Introduction to Database Systems: CS305 Neo4J

Oliver Bonham-Carter
Hang Zhao

28 November 2023

Meaningful Information Should Come From Data

Having data is a small part of it...

- I have raw data to explore
- I want information and *meaning* from this data

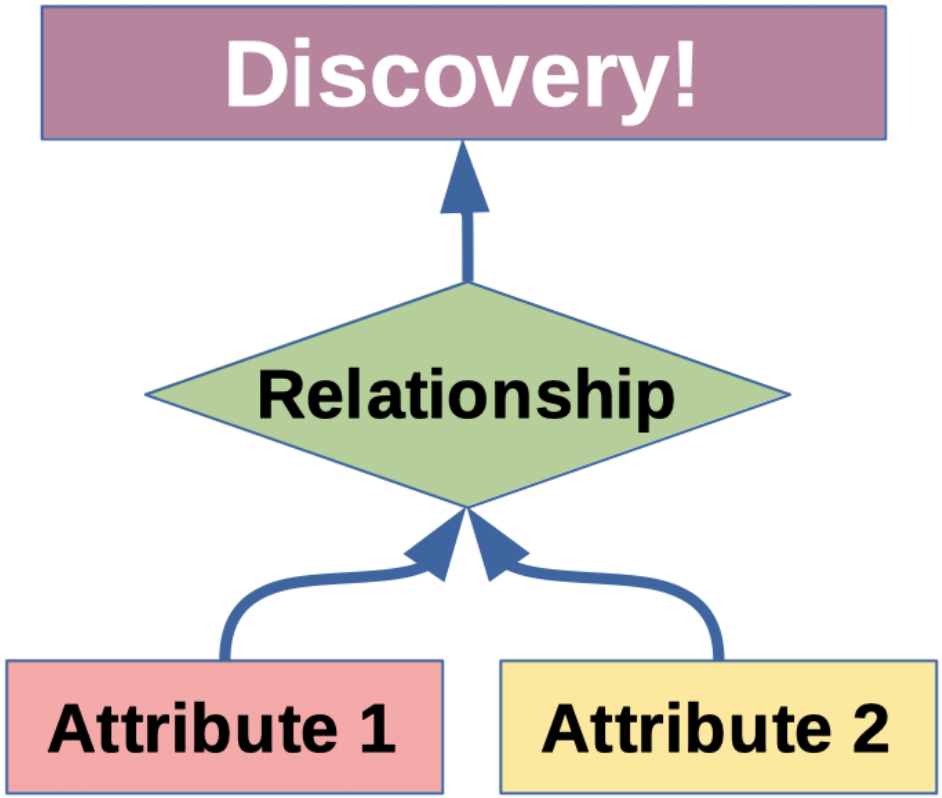
Explore The Data

humanGene	EnsNum	x00511204	x7d9d7119	x93904035
RMND5A	ENSG00000153561.11	16.0546348885	15.6436361402	151243.109382
RAD23A	ENSG00000179262.8	38.9356481105	21.5142980465	775745.038464
RAD17	ENSG00000152942.17	6.71326600879	5.55100617026	151541.361155
TTDN1 (C7orf11)	ENSG00000168303.6	1.85918994126	3.36634373043	49263.8903263
RAD54L	ENSG00000085999.10	0.00970150764521	4.41325732573	15129.8861733
UBE2N	ENSG00000177889.8	10.5477997615	8.83952862957	359788.007983
TMEM30A	ENSG00000112697.14	24.071953429	65.9105478055	702850.166466
POLG	ENSG00000140521.10	11.0086481904	14.6093304994	264802.654955
TIPIN	ENSG00000075131.8	1.0519040137	3.4787739239	46372.2363056
RECQL	ENSG00000004700.14	7.34079033224	13.8899052998	156082.413636
BRCA2 (FANCD1)	ENSG00000139618.13	0.0304680934309	2.60236876714	8123.47419519
RPA3	ENSG00000106399.10	2.73817849196	11.9965343474	98123.2266513
RNASEH2B	ENSG00000136104.17	2.25140800487	2.16690519349	51635.1402182
RAD18	ENSG00000070950.8	1.03082443513	5.06228468473	48787.2494237
CAMKK1	ENSG00000004660.13	0.715650842655	1.95868467159	87931.7903047

- I just collected some data and should store it in a database
- So, I have poured this data into some SQL tables I made
- I should now write some useful queries for some unique purpose
- Intelligence should result from these queries
- Right?

I Want To Know

What is the relationship between ...



- I want to know what relationship(s) exist between my attributes
- This relationship would be an amazing discovery!

