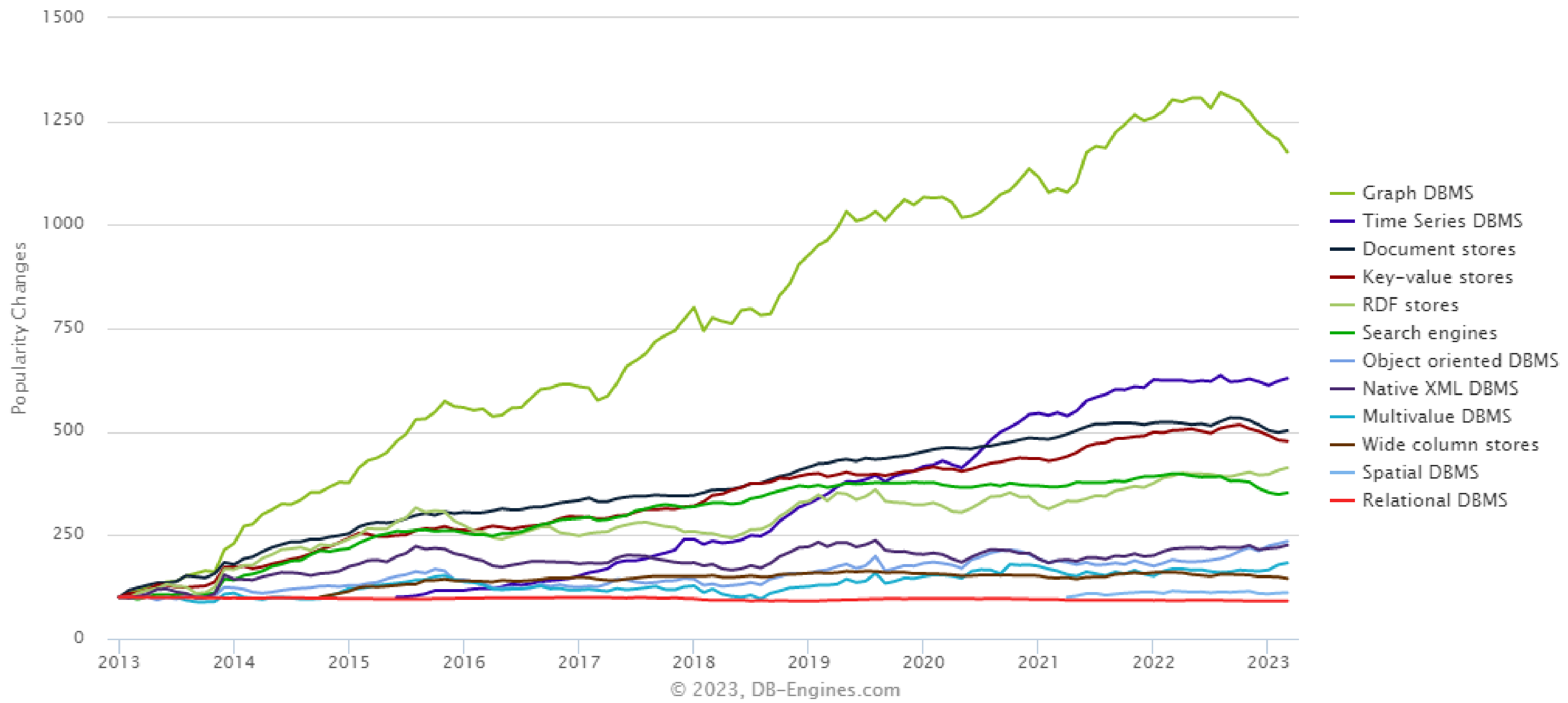
Graph DB

Venkatesh Vinayakarao
venkateshv@cmi.ac.in
http://vvtesh.co.in

Chennai Mathematical Institute

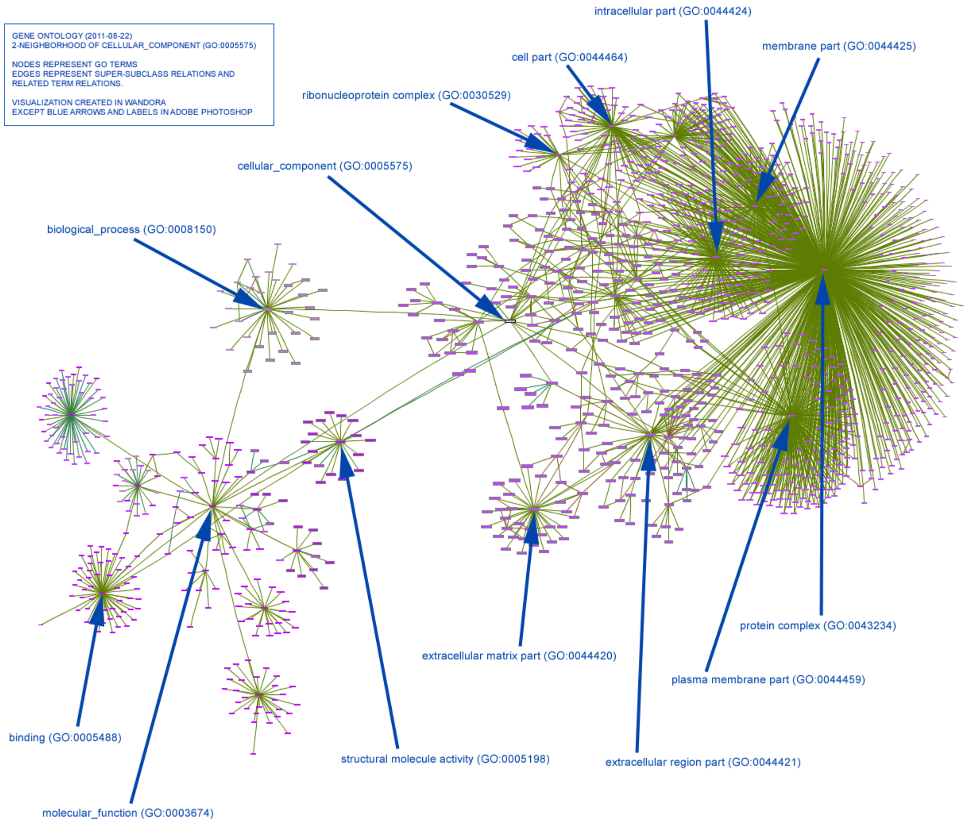
We live in a connected world! . – **Neo4j.**
(Neo4j)-[:LOVES]-(Developers)

