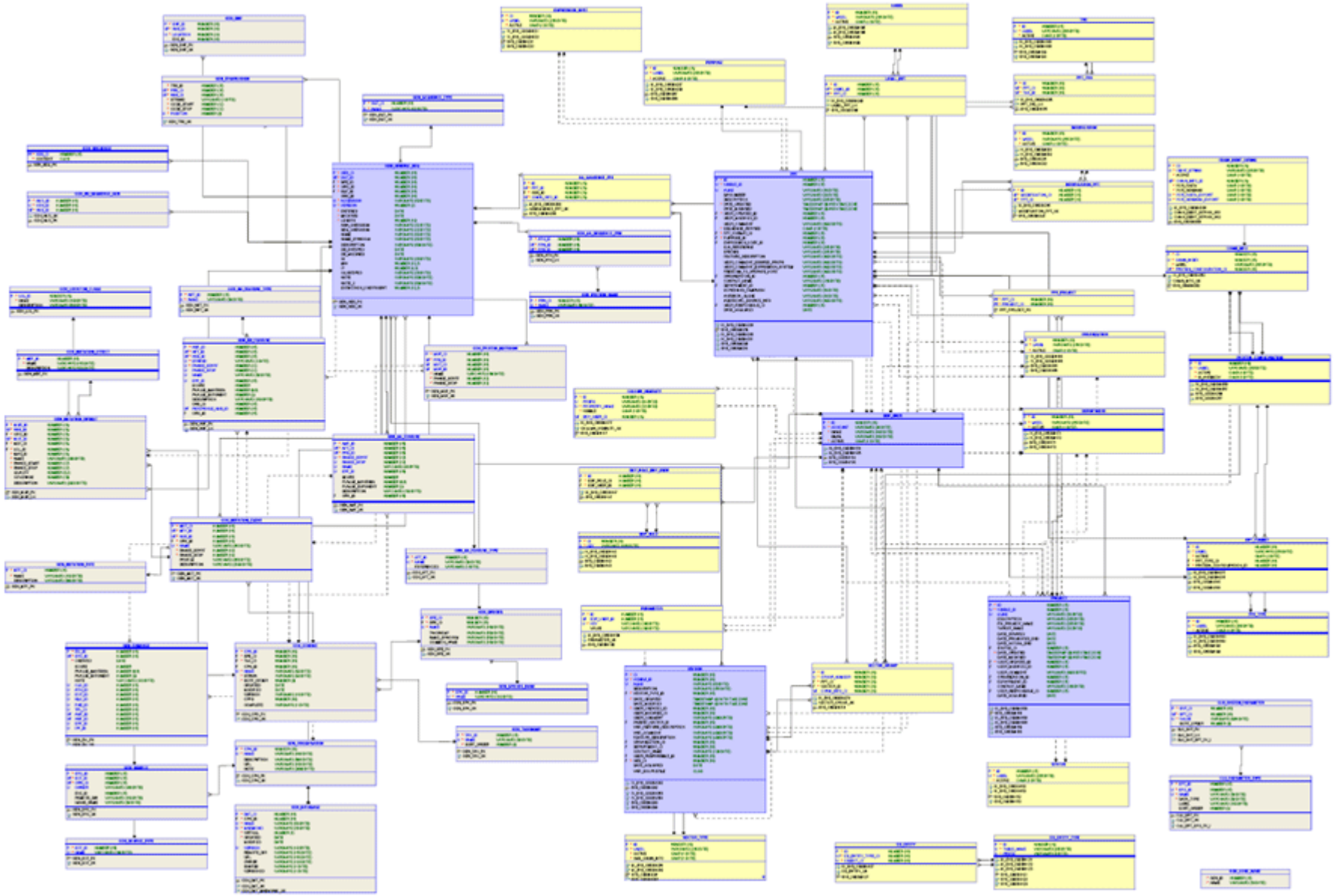
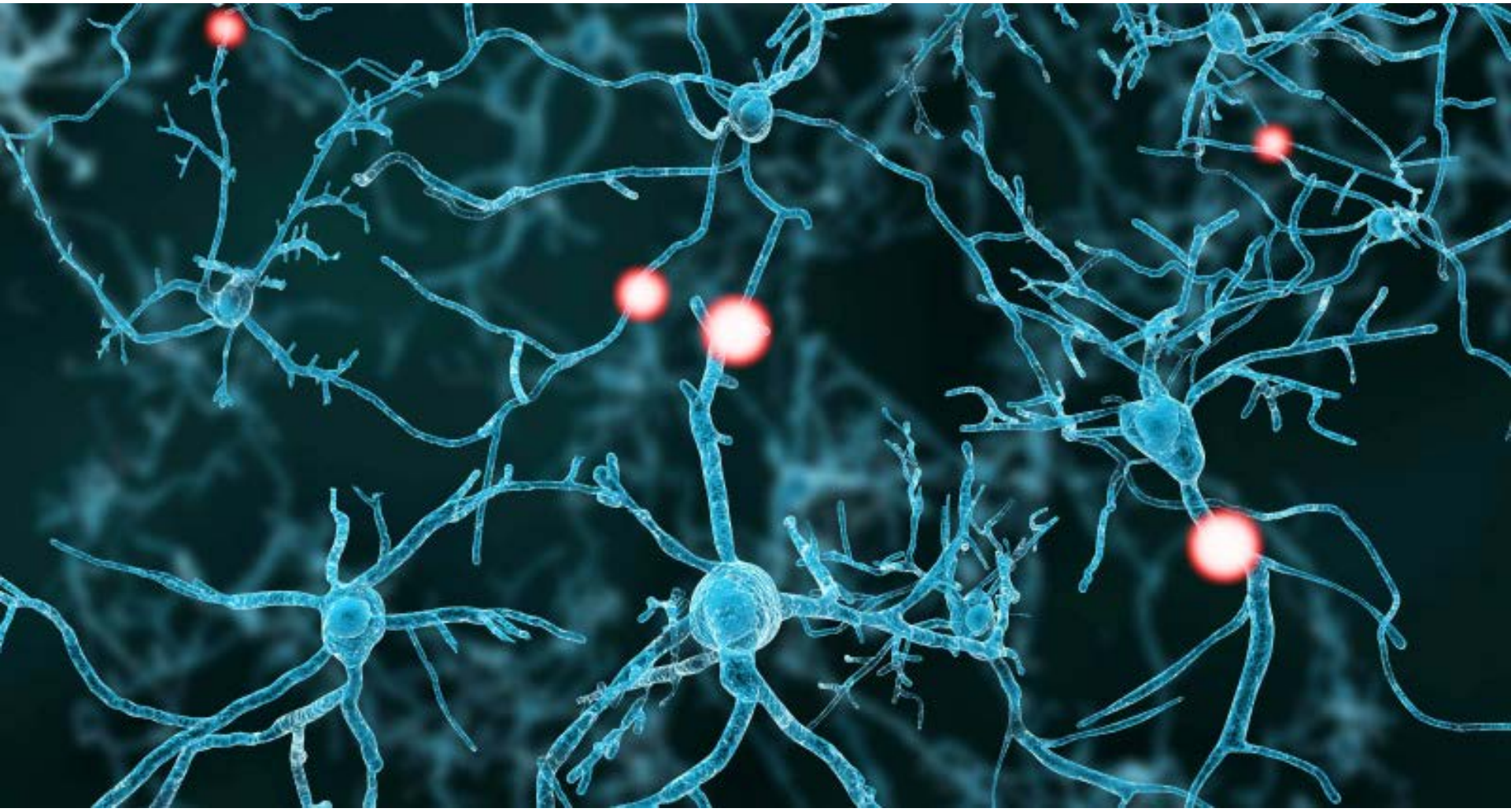
Introduction to Graph Databases and Neo4j

