

Section: Graph Databases

Introduction: New Data Models*

* Relational database model is just one of many

Data Model Notions

formal commercial/trade

- Network --- graph: Neo4j, RDF
 - Hierarchical --- Document: JSON, XML
 - Relational
 - *Object-oriented*

1: Introduction

Data Management & Engineering

Data Model Notions

formal commercial/trade BTW

- Network --- graph: Neo4j, RDF CODASYL
 - Hierarchical – JSON, XML IBM IMS
 - Relational MongoDB Postgres
 - Object oriented all died

1: Introduction

Data Management & Engineering

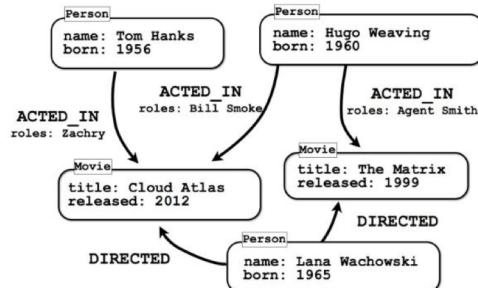
Graph DBs:

Predate Relational Databases: CODASYL standard.

Present:

- Property Graph Model
 - *The hot new thing*
 - *Primitive, no standards*
 - Edge (RDF) Graph Model
 - *Still current. Integrates symbolic AI reasoning*
 - *[original] basis of knowledge graphs*
 - *Stack of standards, all with good formal semantics*
 - *Syntax is challenging for average developers, tools are worse*

A Property Graph



Both vertices and edges, labeled with

- Any list of key value pairs,
- E.g. { (key1, value1), (key2, value2) ... (keyn, valuen)}
- With rare exception – no schema required.

What is a data model?

Definition 1:

Data Model: A data model is a **conceptual representation** of the **data structures** that are required by a **database to support an application**.

Main memory data structures:

- binary tree
- hash-table

Data Model

Definition 2:[text page 17]

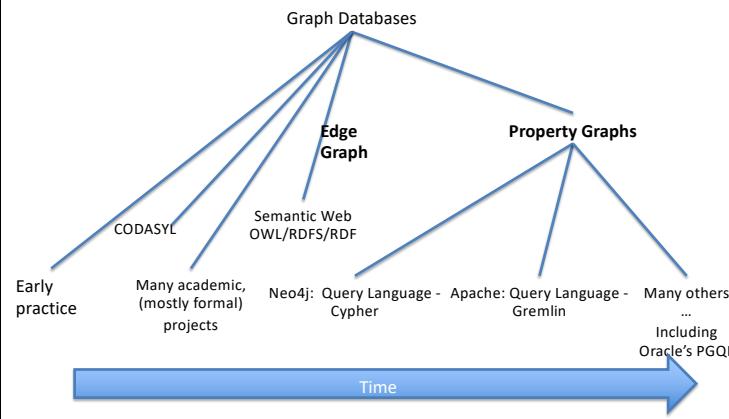
"A *data model* is a notation for describing data or information... generally consisting of three parts"

Data Model (2)

"Generally consists of three parts:

- Structure of the data...
 - Data definition language (DDL)
- Operations on the data ...
 - Data manipulation language (DML)
- Constraints on the data ...
 - (also DDL)

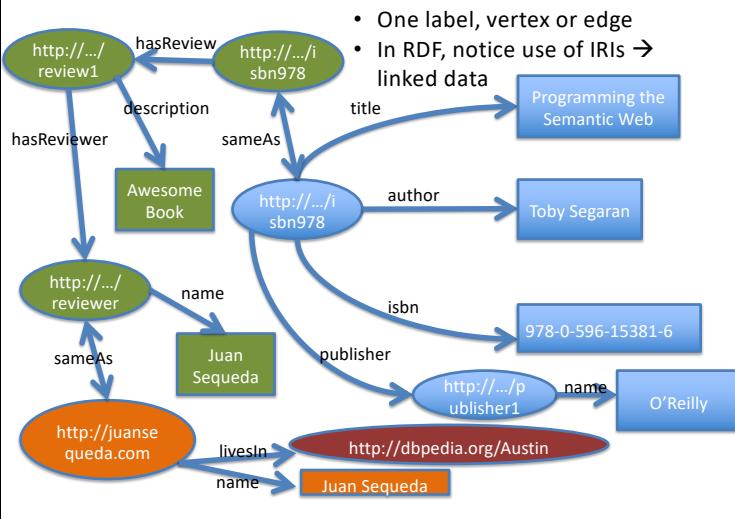
History of Graph (navigational) Databases