Explore The Data

So, I will take a stab using SQL to find meaning ...

humanGene	EnsNum	x00511204	x7d9d7119	x93904035
RMND5A	ENSG00000153561.11	16.0546348885	15.6436361402	151243.109382
RAD23A	ENSG00000179262.8	38.9356481105	21.5142980465	775745.038464
RAD17	ENSG00000152942.17	6.71326600879	5.55100617026	151541.361155
TTDN1 (C7orf11)	ENSG00000168303.6	1.85918994126	3.36634373043	49263.8903263
RAD54L	ENSG00000085999.10	0.00970150764521	4.41325732573	15129.8861733
UBE2N	ENSG00000177889.8	10.5477997615	8.83952862957	359788.007983
TMEM30A	ENSG00000112697.14	24.071953429	65.9105478055	702850.166466
POLG	ENSG00000140521.10	11.0086481904	14.6093304994	264802.654955
TIPIN	ENSG00000075131.8	1.0519040137	3.4787739239	46372.2363056
RECQL	ENSG00000004700.14	7.34079033224	13.8899052998	156082.413636
BRCA2 (FANCD1)	ENSG00000139618.13	0.0304680934309	2.60236876714	8123.47419519
RPA3	ENSG00000106399.10	2.73817849196	11.9965343474	98123.2266513
RNASEH2B	ENSG00000136104.17	2.25140800487	2.16690519349	51635.1402182
RAD18	ENSG00000070950.8	1.03082443513	5.06228468473	48787.2494237
CAMKK1	ENSG00000004660.13	0.715650842655	1.95868467159	87931.7903047

What `EnsNum` do I want ... ?

```
SELECT humanGene WHERE EnsNum LIKE "E%";
SELECT humanGene WHERE x00511204 like "16%";
SELECT err ... what's for lunch?
SELECT a soup and salad, I guess
```

- What was that pattern I was looking for?
- What happened to my quest to *extract meaning from my data*?