What Is a Graph Database?

Roland Guijt
www.rmgsolutions.nl
@rolandguijt

pluralsight
hardcore dev and IT training





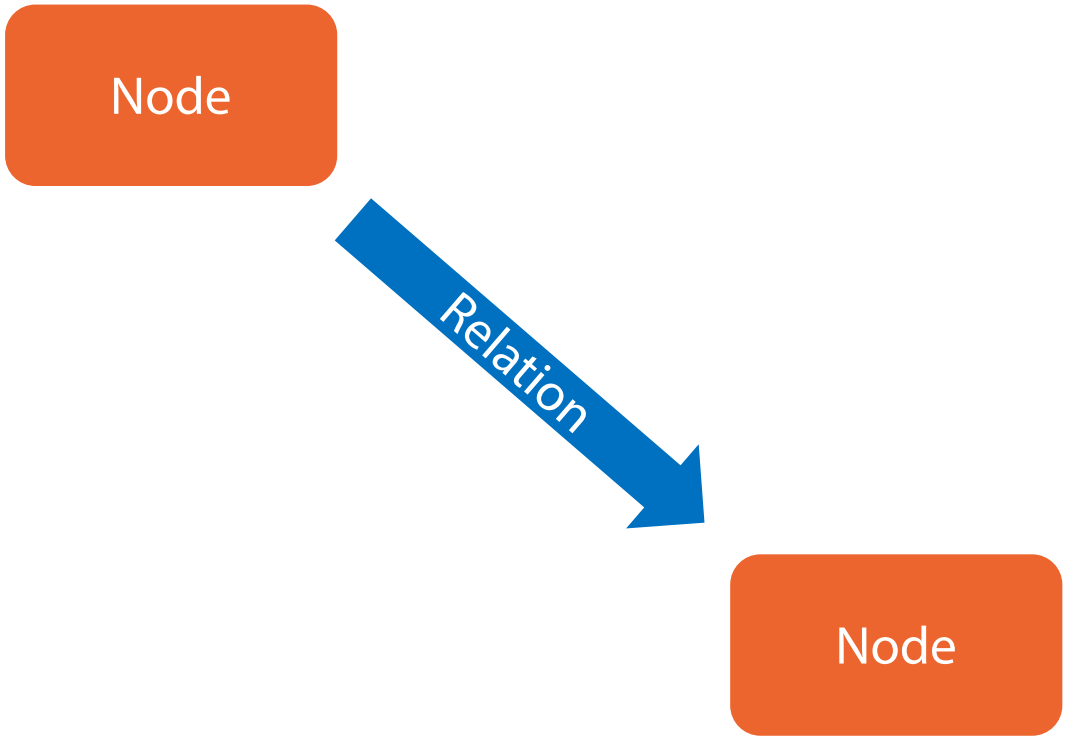
Agenda

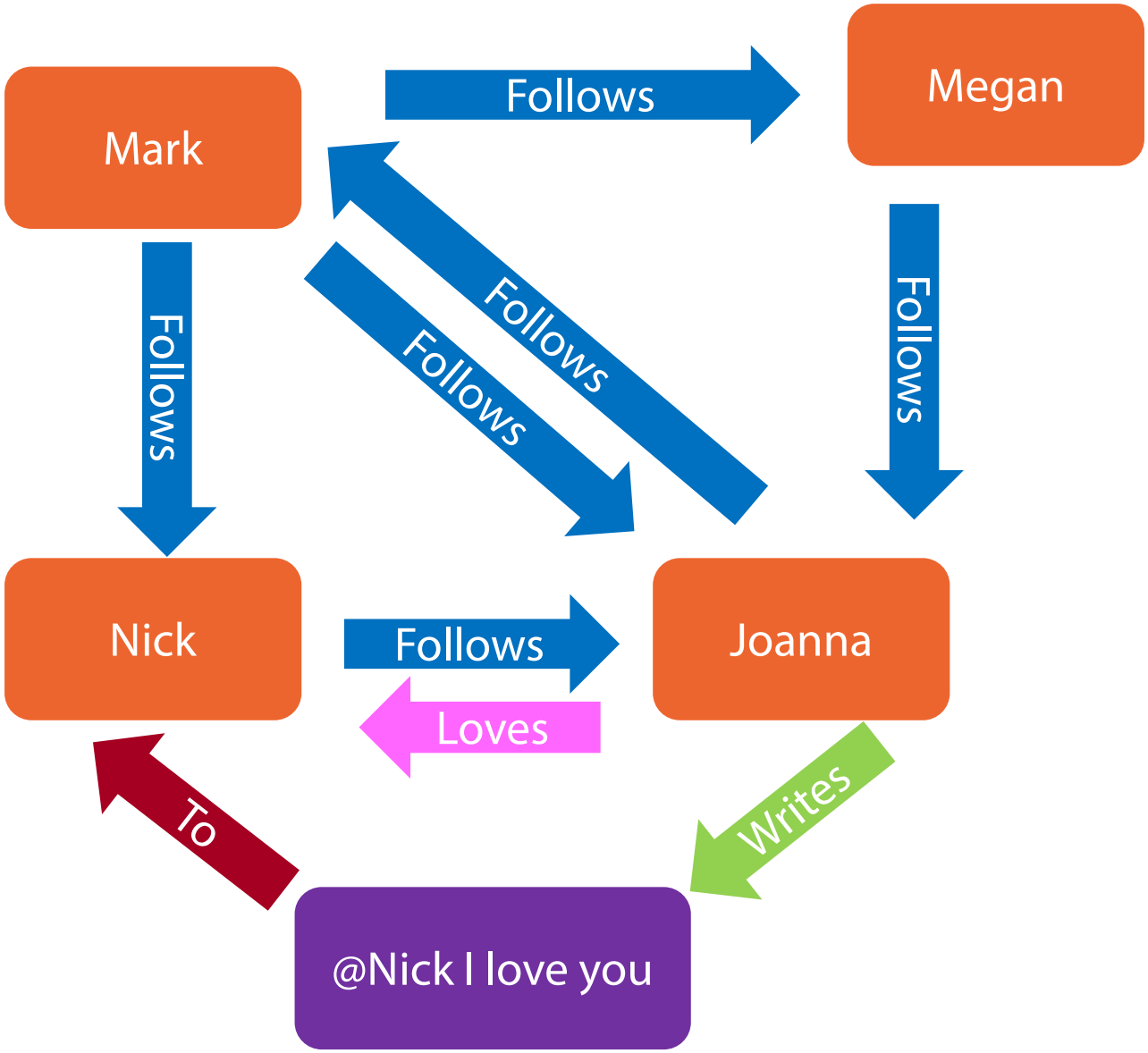
What is a Graph?	What is a Graph Database?	Why a Graph Database?
Graph Databases vs Relational Databases	Graph Databases vs Nosql Databases	Examples of Graph Databases