Change in Popularity



Source: https://db-engines.com/en/ranking_categories

Gene Ontology Model

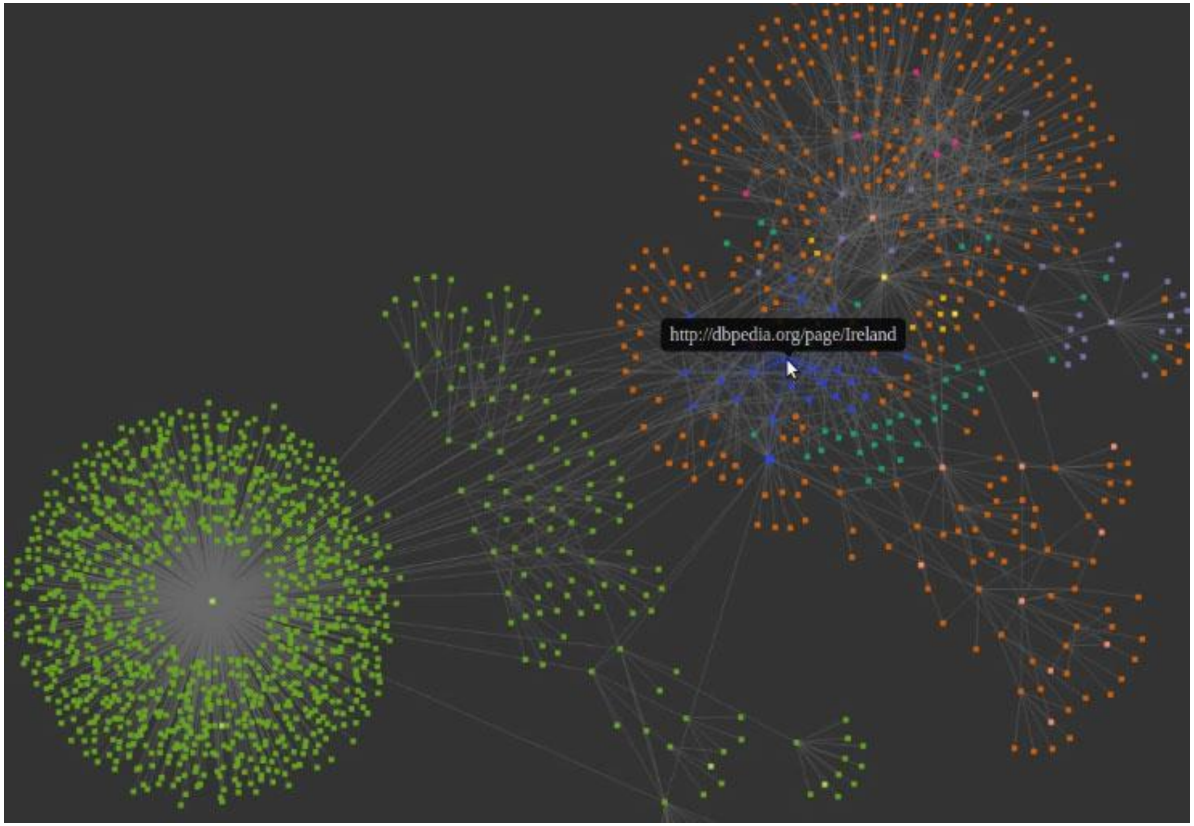

GENE ONTOLOGY (2011-08-22)
2-NEIGHBORHOOD OF CELLULAR_COMPONENT (GO:0005575)
NODES REPRESENT GO TERMS
EDGES REPRESENT SUPER-SUBCLASS RELATIONS AND RELATED TERM RELATIONS.
VISUALIZATION CREATED IN WANDORA EXCEPT BLUE ARROWS AND LABELS IN ADOBE PHOTOSHOP
intracellular part (GO:0044424)
membrane part (GO:0044425)
cell part (GO:0044464)
ribonucleoprotein complex (GO:0030529)
cellular_component (GO:0005575)
biological_process (GO:0008150)
protein complex (GO:0043234)
extracellular matrix part (GO:0044420)
plasma membrane part (GO:0044459)
binding (GO:0005488)
structural molecule activity (GO:0005198)
extracellular region part (GO:0044421)
molecular_function (GO:0003674)

