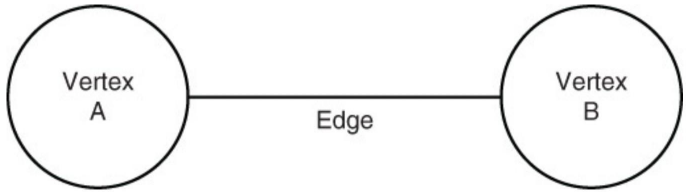
Databases and Information Systems CS303

Graph Databases
04-10-2023

What is Graph?

- Has Vertices (nodes) and Edges
- Vertices can be anything of relevance (Entities)
  - Cities, Employees, Course...
- Connection between vertices are called Edges (directed or undirected)



Graph Databases

No tables

Instead:

Nodes (vertices) and relationships (edges)

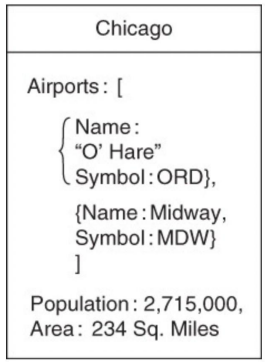
Node: has identifier and attributes

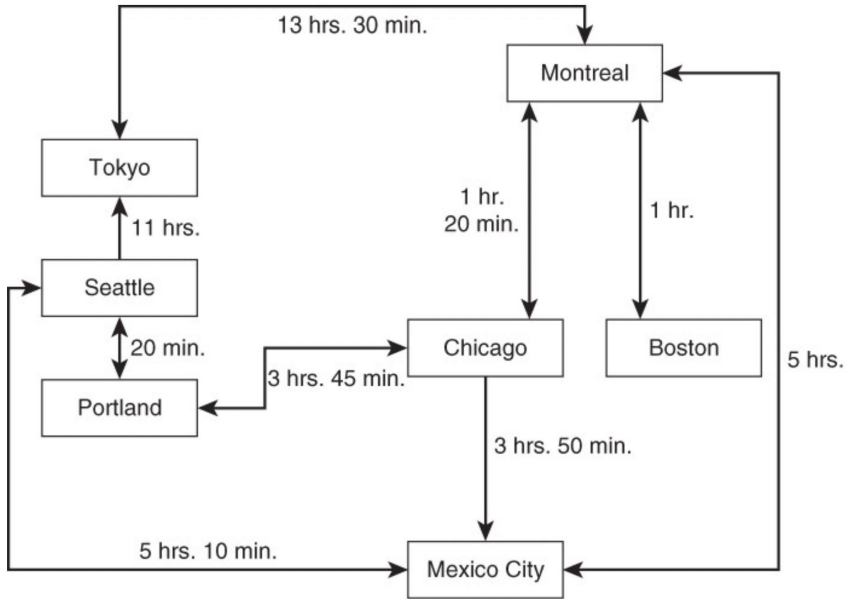
Relationship links two nodes

Graph database: Example
Travel time between cities

Nodes are cities (Nodes can have attributes)