About the Property Graph Model

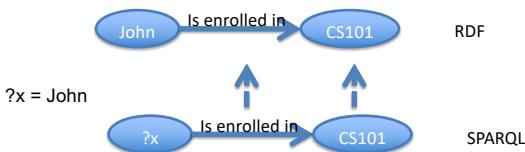
- >20 Commercial Implementations
- Almost none have a schema language.
 - Exception, Tiger graph....
 - Neo4j, Cypher has introduced schema; required when using their SPARK-based implementation
- Without Schema,
 - Can we have operators on graphs?
 - Can we have constraints on graphs?

An RDF (Edge) Graph:



SPARQL

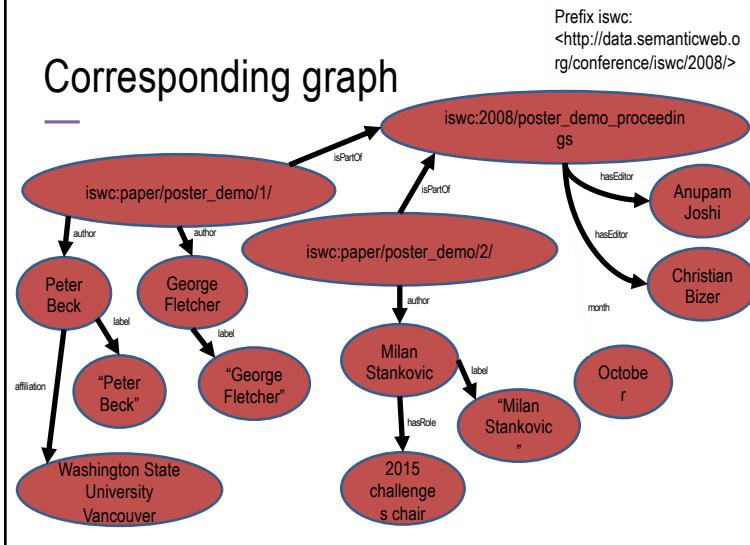
- SPARQL Protocol and RDF Query Language
- Graph Pattern Matching
- Describe sub-graphs of the RDF graph and match it with RDF
- Semantics equivalent to nr-datalog w/ neg



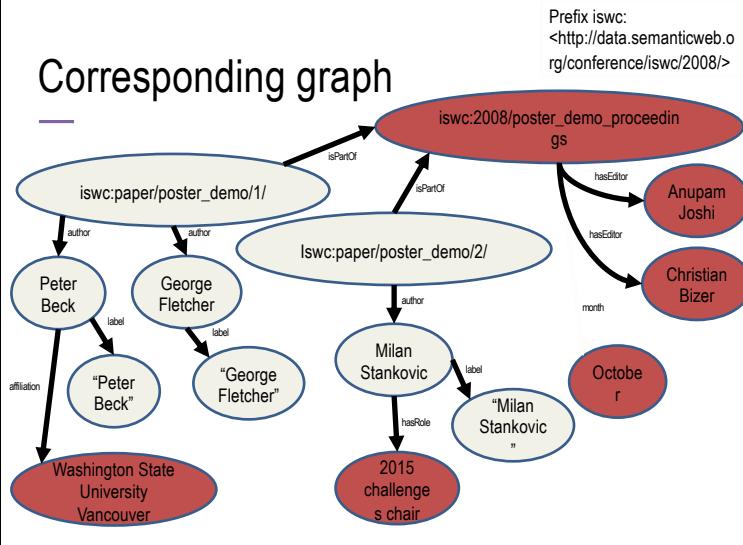
Example SPARQL Query

```
select ?paper ?p ?n
where {
  ?paper swc:isPartOf <http://data.semanticweb.org/conference/iswc/2008/poster_demo_proceedings> .
  ?paper swc:author ?p .
  ?p rdfs:label ?n .
}
```