Number of topics: 177301

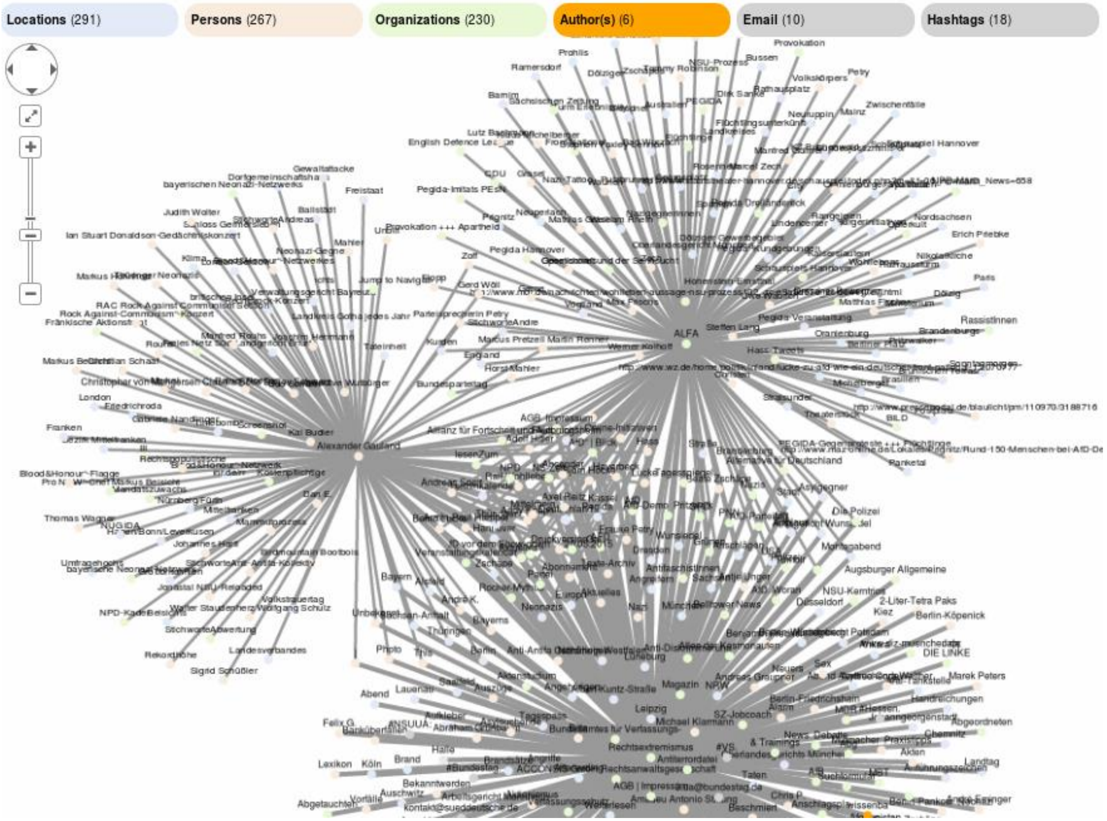
Number of associations: 280198

Source: http://wandora.org/wiki/Topic_map_conversion_of_Gene_Ontology

Knowledge Graphs



Source: https://www.ibm.com/blogs/research/2016/01/from-knowledge-graphs-to-cognitive-computing/



Source: https://www.opensemanticsearch.org/doc/search/graph

Graph Database

Relationships between data is equally as important as the data itself.

Storage and Processing

- Native Graph Storage & Processing
  - Optimized for graph related use cases such as traversals
  - Use index-free adjacency
    - Each node directly references its adjacent (neighboring) nodes
    - Does not have to move through any other type of data structures to find links between the nodes.
  - Not all graph DBs use native storage
    - Some may use other dbs such as RDBMS

Reading: https://neo4j.com/blog/native-vs-non-native-graph-technology/

Graph Compute Engines

- Enables graph computational algorithms
  - Clustering
  - Shortest Path
  - How many flights exist between New Delhi and San Francisco having less than 3 hops?
- Distributed graph compute engines examples
  - Pegasus
  - Giraph

Why Graph DBs?