What Is a Graph Database?

Roland Guijt
www.rmgsolutions.nl
@rolandguijt

pluralsight
hardcore dev and IT training


Node
Relation
Node

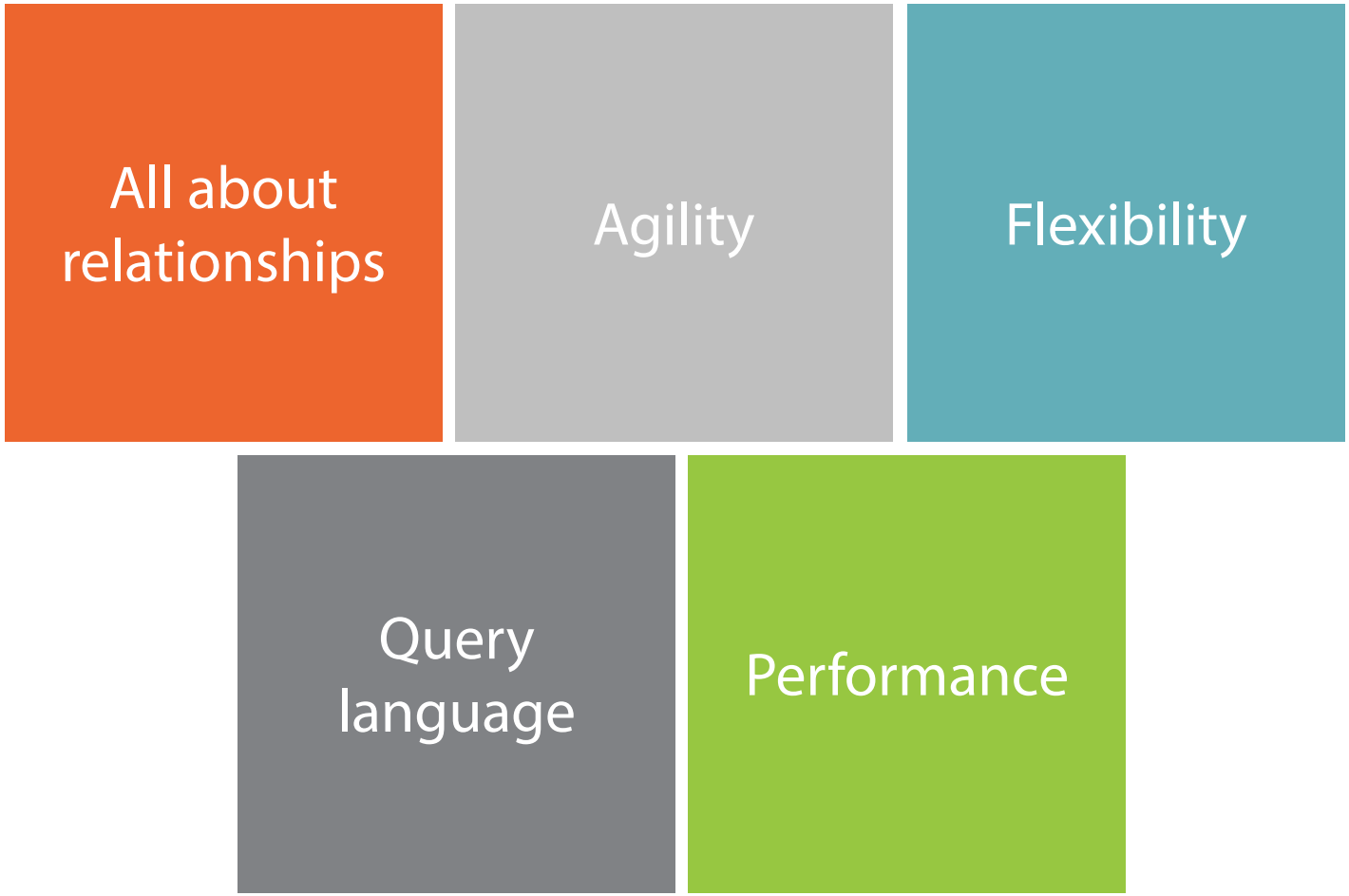

Mark
Follows
Megan
Follows
Follows
Follows
Follows
Nick
Follows
Joanna
Loves
To
Writes
@Nick I love you