Relationships are travel time between cities

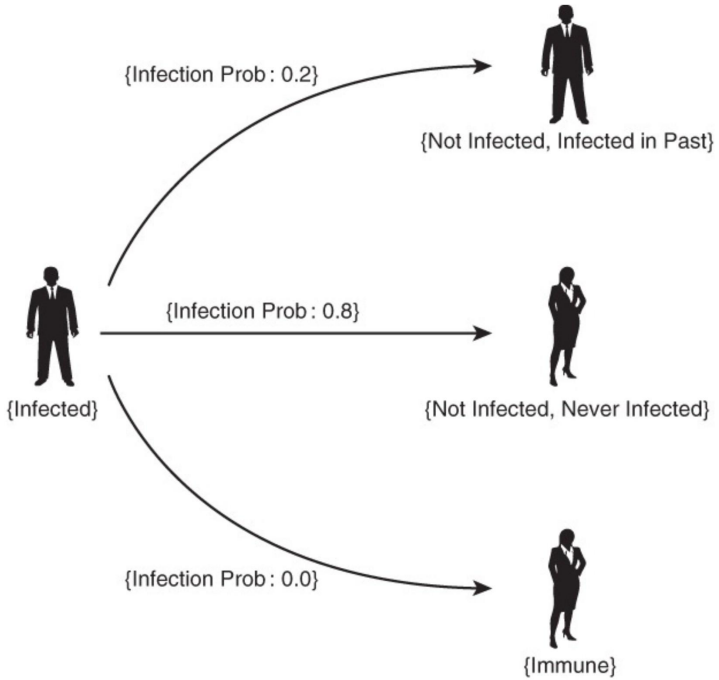




Graph database: Example Spread of Covid

Nodes are people

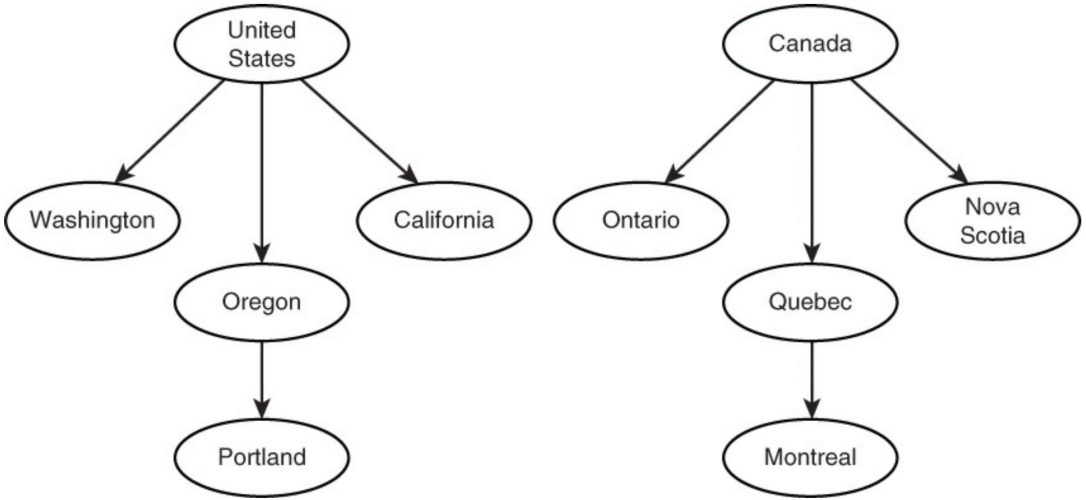
Edges: who can infect whom & with what probability.



Graph database: Example Hierarchy

Nodes are country / state…

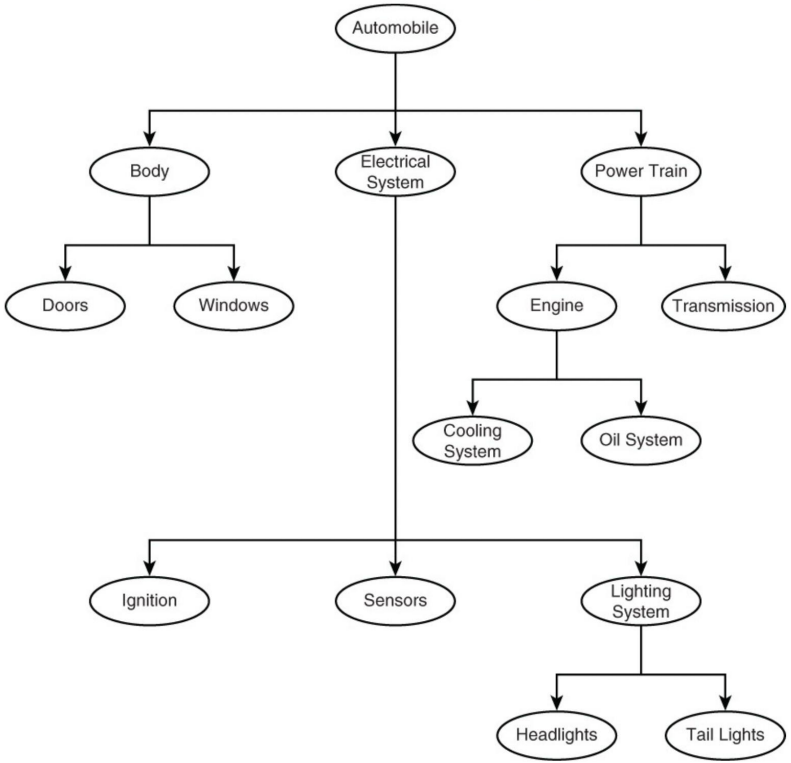
Edges: ‘is in’