Corresponding graph

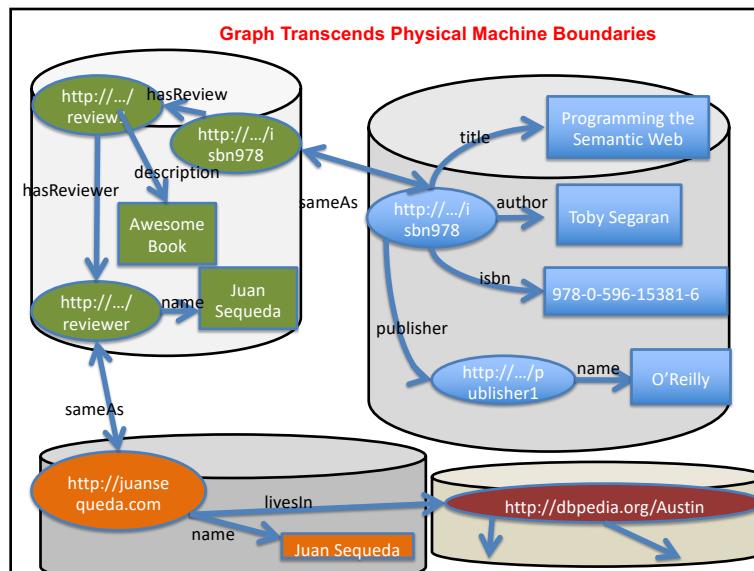


Corresponding graph



Linked Data

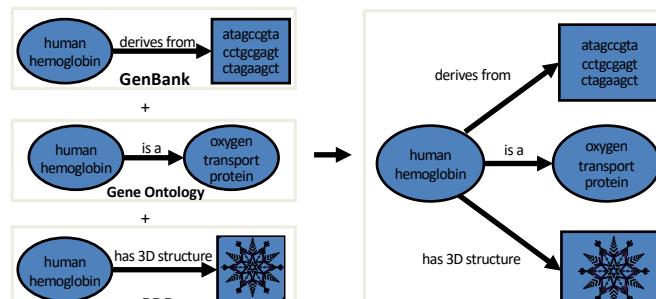
- Data on the Web as graphs and they are interlinked
 - URI's form foreign keys
 - references a table in a different machine!
- Imagine querying all this data together!



Linked Data

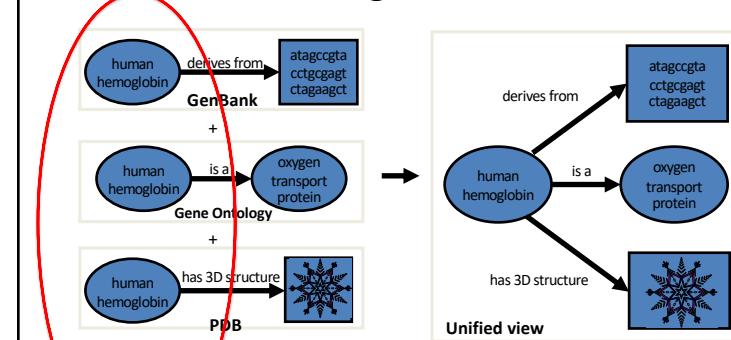
- RDF Graph data model is flexible
 - Can create edges to other nodes from a different graph (database)
 - Merge nodes

Triples are Very Handy for Data Integration



[thanks Neuman and Quan]

Triples are Very Handy for Data Integration



[thanks Neuman and Quan]

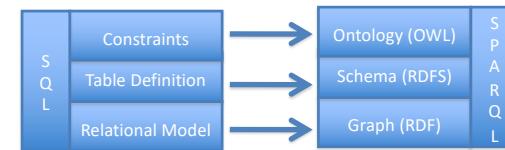
RDF Graphs Knowledge Graphs

Semantic Web
Linked Data

- Distributed Web-scale Data Integration
- Web of Linked Documents → Web of Linked Data



Two Layer Cake Methodology



Next Rebirth Knowledge Graphs

- formerly known as *ontology*
- people have never heard of this, and it can be very confusing.
- people are very scared of cancer
- ontology : oncology

Watershed event

- Google bought Metaweb, a semantic web company
- They came out with...
do this:



About 910,000,000 results (1.37 seconds)

The University of Texas at Austin
<https://www.utexas.edu> ▾
 Like the state it calls home, The University of Texas at Austin is a bold, ambitious leader, home to more than 51,000 students and 3,000 teaching faculty.

Results from utexas.edu

Texas Admissions **Graduate School**
 Admission - Apply - MyStatus - Contact - Explore - Visit - Afford
 Achieving a higher degree of excellence. The Graduate ...

People also ask

What GPA do you need to get into UT? ▾
 Which University of Texas is the best? ▾
 Is University of Texas a good school? ▾
 Is University of Texas at Austin an Ivy League school? ▾

Feedback

UT System Administration Home Page | University of Texas System
<https://www.utsystem.edu> ▾
 The University of Texas System (UT System) is a state university system with a global impact — focusing on higher education, research and health care at 14 ...