Graphs

- Easily extendable and expandable
- Friendly to the human brain
- Whiteboard compatible

A **graph database** is a database that uses graph structures to represent and store data

Graph Databases



Property Graph Model

Contains **nodes** and **relationships**

Nodes and relationships contain **properties**

Relationships are **named** and **directed** with a **start** and **end** node

Why a Graph Database?

"Consider the type of database for every application you're writing"

Why a Graph Database?

Highly **related** data

Flexible schema

Structure and queries are **brain friendly (= easier)**

Graph Databases vs. Relational Databases

Relational	Graph
Tables	Nodes
Schema with nullables	No schema
Relations with foreign keys	Relation is first class citizen
Related data fetched with joins	Related data fetched with a pattern

Relational Databases Advantages

Highly **structured** data

Calculations within one table

Grouping of data

The Foreign Key System

Customer		
CustomerId	Name	City
1	Joanna	Salt Lake City

Order		
OrderId	CustomerId	Date
1	1	2015/1/1

LineItem		
OrderId	ProductId	Quantity
1	1	5

Product		
ProductId	Description	Use
1	Candle	Inside

Partner and Vukotic's Experiment

- Social network
- Friends of Friends structure
- mySql and Neo4j
- 1000.000 people
- Each with an average of 50 friends
- Depth 2: Find all friends of a user's friends
- Depth 3: Find all friends of friends of a user's friends
- Etc.

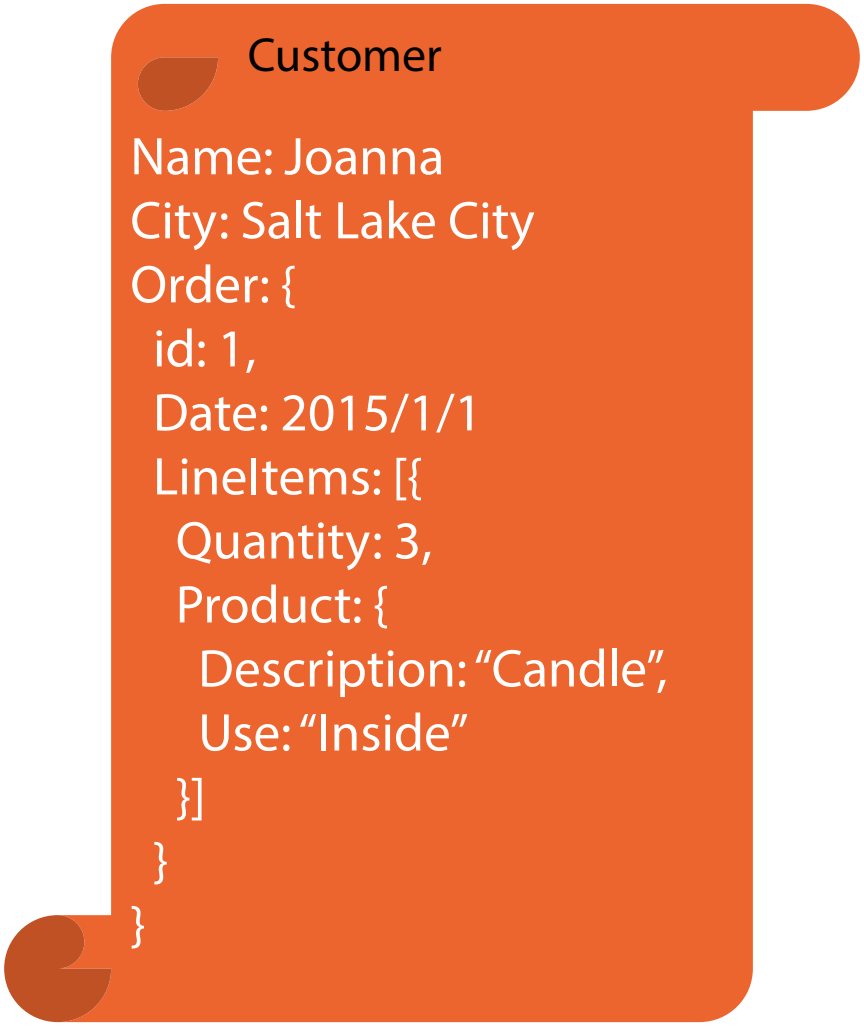
Depth	Rel. Db (s)	Neo4j (s)	# records
2	0,016	0,01	~2500
3	30,267	0,168	~110000
4	1543,505	1,359	~600000
5	Unfinished	2,132	~8000000

