Graph database

Nhuan To

Department of Electrical and Computer Engineering

University of Alberta

Email: Nhuan@ualberta.ca

Outline

1. What is a graph database?
2. Labeled property graph (LPG)
3. Resource Description Framework (RDF) graph
4. Question answering system on RDF graph
5. Q&A

1. What is a graph database?

Question 1: What is a **database**?

Answer: A **database** is an **organized collection of data**, generally **stored and accessed electronically** from a **computer system**. (Wikipedia)

Question 2: What is a **graph**?

Answer: in graph theory, a graph is an order pair $G = (V, E)$ comprising:

- V a set of **vertices;**
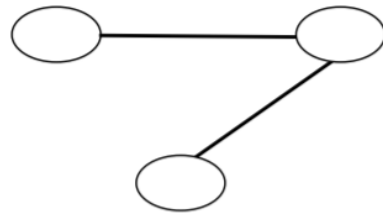- **E** a set of **edges** (pairs of vertices).

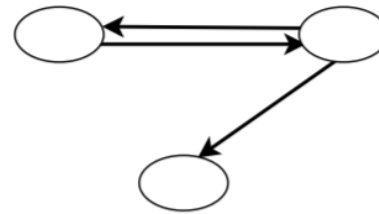Fig 1. A undirected graph
with three nodes and two edges

Fig 2. A directed graph
with three nodes and three edges

A **graph database** is a **database** that uses **graph** structures to represent and store data. (Wikipedia)

2. Labeled property graph database (LPG)

- **Nodes** are often used to represent **entities**. *A node has a set of **property:value** pairs and may have **label(s).***

- **Edges** are often used to represent **relationships** between **entities**. A **relationship** must have exactly one relationship **type** and may have a set of **property:value** pairs describes the relationship.
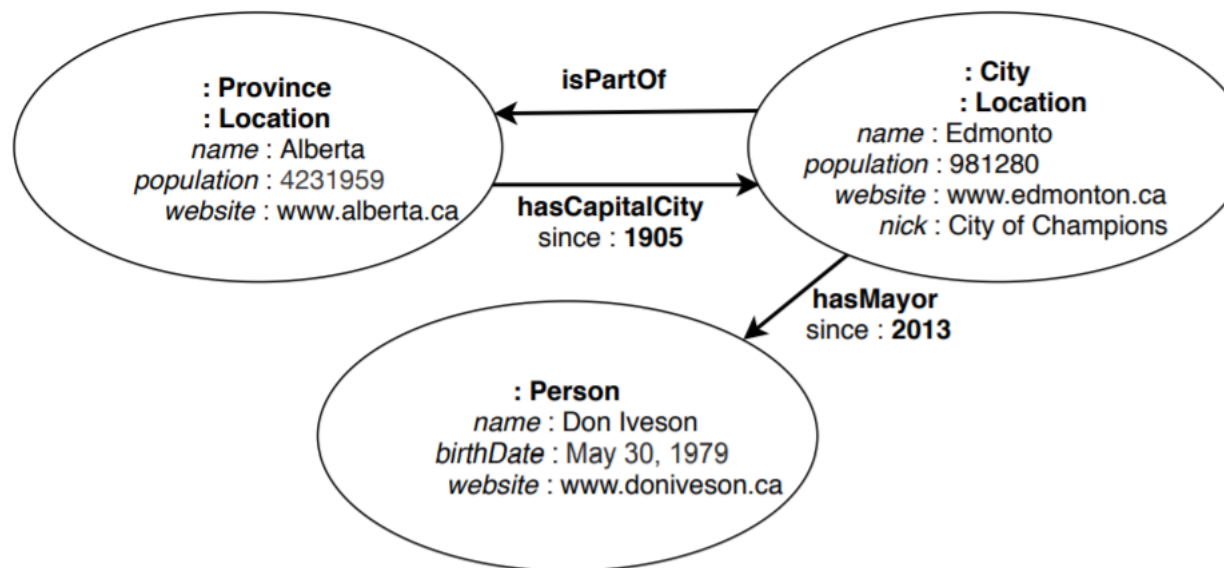

Fig 3. A labeled property graph with three nodes and three edges

2. Labeled property graph database (continue)

- **Query languages:** Cypher, GQL, PGQL, Gremlin

- A example of Cypher query in neo4j:

Match (city:City)-[:isPartOf]->(:Province{name:Alberta})
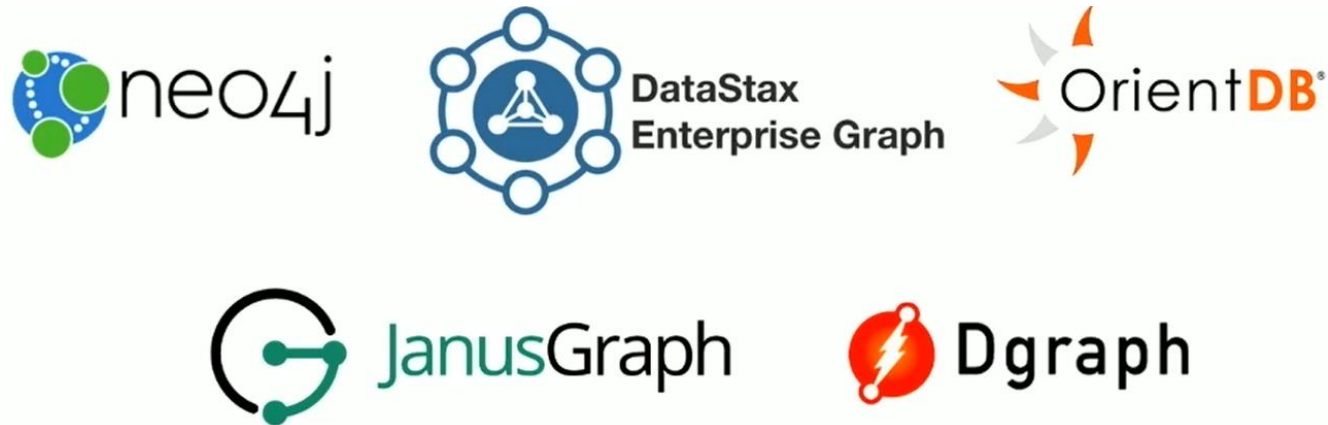Return city

- **LPG datastores:**

Fig 4. Property graph datastores
(source: What are Graph Databases and Why should I care? - Dave Bechberger)

3. Resource Description Framework (RDF) graph

- **RDF** was originally created in early 1999 by World Wide Web Consortium (**W3C**) as a **standard** for **encoding metadata** about **Web resources**.

- A resource is described by a list of **statements**, and each **statement** takes the form of **Subject-Predicate-Object**.

  - The **subject** is the **name** of the resource.
  - The **predicate** *denotes* one of traits or aspects of the resource. It is also called **property.**
  - The **object** is the **property's value**.
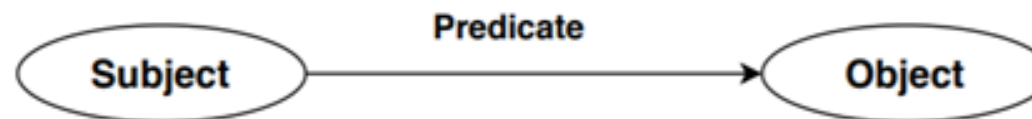

Fig 4. Graph structure of a RDF statement

3. RDF graph   (continue..)

- The **name** of a **resource** must be **global** and should be **identified** by **Uniform Resource Identifier (URI).**
- The name of **predicate** must also be **global** and should be **identified** by **URI** as well.

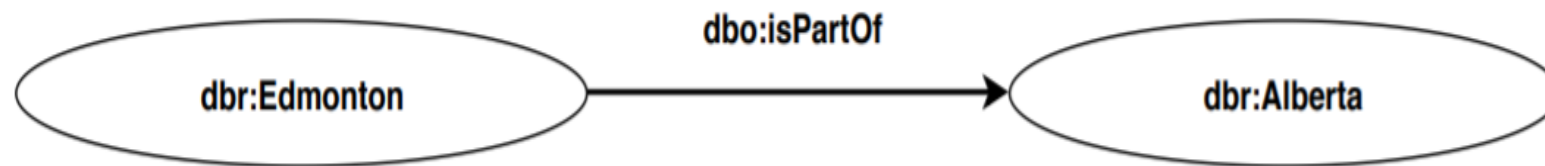Fig 5. Graph structure of a RDF statement in **DBpedia**