- Some data are naturally graphs
- Performance when dealing with graph data
  - Execution time for each query is proportional only to the size of the part of the graph traversed (not the size of the entire graph)
- Ease of maintenance
  - Graphs are naturally additive
    - New labels/relationships/nodes can be added without disturbing existing application features
    - Helps in agility while designing graph based applications
  - Schema free

Universe of Graph DBs

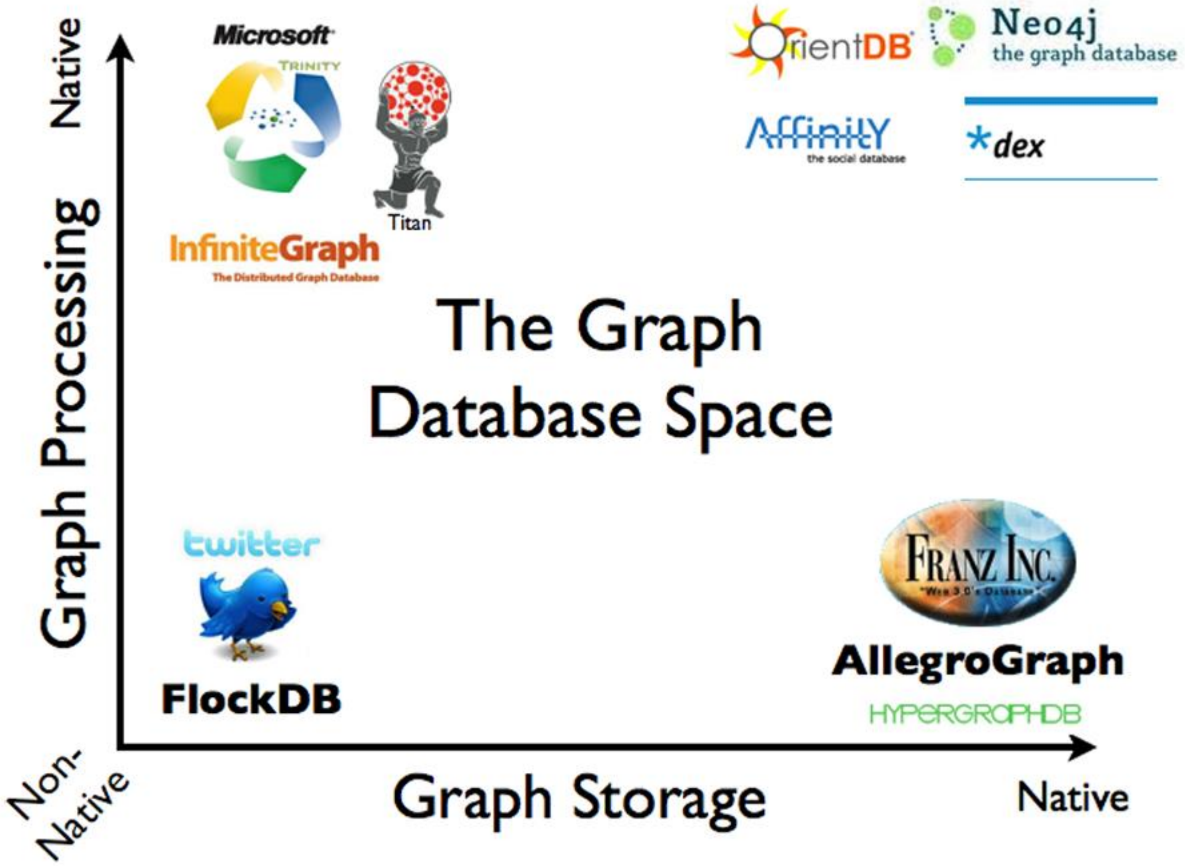


Image Source: Graph Databases, 2nd Edition

Labeled Property Graph Model

- A labelled property graph is made up of nodes, relationships, properties and labels
    - Nodes contain properties
    - Nodes can be tagged with one or more labels
    - Relationships connect nodes
    - Relationships can also have properties

Neo4j

- A leading graph database, with native graph storage and processing.
- Open Source
- NoSQL
- ACID compliant