University of Texas at Austin - Wikipedia
https://en.wikipedia.org/wiki/University_of_Texas_at_Austin ▾
 The University of Texas at Austin (UT Austin) is a public research university in Austin, Texas. It was founded in 1883 and is the flagship institution of the ...
 Students: 51,525 (Fall 2017) Former names: The University of Texas; (1881...
 Campus: Urban, 431 acres (1.74 km²). Location: **Austin, Texas**, United States
 History · Academics · Research · Student life

Address: Austin, TX 78712 **Phone:** (512) 471-3434 **Undergraduate enrollment:** 39,057 (2015–16)

This part of the result is called the knowledge panel

It comes out of their *knowledge graph*

It contains structured data (not links to documents)

The University of Texas at Austin
[Website](#) [Directions](#) [Save](#)
 Public university in Austin, Texas

The University of Texas at Austin is a public research university in Austin, Texas. It was founded in 1883 and is the flagship institution of the University of Texas System.
[Wikipedia](#)

Avg cost after aid	Graduation rate	Acceptance rate
\$18K	81%	39%

Graduation rate is for first-time, full-time undergraduate ... more ▾
 Source: US Dept of Education - Learn more

Address: Austin, TX 78712 **Phone:** (512) 471-3434 **Undergraduate enrollment:** 39,057 (2015–16)

University of Texas at Austin - Wikipedia
https://en.wikipedia.org/wiki/University_of_Texas_at_Austin ▾
 The University of Texas at Austin (UT Austin) is a public research university in Austin, Texas. It was founded in 1883 and is the flagship institution of the ...
 Students: 51,525 (Fall 2017) Former names: The University of Texas; (1881...
 Campus: Urban, 431 acres (1.74 km²). Location: **Austin, Texas**, United States
 History · Academics · Research · Student life

Also structured data

named data types (tags)

also "reserved" tags

```
<div>
<h1>Avatar</h1>
<span>Director: James Cameron (born August 16, 1954)</span>
<span>Science fiction</span>
<a href="#">Trailer</a>
</div>
```

Where to these tags come from?

- One source, embedded in anyone's web page, **Breed & Company**

Breed & Company
[Website](#) [Directions](#) [Save](#)
 4.7 ★★★★☆ 187 Google reviews
 \$\$\$ · Hardware store in Austin, Texas

Old-fashioned hardware store with a wide range of items from home repair necessities to decorations.
Address: 718 W 29th St, Austin, TX 78705
Hours: Open - Closes 7PM *
Phone: (512) 474-6679

- Particular tags come from <http://schema.org>

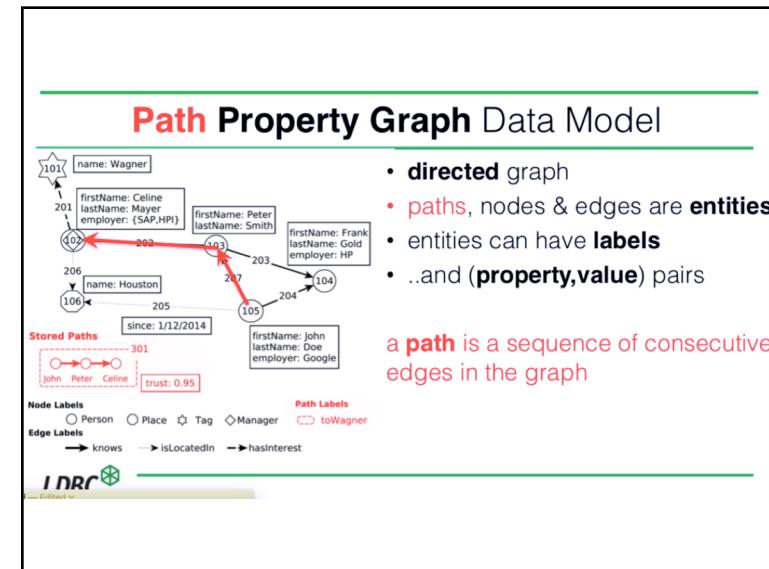
The screenshot shows the schema.org Person schema page. At the top, there's a navigation bar with links for Home, Schemas, and Documentation. Below the navigation, the title "Person" is displayed under "Thing > Person". A brief description follows: "A person (alive, dead, undead, or fictional)." There's a "[more...]" link. Below this is a table titled "Properties from Person" with columns for "Property", "Expected Type", and "Description". The properties listed include:

Property	Expected Type	Description
<code>additionalName</code>	Text	An additional name for a Person, can be used for a middle name.
<code>address</code>	PostalAddress or Text	Physical address of the item.
<code>affiliation</code>	Organization	An organization that this person is affiliated with. For example, a school/university, a club, or a team.
<code>alumniOf</code>	EducationalOrganization or Organization	An organization that the person is an alumni of. Inverse property: <code>alumni</code> .
<code>award</code>	Text	An award won by or for this item. Supersedes <code>awards</code> .
<code>birthDate</code>	Date	Date of birth.
<code>birthPlace</code>	Place	The place where the person was born.