Relational Database Normalization

Created when disk space was **expensive**

Normalization is encouraged

A Document


Customer
Name: Joanna
City: Salt Lake City
Order: {
id: 1,
Date: 2015/1/1
LineItems: [{
Quantity: 3,
Product: {
Description: “Candle”,
Use: “Inside”
}]
}
}

Document Databases

All **related** data in one entity

Duplication of data is not something to avoid

Copy master data

Documents

Customer

```
Name: Joanna
City: Salt Lake City
Order: {
  id: 1,
  Date: 2015/1/1
  LineItems: [{
    Quantity: 3,
    Product: {
      Description: “Ca
      Use: “Inside”
    }]
  }
}
```

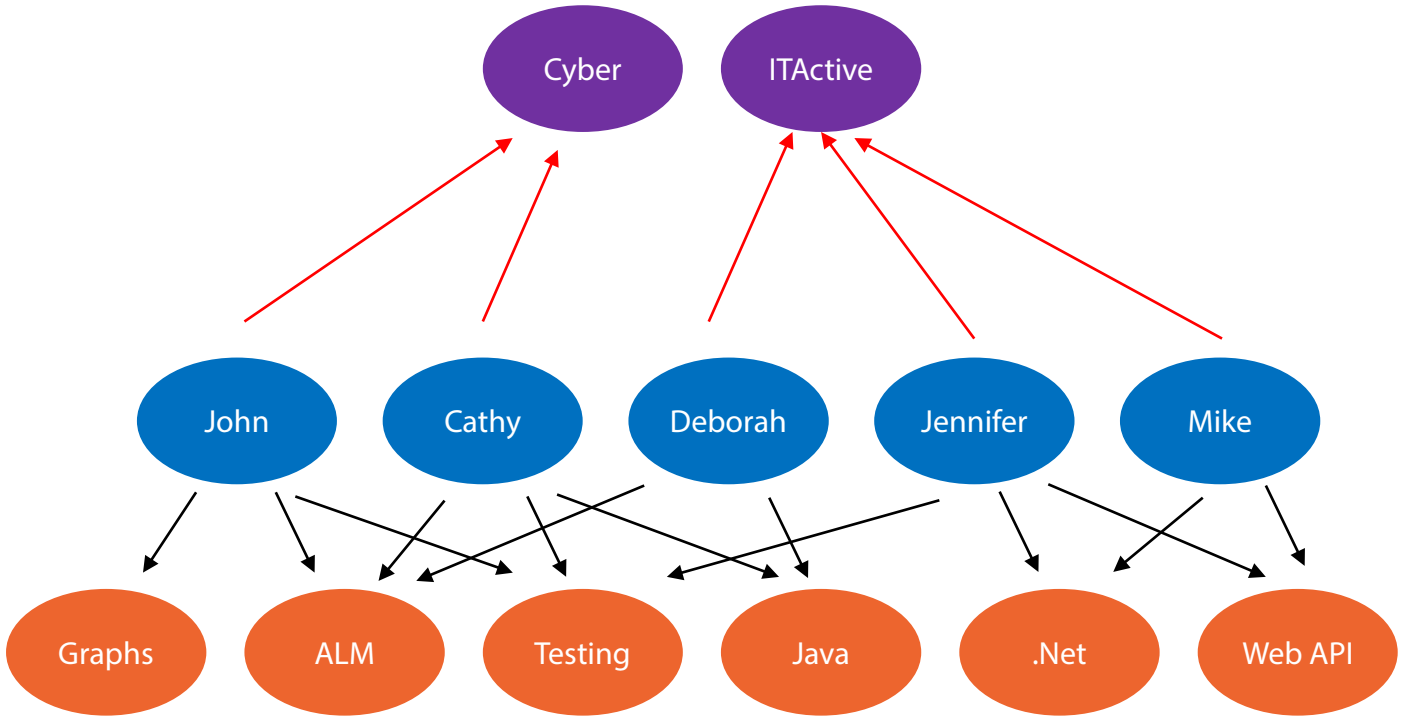
Customer

```
Name: Peter
City: Dallas
Order: {
  id: 2,
  Date: 2015/2/1
  LineItems: [{
    Quantity: 2,
    Product: {
      Description: “Matches”,
      Use: “Inside”
    }]
  }
}
```