Graph database: Example Parts of an automobile

Nodes are components

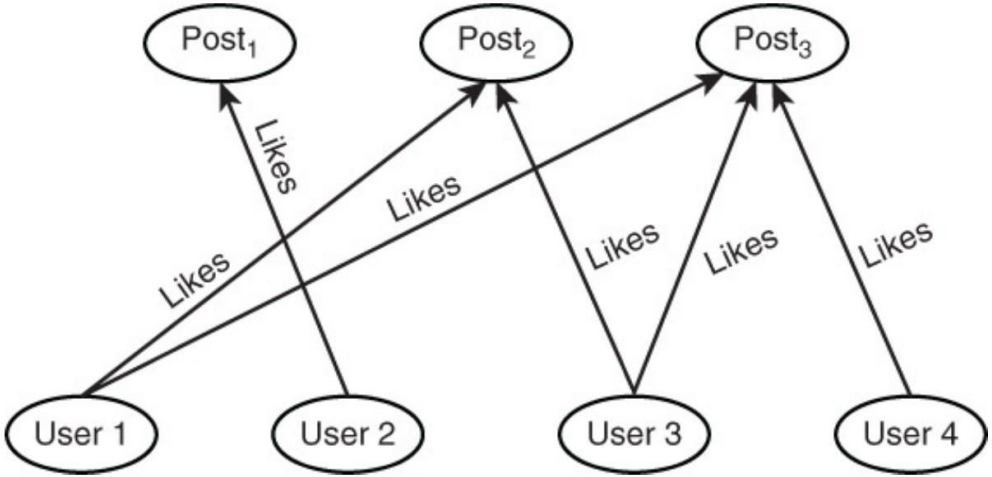
Edges: ‘composed of’



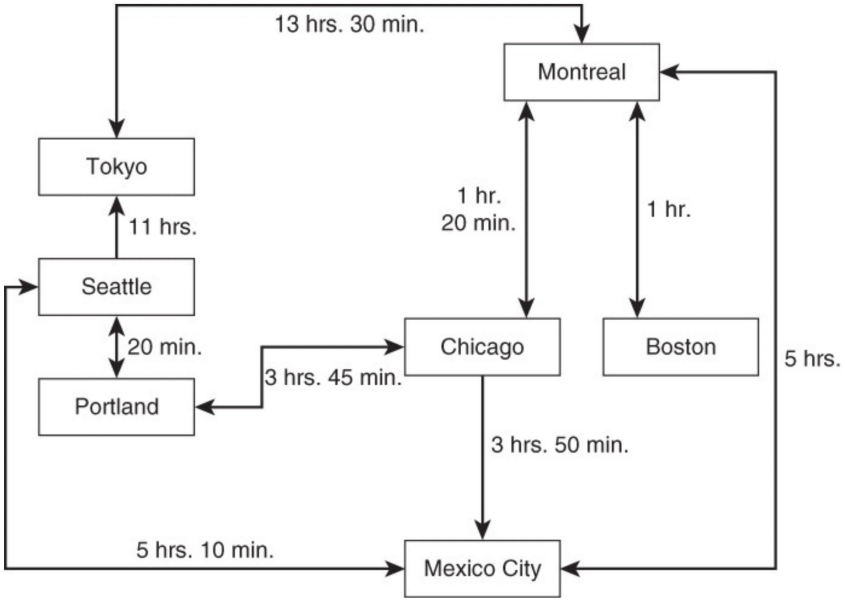
Graph database: Example
Social Media

Nodes are users and posts

Edges: ‘likes’