Neo4j Sandbox

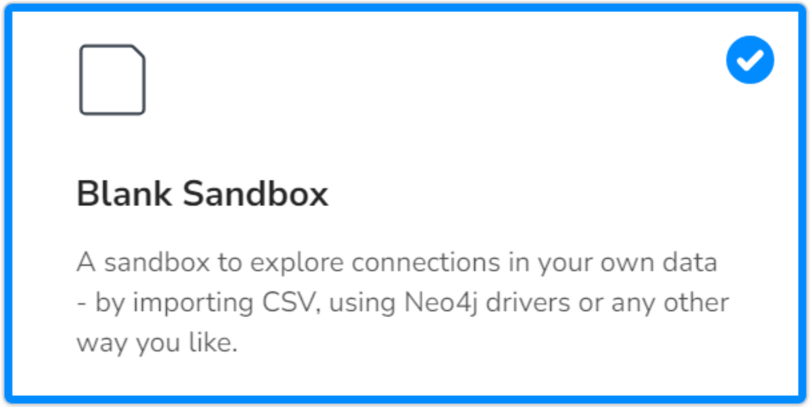
https://sandbox.neo4j.com/

Neo4j Desktop

https://neo4j.com/download

Using Neo4J Sandbox

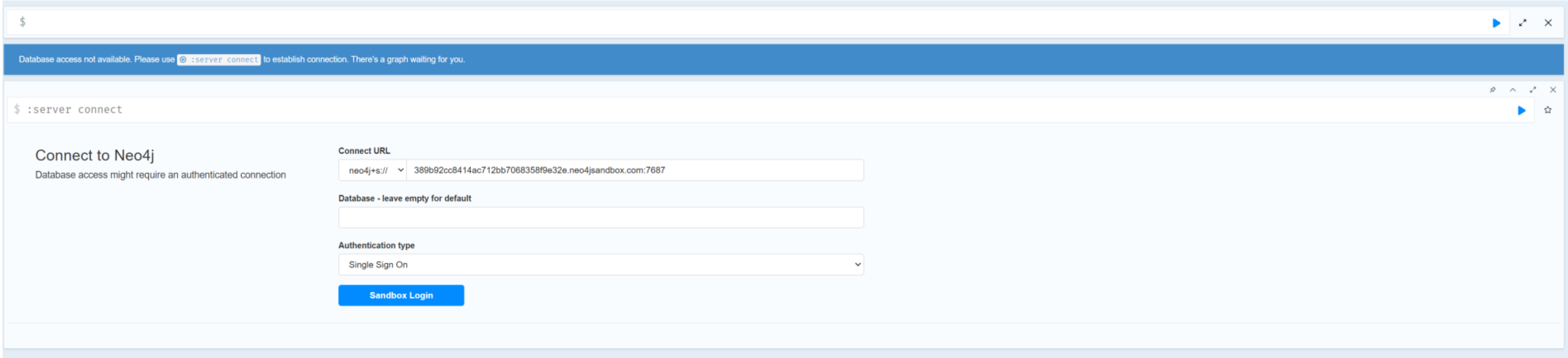
- Sign up at https://sandbox.neo4j.com/
- Select a project – Blank Sandbox



Blank Sandbox

A sandbox to explore connections in your own data - by importing CSV, using Neo4j drivers or any other way you like.

- Open the project



```
$
```

Database access not available. Please use :server connect to establish connection. There's a graph waiting for you.

```
$ :server connect
```

Connect to Neo4j

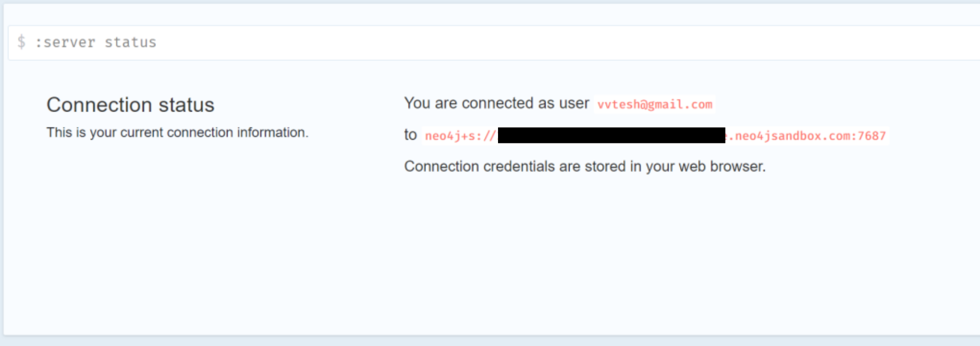
Database access might require an authenticated connection

Connect URL: neo4j+s:// 389b92cc8414ac712bb7068358f9e32e.neo4jsandbox.com:7687

Database - leave empty for default:

Authentication type: Single Sign On

- Hit the Sandbox Login and authenticate once again



```
$ :server status
```

Connection status

This is your current connection information.

You are connected as user vvtesh@gmail.com

to neo4j+s://[redacted].neo4jsandbox.com:7687

Connection credentials are stored in your web browser.

Data Model

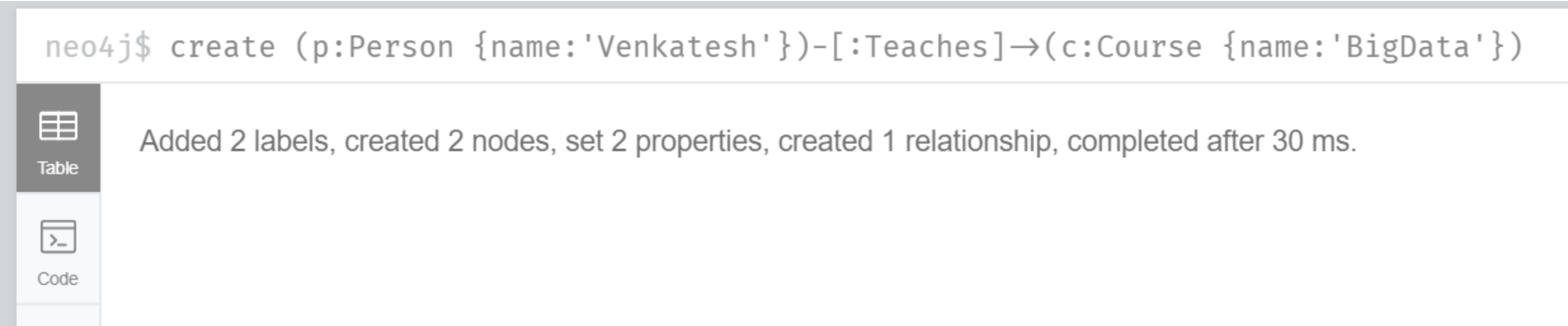
- create (p:Person {name:'Venkatesh'})-[:Teaches]->(c:Course {name:'BigData'})

Query Language

- Cypher Query Language
  - Similar to SQL
  - Optimized for graphs
  - Used by Neo4j, SAP HANA Graph, Redis Graph, etc.