Introduction to Database Systems: CS305 Neo4J

Oliver Bonham-Carter
Hang Zhao

A Missed Discovery

Neo4J

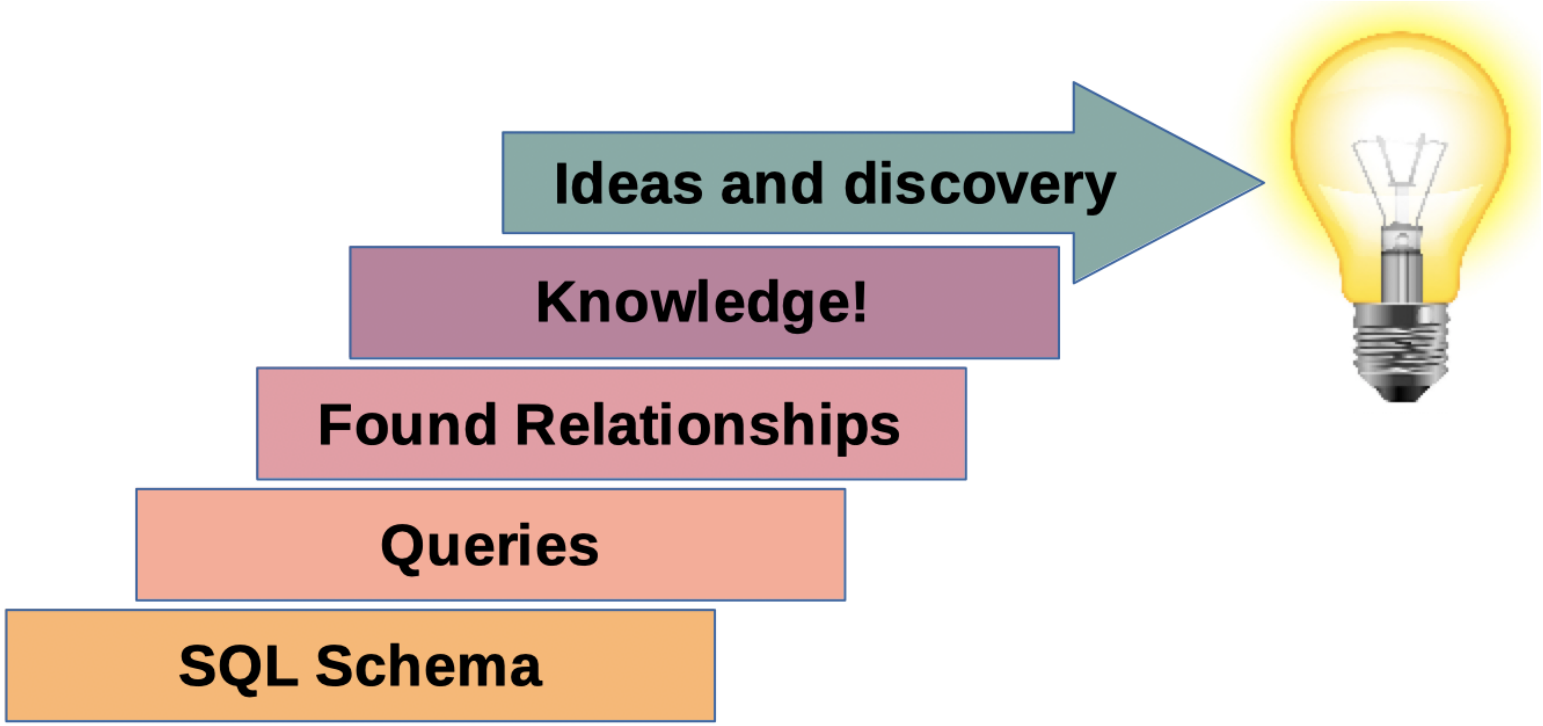
Start Neo4j in Docker

Open a Project

Play Time
Finding Kevin Bacon

Shutting Down

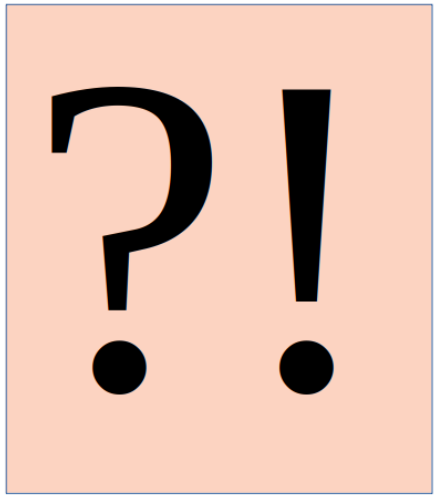
Consider this



humanGene	EnsNum	x00511204	x7d9d7119	x93904035
RMND5A	ENSG00000153561.11	16.0546348885	15.0430361402	151243.109382
RAD23A	ENSG00000179262.5	38.9356481105	21.5142980486	775745.036484
RAD17	ENSG00000152942.17	6.71326600079	5.55100617028	151541.361155
TTDN1 (C7orf11)	ENSG00000168303.5	1.85918994126	3.36634373043	49283.8903263
RAD54L	ENSG00000085999.10	0.009701507/64521	4.41325732573	15129.8861733
UBE2N	ENSG00000177889.9	10.5477997015	8.83952862957	359788.007883
TMEM30A	ENSG00000112697.14	24.071953429	65.9105478056	702950.166466
POLG	ENSG00000140521.10	11.0086481904	14.6093304994	264902.854955
TIPIN	ENSG00000075131.5	1.0519040137	3.4787739239	46372.2363056
RECQL	ENSG00000004700.14	7.340/9033224	13.889906/998	156082.413536
BRCA2 (FANCD1)	ENSG00000139618.13	0.0304680934309	2.60230876714	8123.47419519
RPA3	ENSG00000106399.10	2.73517849196	11.9965343474	98123.2266513
RNASEH2B	ENSG00000136104.17	2.25140800487	2.16690519349	51635.1402182
RAD18	ENSG00000070950.5	1.03082443513	5.06228468473	48757.2494237
CAVIK1	ENSG00000004660.13	0.715850842656	1.9586846/159	87931.790304/

Missing Discoveries?

Where did my idea go?



What stumped my discovery?

- Discoveries in data are first imagined, then verified
- The patterns that we can find are limited by our imaginations to find a *testable* cases to query
- Is there a way to find relationships without first knowing that they could exist?!