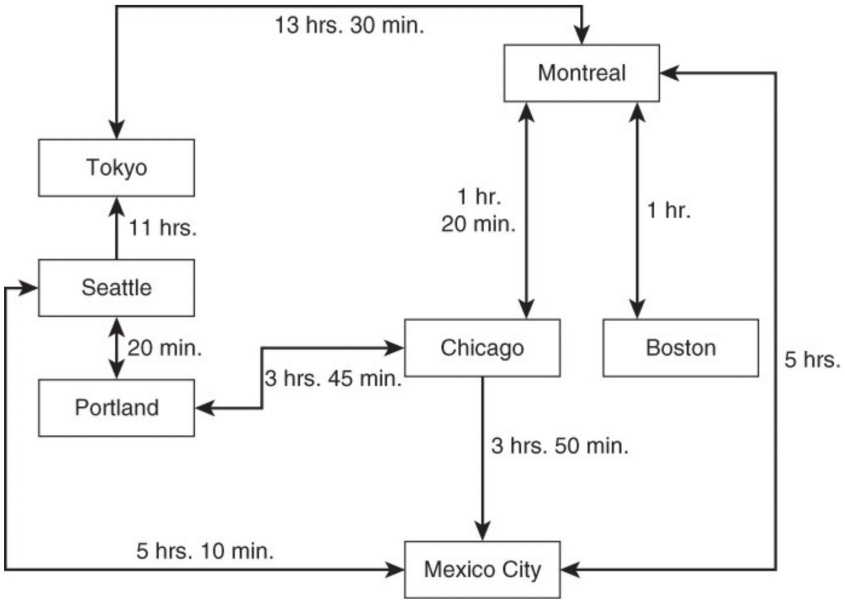


Difference between Relational and Graph Databases



City 1	City 2	Flying Time
Montreal	Boston	1 hr.
Montreal	Chicago	1 hr. 20 min.
Montreal	Tokyo	13 hr. 30 min.
Montreal	Mexico City	5 hr.
Chicago	Mexico City	3 hr. 50 min.
Chicago	Portland	3 hr. 45 min.
Portland	Seattle	20 min.
Seattle	Tokyo	11 hr.
Seattle	Mexico City	5 hr. 10 min.

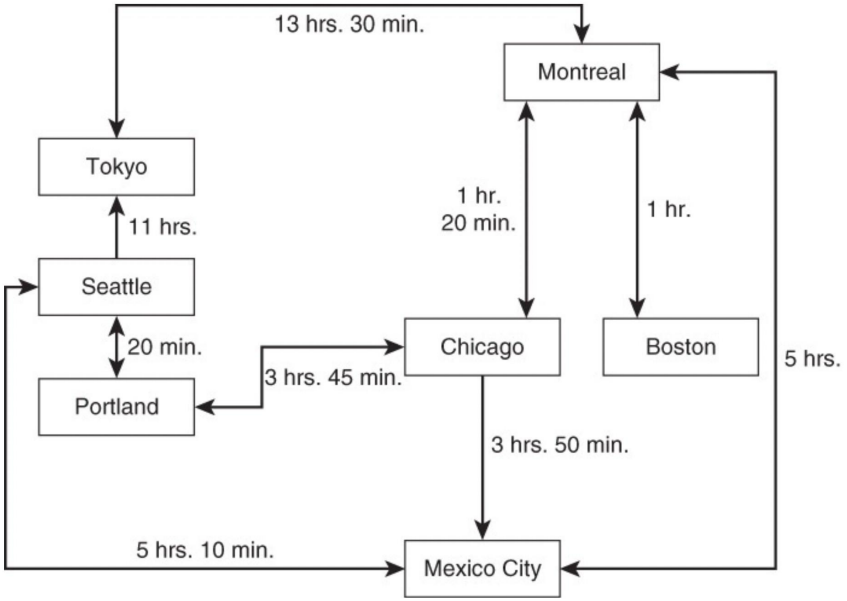
Difference bet ween Relational and Graph Databases



City 1	City 2	Flying Time
Montreal	Boston	1 hr.
Montreal	Chicago	1 hr. 20 min.
Montreal	Tokyo	13 hr. 30 min.
Montreal	Mexico City	5 hr.
Chicago	Mexico City	3 hr. 50 min.
Chicago	Portland	3 hr. 45 min.
Portland	Seattle	20 min.
Seattle	Tokyo	11 hr.
Seattle	Mexico City	5 hr. 10 min.