CQL

- create (p:Person {name:'Venkatesh'})-[:Teaches]->(c:Course {name:'BigData’})
- Don’t forget the single quotes.

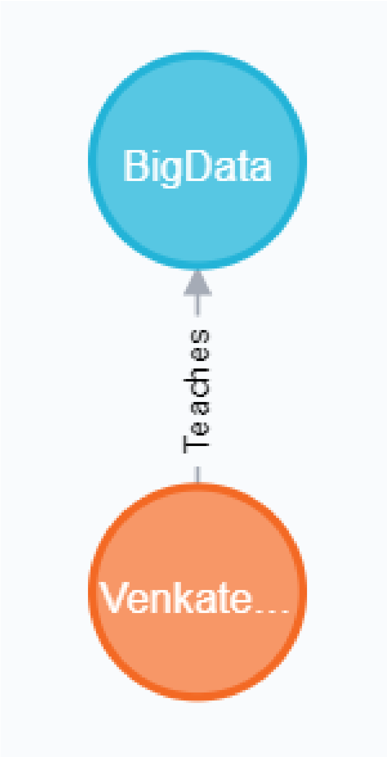


```
neo4j$ create (p:Person {name:'Venkatesh'})-[:Teaches]→(c:Course {name:'BigData'})
```

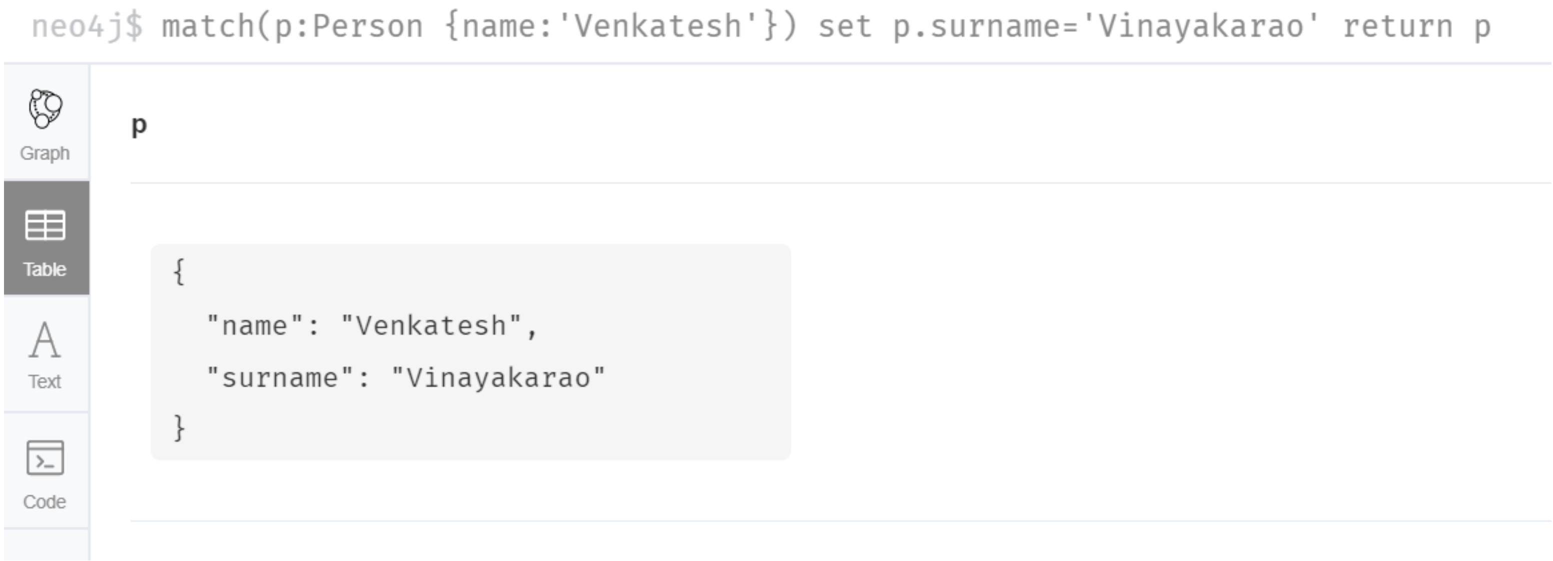
Added 2 labels, created 2 nodes, set 2 properties, created 1 relationship, completed after 30 ms.