Fig 6. Graph structure of a RDF statement in **Dbpedia** using **XML's QName**

- If a resource is not identified by a URI, the node repressing the resource in RDF graph is called a **blank node** (Bnode)**.**
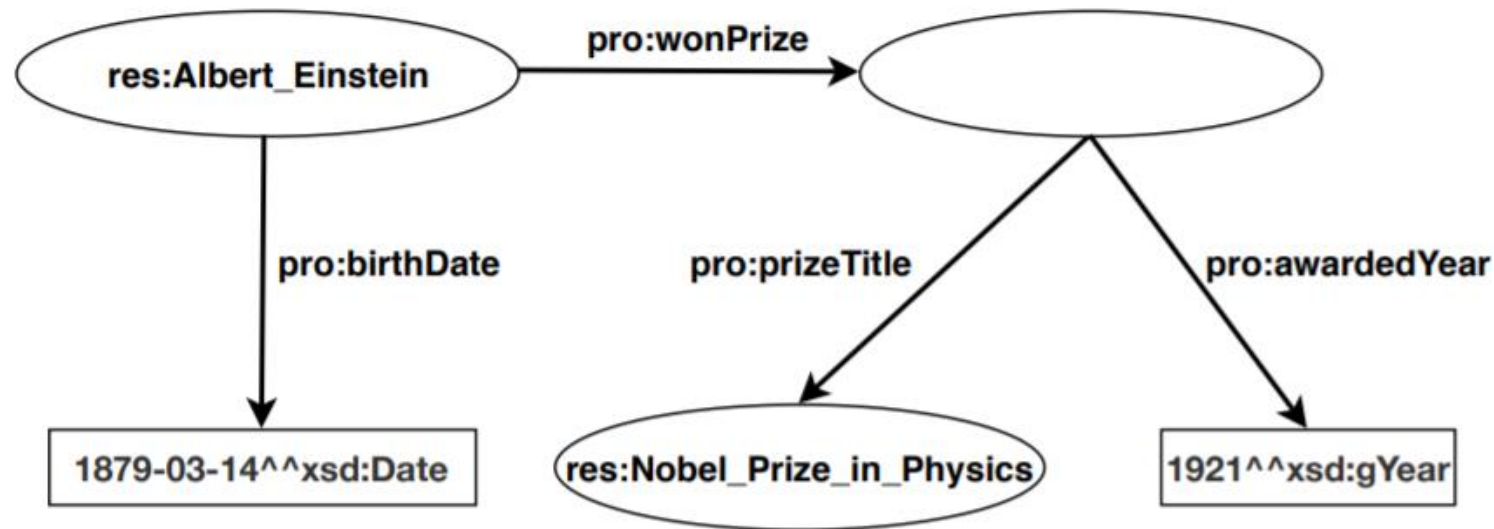

Fig 7. RDF Graph containing a blank node

3.1. RDF schema

- RDF refers a set of URIs to an RDF vocabulary.

- RDFS is recommendation from W3C and it is a language that one can use to create a vocabulary for describing classes, subclasses and properties of RDF resources. It provides

  - rdfs:label

  - rdfs:comment

  - rdfs:domain

  - rdfs:range

  - ...

3.2. Ontology and Web Ontology Language (OWL)

- An **ontology** formally defines a common set of terms that are used to describe and represent a domain.
    - Ontology is **domain-specific**
    - Ontology contains **terms (classes, properties)** and the relationships among these terms.
- **OWL** is a **language** to use when **creating ontologies.**
    - Has same purpose as RDF Schema
    - OWL = RDF Schema + new constructs for better expressiveness
- Some predefined ontologies:
    - **Schema.org**: a well accepted ontology for common things in life founded by Google, Microsoft, Yahoo and Yandex.
    - **FOAF:** a machine-readable ontology describing persons, their activities and their relations to other people and objects
    - …

3.3. RDF Serialization

Provide syntax for syntax for creating and reading concrete RDF data.

- **RDF/XML**: represents an RDF graph as an XML document.

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:dbo="http://dbpedia.org/ontology/"
    <rdf:Description rdf:about="http://dbpedia.org/resource/Edmonton">
        <rdf:type rdf:resource="http://dbpedia.org/ontology/Location" />
        <rdf:type rdf:resource="http://dbpedia.org/ontology/City" />
        <dbo:areaCode>780,587, 825</dbo:areaCode>
```

- **Turtle**: a compact, human-friendly format.

```
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbr:    <http://dbpedia.org/resource/> .
@prefix dbo:    <http://dbpedia.org/ontology/> .
dbr:Edmonton    rdf:type        dbo:City ;
                dbo:isPartOf    dbr:Alberta ;
```

- Other RDF serializations: **Notation 3 (N3), RDF/JSON, JSON-LD, N-Triples, N-Quads**

3.4. RDF query language: SPARQL

- Is a **SQL-like** query language for RDF graph data provides 6 main query types:
  - SELECT which returns tabular results.
  - CONSTRUCT creates a new RDF graph.
  - ASK which returns 'yes' if the query has a solution, otherwise, 'no'.
  - DESCRIBE which return RDF graph data about a resource.
  - INSERT which inserts triples into a graph.
  - DELETE which deletes triples from a graph.
- **Example:**

  SELECT DISTINCT ?city

  WHERE {?city rdf:type dbo:City. ?city dbo:isPartOf dbr:Alberta.}

3. RDF graph
3.5. RDF datastores

- **RDF datastores:**

Fig 7. RDF datastores
(source: What are Graph Databases and Why should I care? - Dave Bechberger)

Fig 8. Multi-model datastores
(source: What are Graph Databases and Why should I care? - Dave Bechberger)

4. Question answering system on RDF graph

- Take a natural language question from user.
- Translate the question into a SPARQL query.
- Send the query to SPARQL endpoint (DBpedia/Wikidata)
- Get the query result and represent it to the user.