All possible ways to fly from Montreal to Mexico City?

Difference between Relational and Graph Databases



All possible ways to fly from Montreal to Mexico City?

Can be found looking at paths in the graph

Difference bet ween Relational and Graph Databases

City 1	City 2	Flying Time
Montreal	Boston	1 hr.
Montreal	Chicago	1 hr. 20 min.
Montreal	Tokyo	13 hr. 30 min.
Montreal	Mexico City	5 hr.
Chicago	Mexico City	3 hr. 50 min.
Chicago	Portland	3 hr. 45 min.
Portland	Seattle	20 min.
Seattle	Tokyo	11 hr.
Seattle	Mexico City	5 hr. 10 min.

All possible ways to fly from Montreal to Mexico City?

Difficult to analyse in relational database

Graph databases : Features

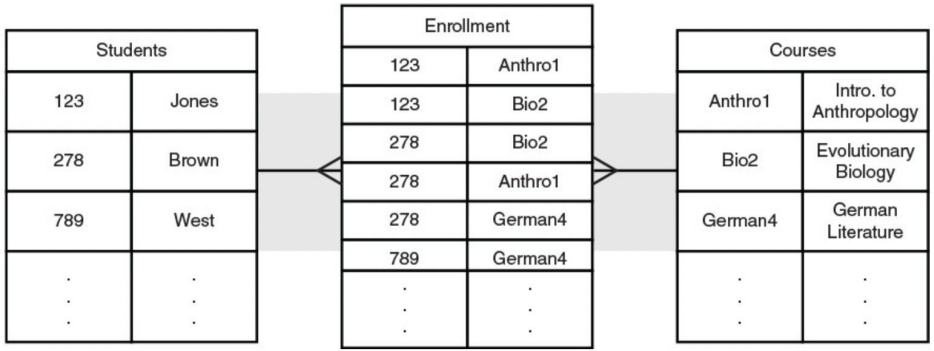
Allows efficient querying when paths in the graph are involved.

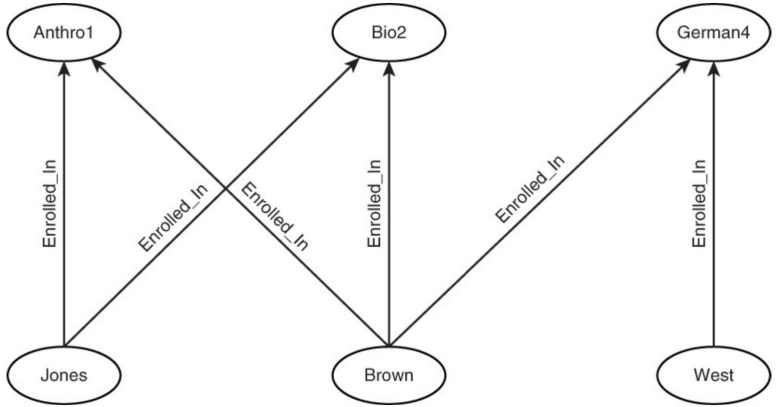
In such cases, minimises the amount of code and efficiently computes answers.

Graph databases : Advantages

Query faster by avoiding Joins:

Joins are replaced by following edges from vertex to vertex





Retrieve student names along with course names that they have enrolled in.

Graph databases : Advantages

Simplified modelling:

Can avoid many to many relations

People

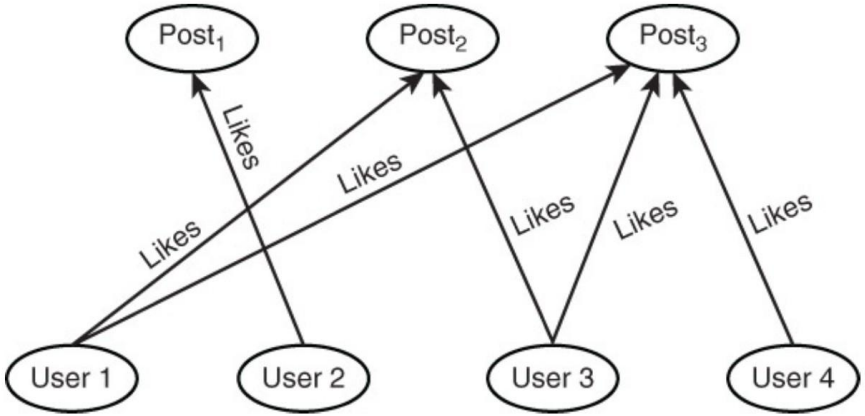
U1	User1
U2	User2
U3	User3
	.
	.
	.
	.
UN	UserN