CQL

- Match (n) return n



- match(p:Person {name:'Venkatesh'}) set p.surname='Vinayakarao' return p

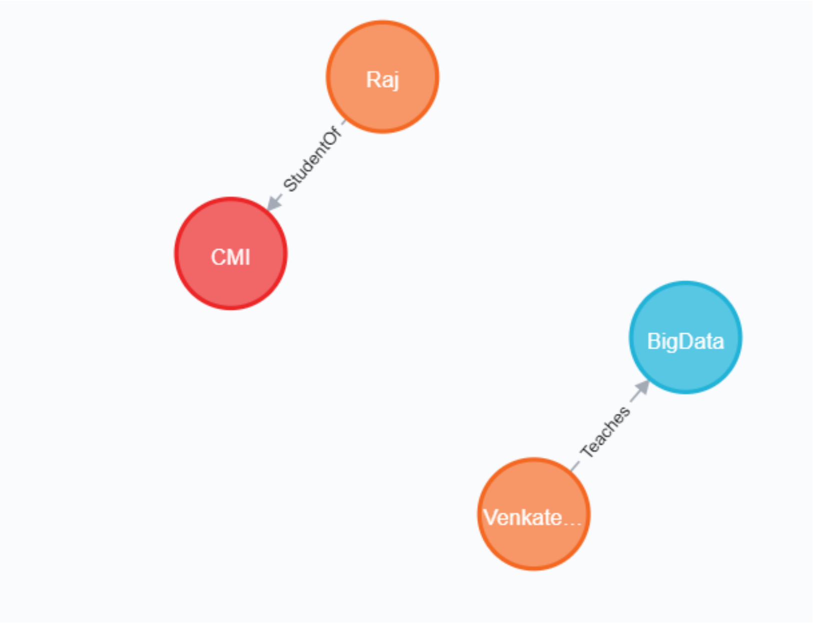


```
neo4j$ match(p:Person {name:'Venkatesh'}) set p.surname='Vinayakarao' return p
```

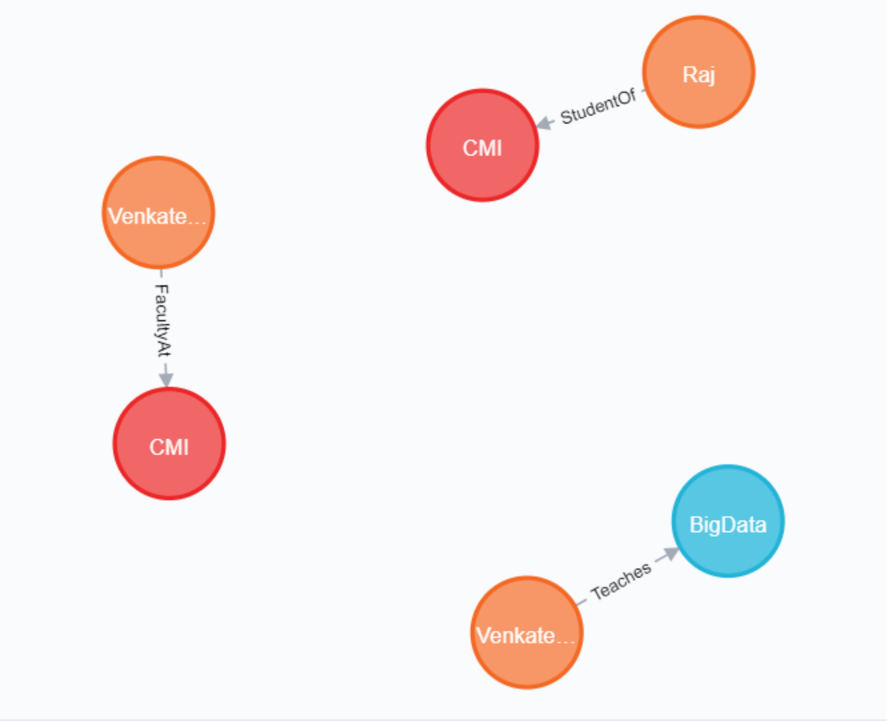
p

```
{
  "name": "Venkatesh",
  "surname": "Vinayakarao"
}
```

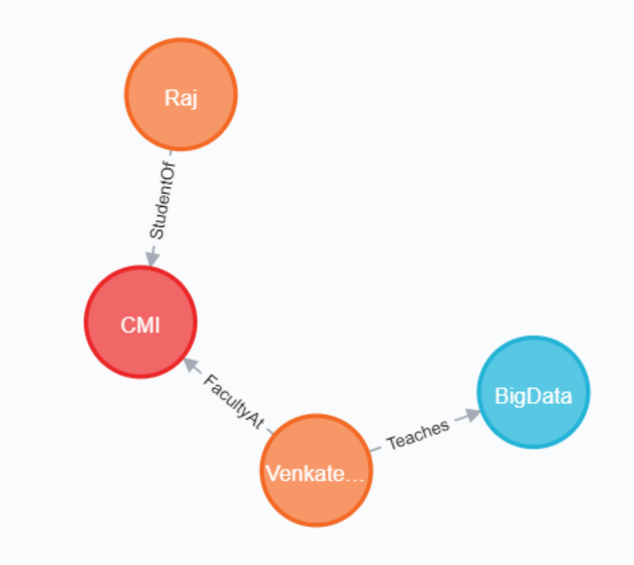
- Create (p:Person {name:’Raj’})-[:StudentOf]->(o:Org {name:’CMI’}
- Match (n) return n



- create (p:Person {name:'Venkatesh'})-[:FacultyAt]->(o:Org {name:'CMI’})
- Match (n) return n