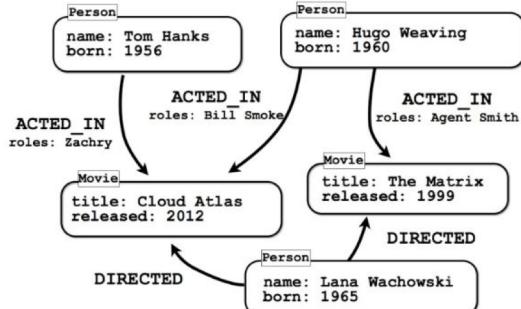
But Property Graphs are the New Thing

Some Things to Know

- No property standards, >20 commercial systems → many languages
- Two dominate
 - Gremlin (Apache Open Source)
 - Cypher
- Standards are coming!



A Property Graph



Both vertices and edges, labeled with

- Any list of key value pairs,
- E.g. { (key1, value1), (key2, value2) ... (keyn, valuen)}
- With for one vendor exception – No schema required.

Cypher

```

MATCH (nicole:Actor {name: 'Nicole Kidman'})-[:ACTED_IN]->(movie:Movie)
WHERE movie.year < $yearParameter
RETURN movie
  
```

Gremlin

```

g.V().match(
  as("a").has("name","gremlin"),
  as("a").out("created").as("b"),
  as("b").in("created").as("c"),
  as("c").in("manages").as("d"),
  where("a",neq("c"))).
select("d").
groupCount().by("name")
  
```

Cypher

```

MATCH (nicole:Actor {name: 'Nicole Kidman'})-[:ACTED_IN]->(movie:Movie)
WHERE movie.year < $yearParameter
RETURN movie
  
```

Gremlin

```

g.V().match(
  as("a").has("name","gremlin"),
  as("a").out("created").as("b"),
  as("b").in("created").as("c"),
  as("c").in("manages").as("d"),
  where("a",neq("c"))).
select("d").
groupCount().by("name")
  
```

Declarative: Associative Look-up

Navigational: User keeps track of graph edges (like they are pointers.... That's what we do in C programming, even some in C++)

What About Semantics?

- **Formal Semantics** [\[wikipedia\]](#)

Variations

Some variations of formal semantics include the following:

- Action semantics is an approach that tries to modularize denotational and predefining three semantic entities (actions, data and yielders) to
- Algebraic semantics is a form of **axiomatic semantics** based on **algebraic** structures
- Attribute grammars define systems that systematically compute "meta" grammars can be understood as a denotational semantics where the from formal semantics, attribute grammars have also been used for context-sensitive conditions;
- Categorical (or "functorial") semantics uses **category theory** as the conceptual framework for semantics
- Concurrency semantics is a catch-all term for any formal semantics that included the Actor model and process calculi;
- Game semantics uses a metaphor inspired by game theory.
- Predicate transformer semantics, developed by Edsger W. Dijkstra, defines a function that takes a predicate and returns another predicate, corresponding to the precondition needed to establish it.

June 2018 The beginning of consolidation

- G-CORE: A Core for Future Graph Query Languages
 - published at SIGMOD by a committee of both academic researchers and a few industry representatives
 - <https://arxiv.org/pdf/1712.01550.pdf>
 - Formal semantics

September GQL an official ISO project.

[GQL Is Now a Global Standards Project alongside SQL - Neo4j](#)
<https://neo4j.com/blog/gql-standard-query-language-property-graphs/> ▾
 Sep 16, 2019 - Learn how Graph Query Language (GQL) has just been approved to ... Now, it's

Graph Query Language (GQL) is a new language being developed and maintained by the same international working group that also maintains the SQL standard.

Now, it's official: Earlier this week, the ballot closed and the proposal passed, with seven countries putting experts forward to work on the four-year project.

GQL draws heavily from existing languages, but the main inspiration has so far been Cypher (now with [over 10 implementations](#), including six commercial products),

September: GQL an official ISO project.