Graph Databases vs. Document Databases

Document	Graph
Document	Nodes
No schema	No schema
Relations with foreign keys or embedded	Relation is first class citizen
Related data fetched with joins or embedded	Related data fetched with a pattern

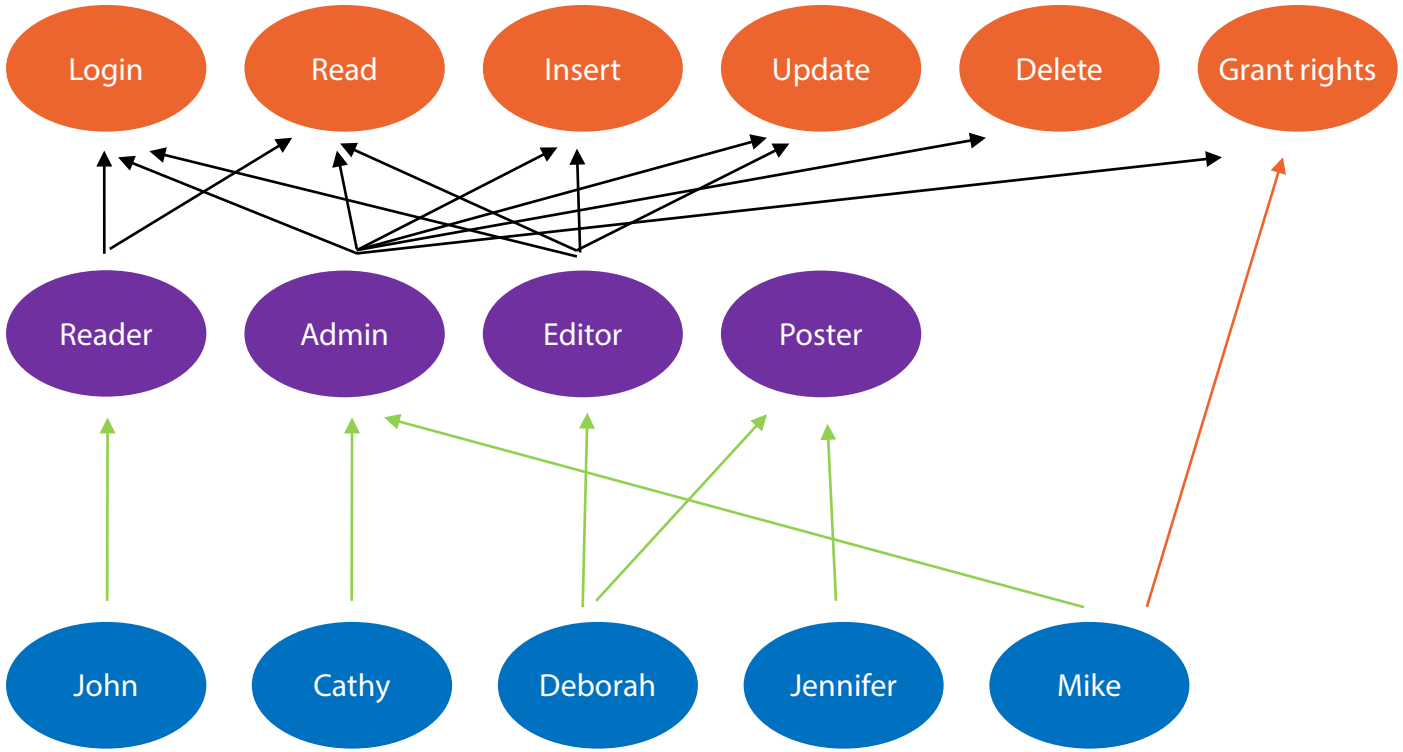
A Social Graph



Who shares Cathy's skills?

Who works in the same company as Cathy and shares the most skills?