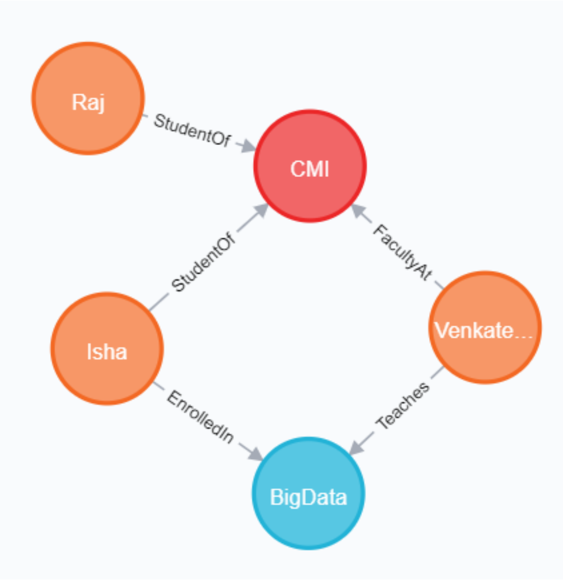
- MATCH (p:Person) where ID(p)=4
- DELETE p
- MATCH (o:Org) where ID(o)=5
- DELETE o



- MATCH (a:Person),(b:Org)
- WHERE a.name = 'Venkatesh' AND b.name = 'CMI'
- CREATE (a)-[:FacultyAt]->(b)

- MATCH (a:Person),(b:Course)
- WHERE a.name = 'Isha' and b.na
- CREATE (a)-[:StudentOf]->(b)

- MATCH (a:Person)-[o:StudentOf] ID(o)=4
- DELETE o



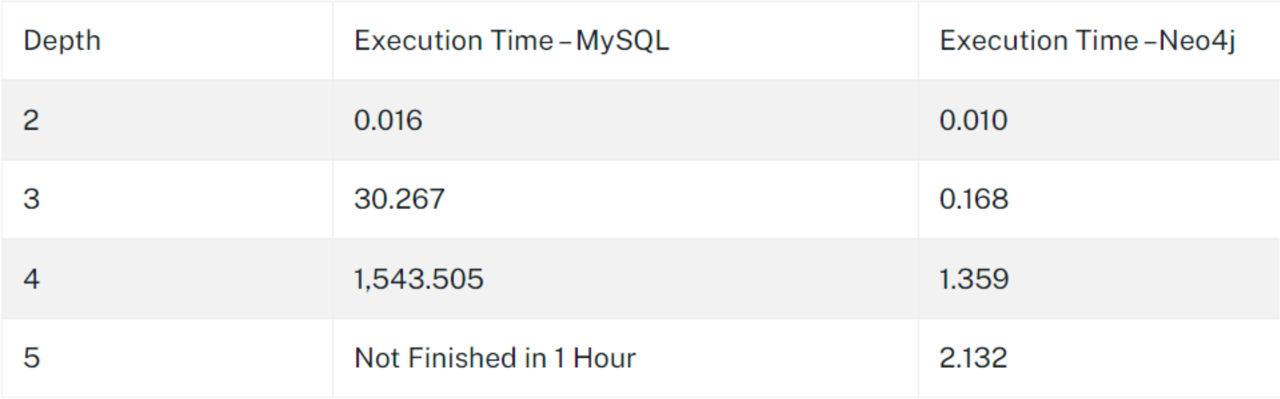
- MATCH (a:Person),(b:Course)
- WHERE a.name = 'Isha' and b.name = 'BigData'
- CREATE (a)-[:EnrolledIn]->(b)

Graph Query Languages

- Just like Cypher, there are other graph query languages
    - Gremlin
    - GSQL
    - Morpheus

How fast are graph DBs?

- Social network: Find all friends of a user’s friends, and friends of friends of friends.
- Built the query in both MySQL and Neo4j with a DB of 1M users.

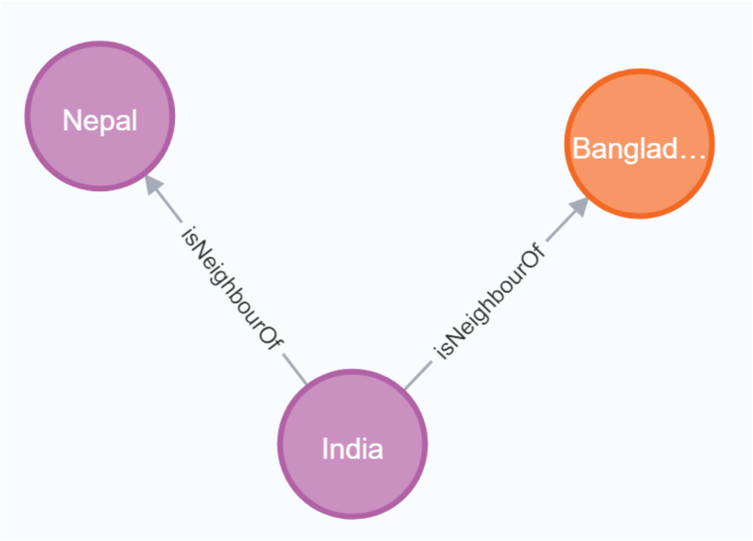


Depth	Execution Time – MySQL	Execution Time –Neo4j
2	0.016	0.010
3	30.267	0.168
4	1,543.505	1.359
5	Not Finished in 1 Hour	2.132

https://neo4j.com/news/how-much-faster-is-a-graph-database-really/

Exercise

- Sign up at neo4j
- Create the following graph



Note

- All relationships are directional in Neo4J
- However, at query time, you may ignore the direction
  - MATCH (x)-[:isNeighbourOf]-(y)

Thank You