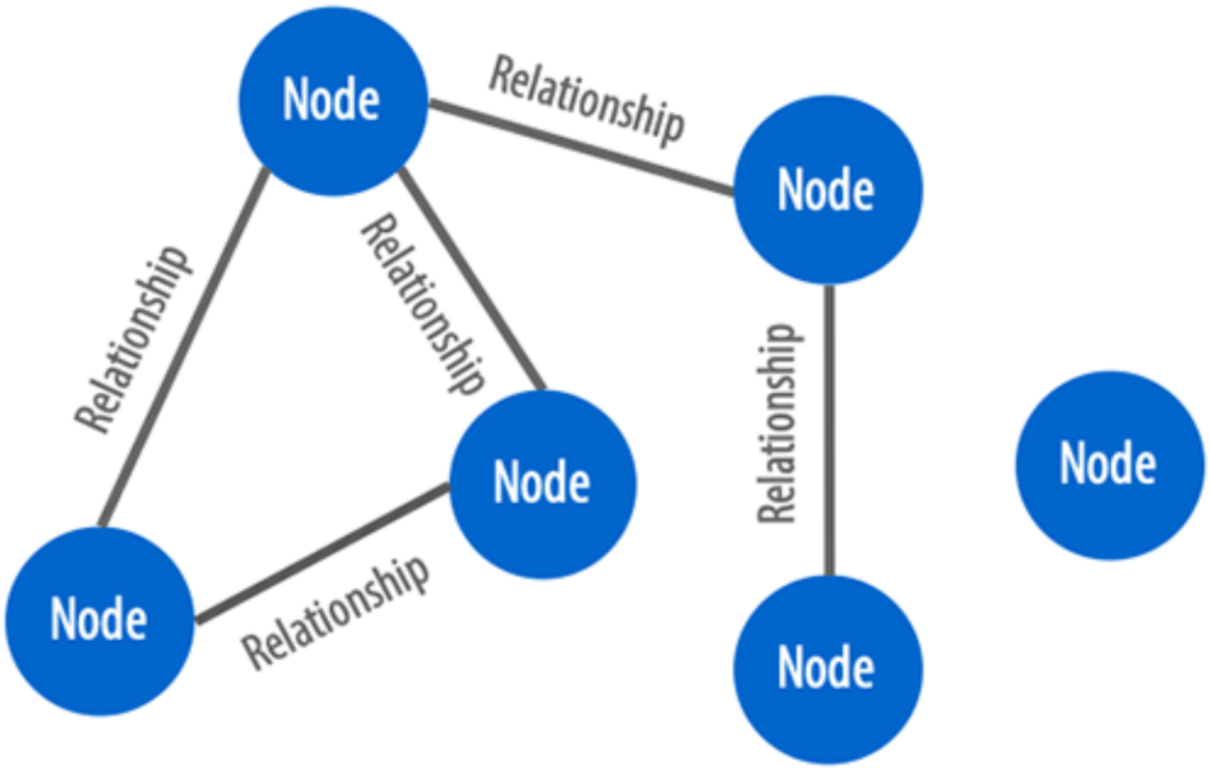
Databases, Visually

- A visual database system using methods from graph theory to use networks to determine relationships (edges) and discover meaning from connected data-points (nodes). Users are able to interact with the data in a network.