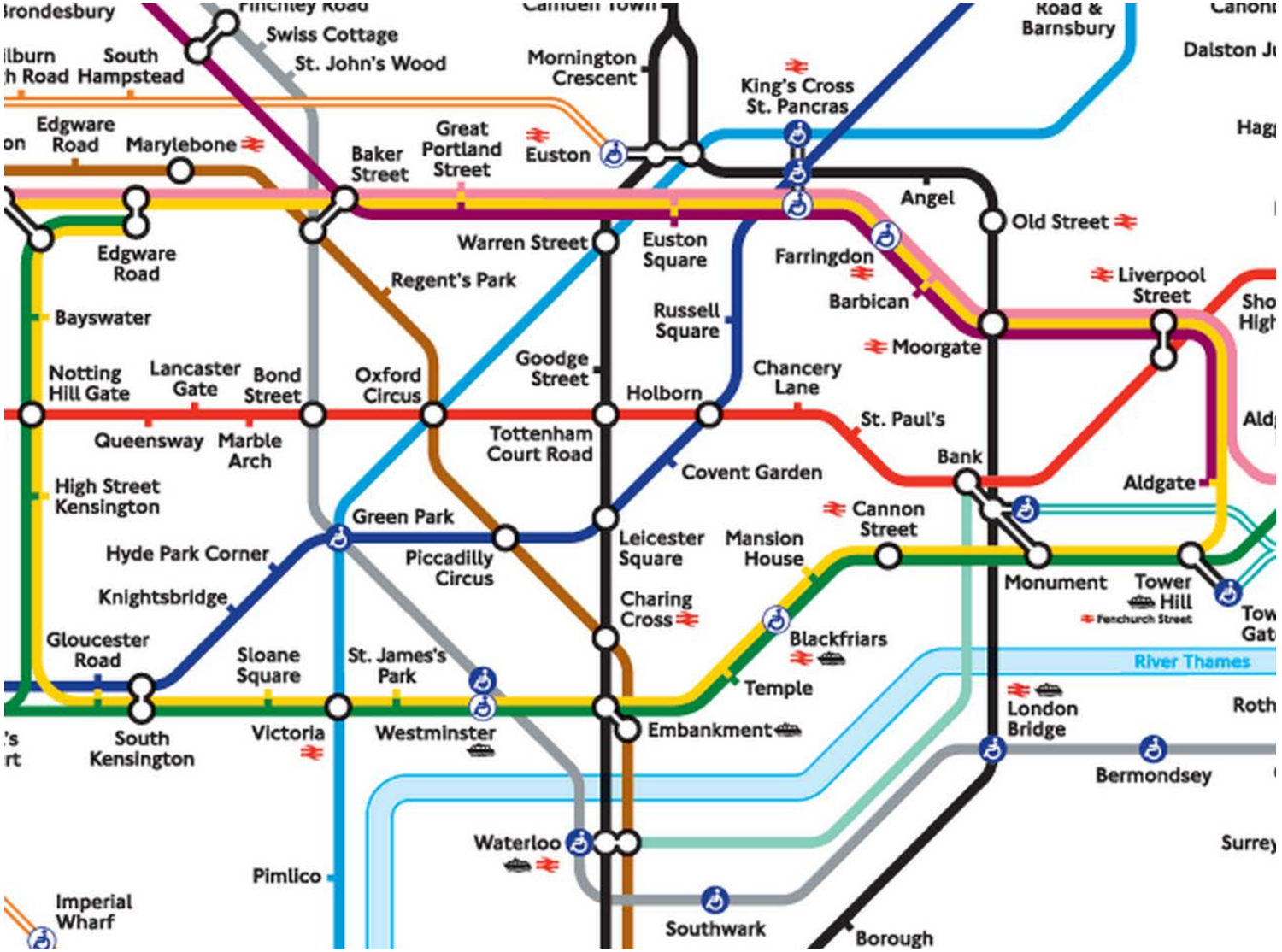
Security



Which rights does Deborah have?

Who edited a blog post and when?

Logistics


Swiss Cottage
St. John's Wood
South Hampstead
Mornington Crescent
King's Cross St. Pancras
Edgware Road
Marylebone
Baker Street
Great Portland Street
Euston
Angel
Old Street
Edgware Road
Warren Street
Euston Square
Farringdon
Liverpool Street
Regent's Park
Barbican
Bayswater
Russell Square
Moorgate
Notting Hill Gate
Lancaster Gate
Bond Street
Oxford Circus
Goodge Street
Holborn
Chancery Lane
St. Paul's
Queensway
Marble Arch
Tottenham Court Road
Bank
High Street Kensington
Covent Garden
Aldgate
Green Park
Cannon Street
Hyde Park Corner
Piccadilly Circus
Leicester Square
Mansion House
Monument
Tower Hill
Knightsbridge
Charing Cross
Fenchurch Street
Gloucester Road
Sloane Square
St. James's Park
Blackfriars
River Thames
Temple
London Bridge
South Kensington
Victoria
Westminster
Embankment
Bermondsey
Waterloo
Pimlico
Imperial Wharf
Southwark
Borough

Summary

- A graph is a collection of nodes connected by relationships
- Graph databases are flexible and performant with highly related data
- All database types have their place
- Relational database suitable for reporting and calculation on a single table. Weak point: related tables
- Document database suitable to store objects. Weak point: related documents
- Graph databases are great in many scenarios, but not all

What's Next?

- Neo4j