Likes

U1	P1
U1	P2
U1	P3
U2	P3
U3	PN
U3	P2
UN	P1
UN	P2

Posts

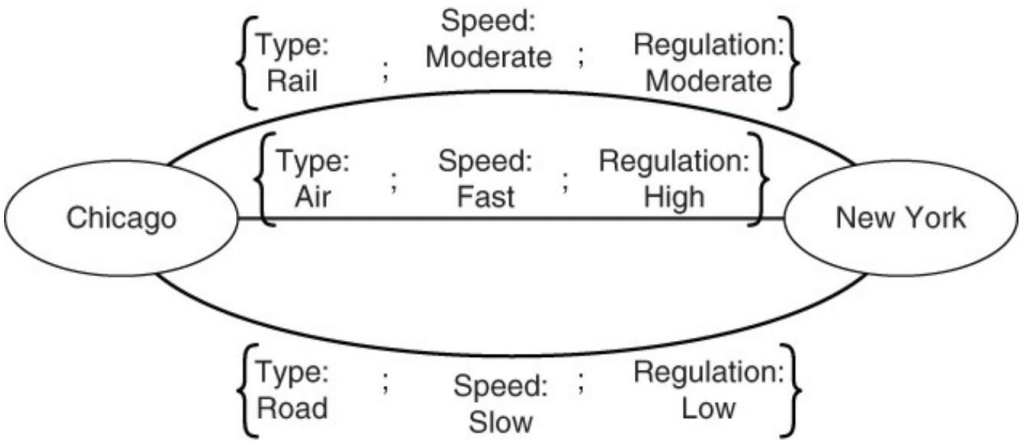
P1	Post1
P2	Post2
P3	Post3
.	.
.	.
.	.
.	.
PN	PostN



Easy to understand in graphs compared to relational model

Graph databases : Advantages

Multiple relationship between entities



Easy to understand in graphs compared to relational model

Designing a Graph Database

Here also we have Entities (nodes)

Identifying relations (edges) between entities

Identifying common properties of edges from a node

Calculating aggregate properties of edges from a node

Calculating aggregate properties of nodes

Example

Social Network

Users can post, reply to a post and follow each other

Example

Social Net work for NoSQL Developers

* Entities: Developers and Posts (You can add something new later)

* Properties (Attributes) of Developers:

Example
Social Network

Entities:

Developers and Posts (You can add something new later)

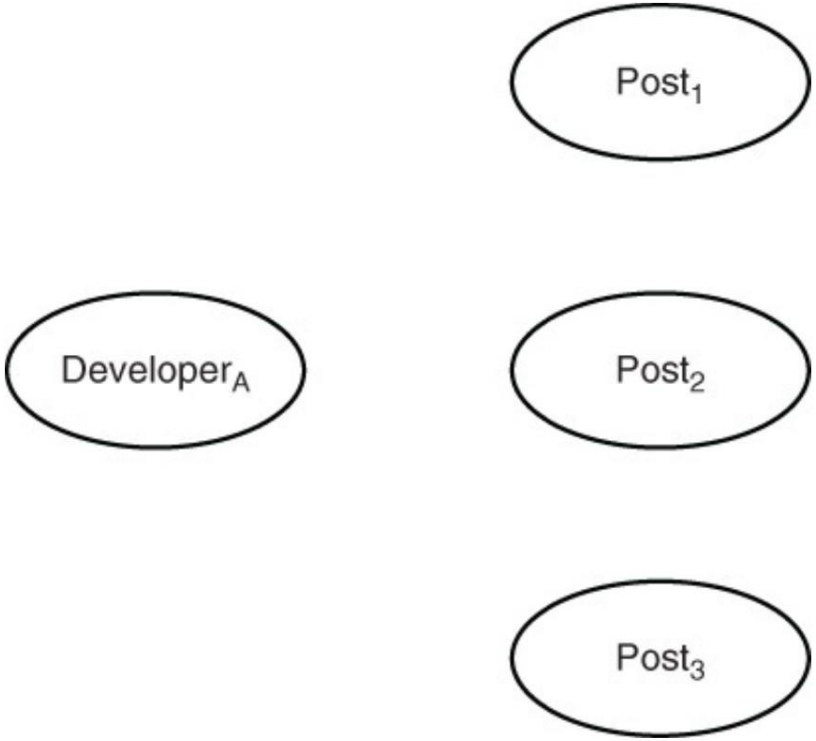
Properties (Attributes) of Users :