- `https://neo4j.com/`
- Graphgists Projects: `https://neo4j.com/graphgists/`

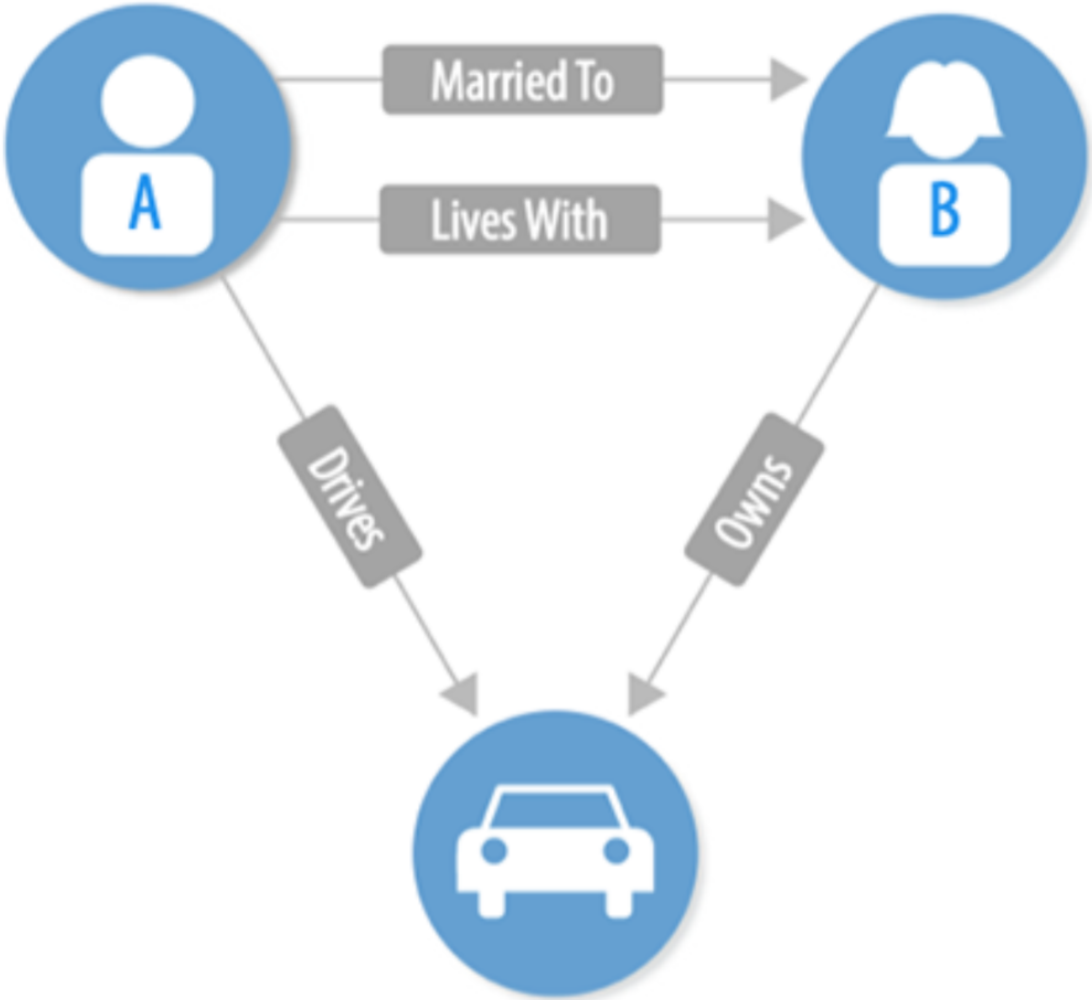
Networks Of Data

Relationships exist by connectivity



- Nodes and edges represent inter-relationships
- Relationships are described by connections between nodes
- Single nodes have no immediate relationships with the others