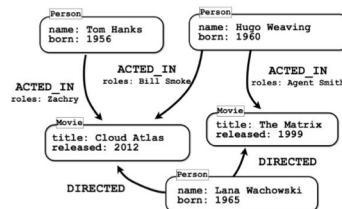
[GQL Is Now a Global Standards Project alongside SQL - Neo4j](#)
<https://neo4j.com/blog/gql-standard-query-language-property-graphs/> ▾
 Sep 16, 2019 - Learn how Graph Query Language (GQL) has just been approved to ... Now, it's

Graph Query Language (GQL) is a new language being developed and maintained by the same international working group that also maintains the SQL standard.

Now, it's official: Earlier this week, the ballot closed and the proposal passed, with seven countries putting experts forward to work on the four-year project.

GQL draws heavily from existing languages, but the main inspiration has so far been Cypher (now with [over 10 implementations](#), including six commercial products),

A Property Graph

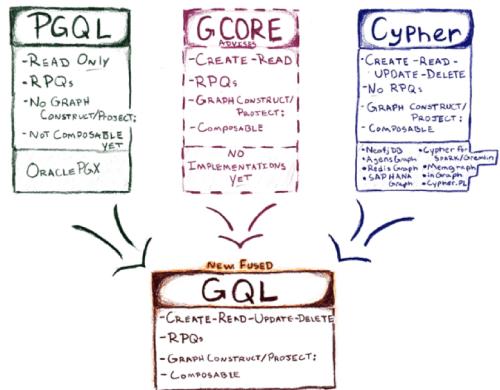


Commonly Multi-Modal: Put “anything you want” on vertices and edges

- JSON
- XML
- Implement Semantic Web’s RDF
 - Borrow from RDF Schema
 - Borrow from OWL
 - there is at least one implementation of OWL layered on property graphs

From a Neo4j sponsored blog

<https://gql.today/comparing-cypher-pgql-and-g-core/>



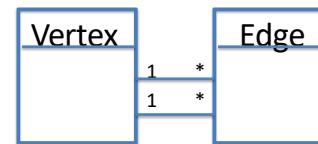
What About Graphs and The Relational Model?

What is a graph?

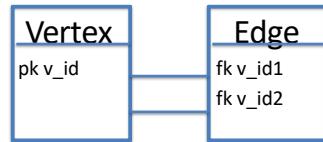
A **graph** ... is a pair $G = (V, E)$, where V is a set whose elements are called *vertices* (singular: *vertex*), and E is a set of two-sets (sets with two distinct elements) of vertices, whose elements are called *edges* (sometimes *links* or *lines*).

- The vertices x and y of an edge $\{x, y\}$ are called the *endpoints* of the edge.

Logical Model



Physical Model



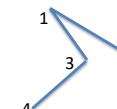
Database Instance

The Database Instance diagram shows the physical model implemented in a database. It includes two tables: 'Vertex' and 'Edge'. The 'Vertex' table has four rows with 'v_id' values 1, 2, 3, and 4. The 'Edge' table has four rows with 'v_id1' and 'v_id2' values: (1, 2), (1, 3), (3, 4), and (4, 2). To the right, a directed graph is shown with nodes 1, 2, 3, and 4. Node 1 is at the top, node 2 is to its right, node 3 is below node 1, and node 4 is at the bottom. Directed edges exist from node 1 to node 2, node 1 to node 3, node 3 to node 4, and node 4 to node 2.

Actually A Small Problem When Taking This All the Way to SQL

```
** All paths of length 2
Select E1.v_id1, E1.v_id2, E2.v_id2
From Edge E1, Edge E2
Where E1.v_id2 = E2.v_id1;
```

The tables for the SQL query are identical to those in the Database Instance. The 'Vertex' table has four rows with 'v_id' values 1, 2, 3, and 4. The 'Edge' table has four rows with 'v_id1' and 'v_id2' values: (1, 2), (1, 3), (3, 4), and (4, 2).



As Illustrated, Edges are Directed

```
** All paths of length 2
Select E1.v_id1, E1.v_id2, E2.v_id2
From Edge E1, Edge E2
Where E1.v_id2 = E2.v_id1;
```

The tables for the SQL query are identical to those in the Database Instance. The 'Vertex' table has four rows with 'v_id' values 1, 2, 3, and 4. The 'Edge' table has four rows with 'v_id1' and 'v_id2' values: (1, 2), (1, 3), (3, 4), and (4, 2).