Name, Location, NoSQL DB used, Expertise…

Properties (Attributes) of Posts:

Posted Date, Topic, Title, Body Type (Question / Answer)

Example
Social Network



Example
Social Network

Relations:

Developer to Developer (Follows…)

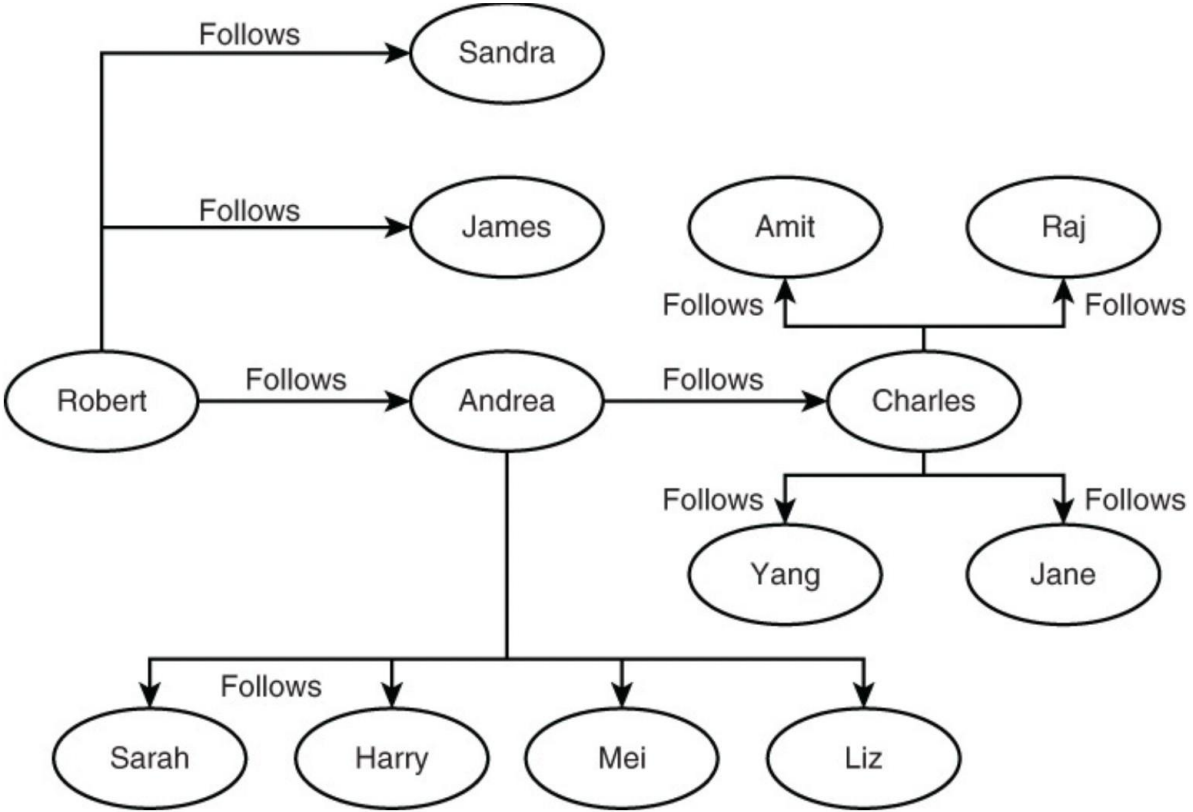
Developer to Post (Creates, Likes..)

Post to Developer (Created by..)