Networks In Neo4J



- An acting schema: The relationships between nodes are built into the network

Networks In Neo4J

Introduction to Database Systems: CS305 Neo4J

Oliver Bonham-Carter Hang Zhao

A Missed Discovery

Neo4J

Start Neo4j in Docker

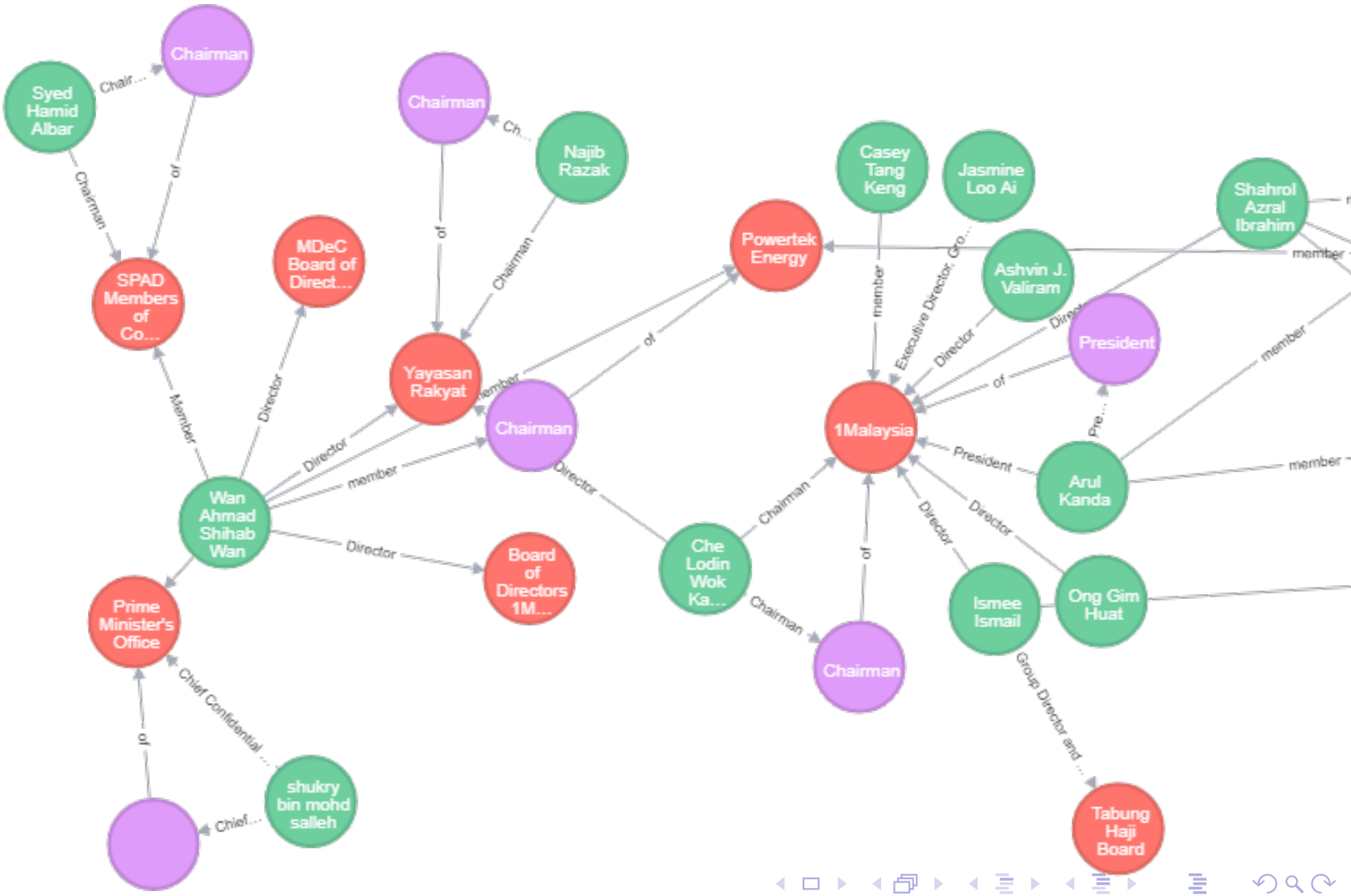
Open a Project

Play Time

Finding Kevin Bacon

Shutting Down

Consider this



Networks Of Data

Relationships exist by connectivity

Introduction to Database Systems: CS305 Neo4J

Oliver Bonham-Carter
Hang Zhao

A Missed Discovery

Neo4J

Start Neo4j in Docker

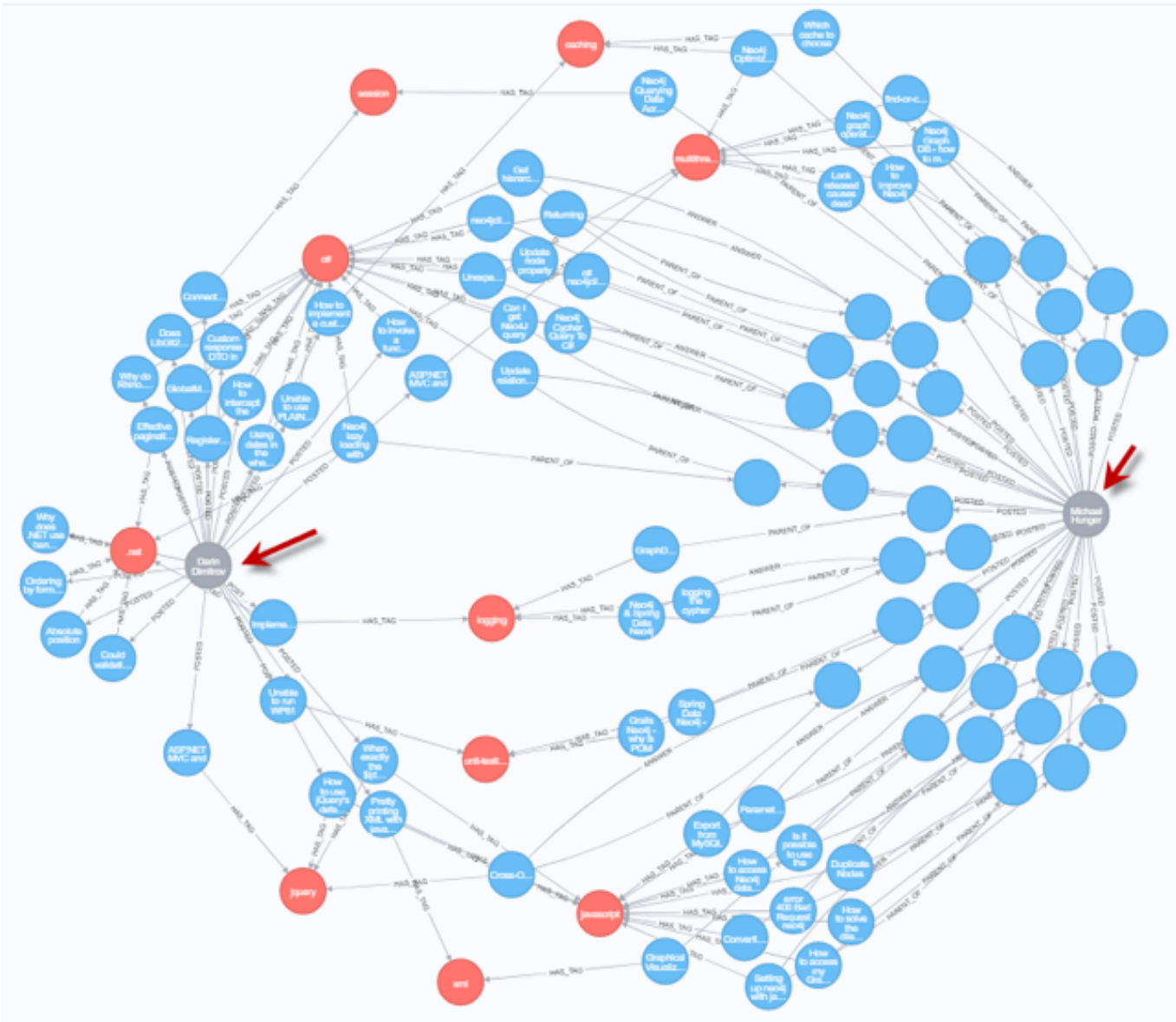
Open a Project

Play Time

Finding Kevin Bacon

Shutting Down

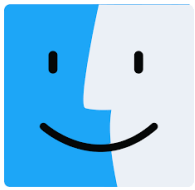
Consider this



Getting started with Neo4j in Docker

These files are located in `sandbox/`





Windows

```
build_neo4j_windows.bat
```

MacOS and Linux

```
sh build_neo4j_macOSAndLinux.sh
```

You can **build** and **start** the container with this script. You will have to manually stop the container, as necessary.