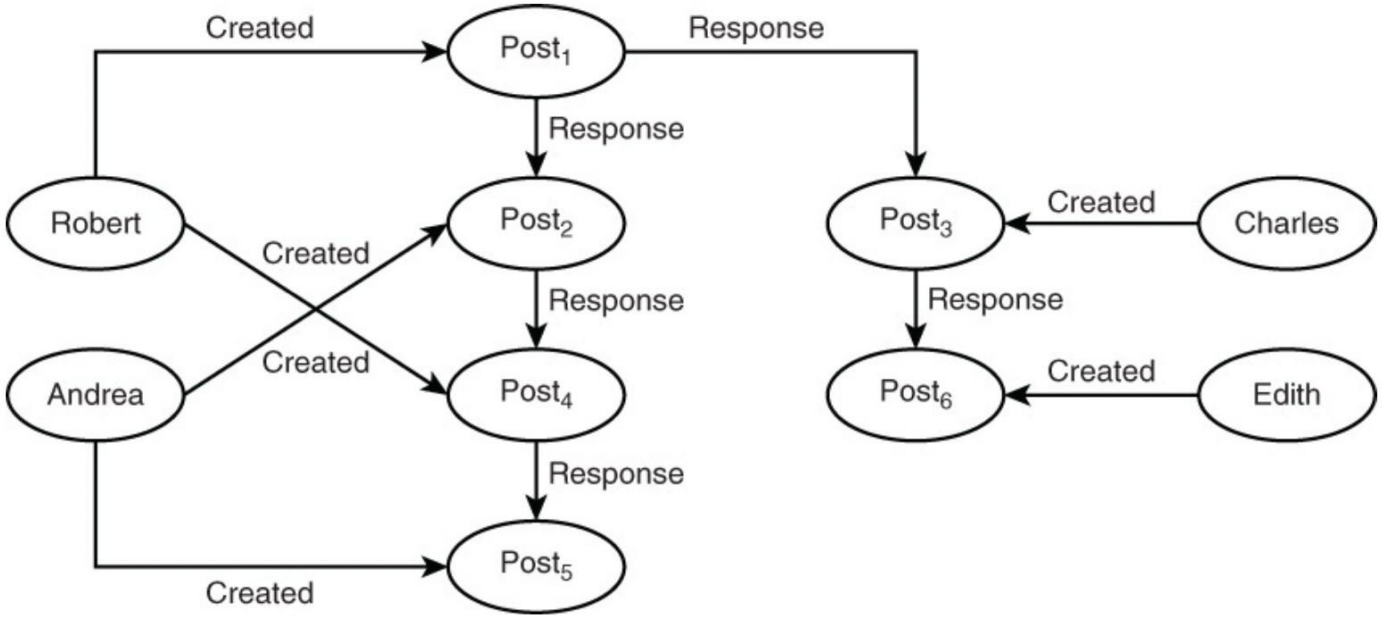
Post to Post (in response to..)

Example

Social Network



Example
Social Network



Example
Social Network

Graph-Centric Query	Domain-Specific Query
How many hops (that is, edges) does it take to get from vertex A to vertex B?	How many follows relations are between Developer A and Developer B?
How many incoming edges are incident to vertex A?	How many developers follow Andrea Wilson?
What is the centrality measure of vertex B?	How well connected is Edith Woolfe in the social network?
Is vertex C a bottleneck; that is, if vertex C is removed, which parts of the graph become disconnected?	If a developer left the social network, would there be disconnected groups of developers?

Example
Social Network

* Creating Entities:

  CREATE (alice: User {name: ‘Alice Wonderland’})

  CREATE (charlie: User {name: ‘Charlie Chaplin’})

* Creating Relationship:

  CREATE (alice)-[follows]->(charlie)

Example
Social Network

* Retrieving queries:

  MATCH (d: developer) RETURN d

  SELECT * FROM developer (SQL equivalent)

Example
Social Network

Retrieving queries:

MATCH (d: developer) RETURN d

SELECT * FROM developer (SQL equivalent)

WHERE, ORDER BY, LIMIT, COUNT, SUM, AVG…

Reference:

NoSQL for mere mortals by Dan sullivan
Chapter 14