Getting started with Neo4j in Docker

Specific Terminal commands

Introduction to Database Systems: CS305 Neo4J

Oliver Bonham-Carter Hang Zhao

A Missed Discovery

Neo4J

Start Neo4j in Docker

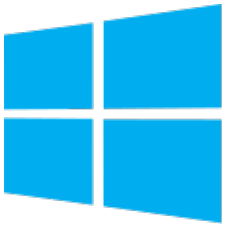
Open a Project

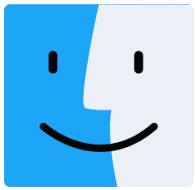
Play Time

Finding Kevin Bacon

Shutting Down

Consider this





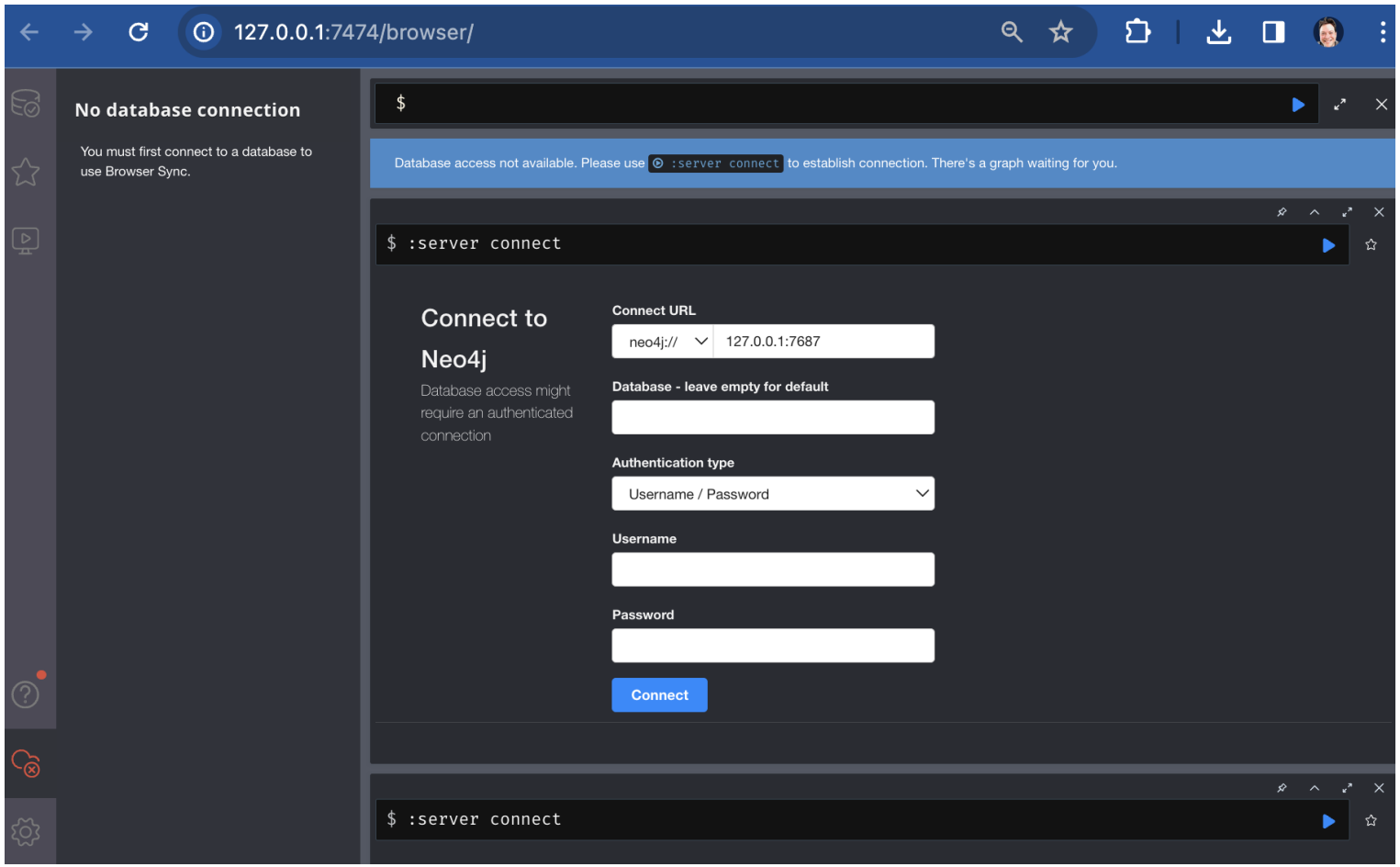
Terminal Command to START Neo4j

```
docker start testneo4j # windows
sudo docker start testneo4j # MacOS and Linux
```

Terminal Command to STOP Neo4j

```
docker stop testneo4j # windows
sudo docker stop testneo4j # MacOS and Linux
```

Login



- Open your browser and head to: `http://127.0.0.1:7474/browser/`

User and Password

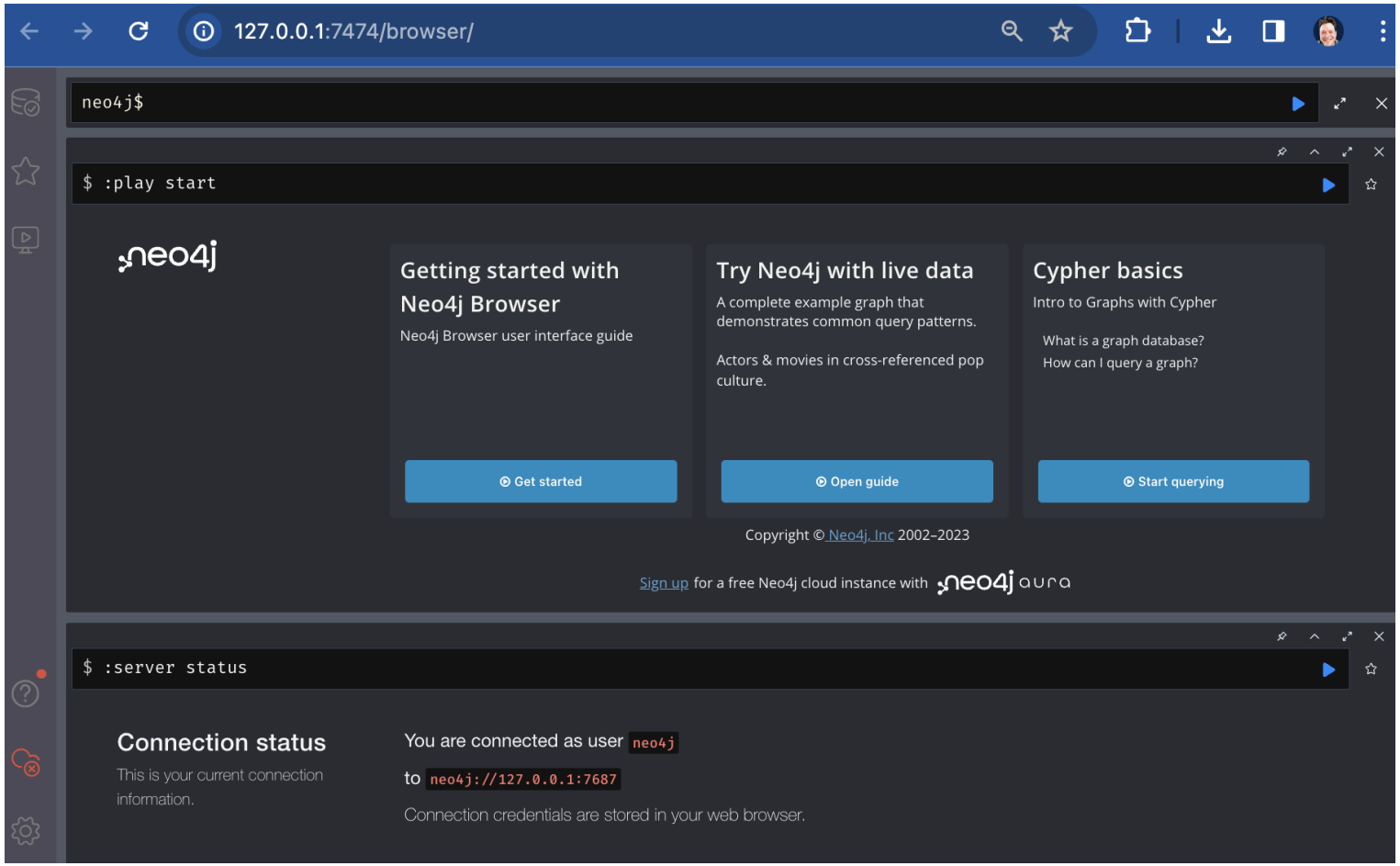
Note: The user and password variables are defined in the *build* files we used to create the Docker container.

- Your first login
  - **User**: `neo4j`
  - **Password**: `password`

Parameter in the build file

```
--env NEO4J_AUTH=neo4j/password
```

Ready!



- If all has gone well, you should be ready to work

Ready!

```
$ :play movie graph
```

- Type `:play movie graph` in the editor at the top.
- Now click right arrow

```
$ :play movie graph
```

Movie Graph

Pop-cultural connections between actors and movies

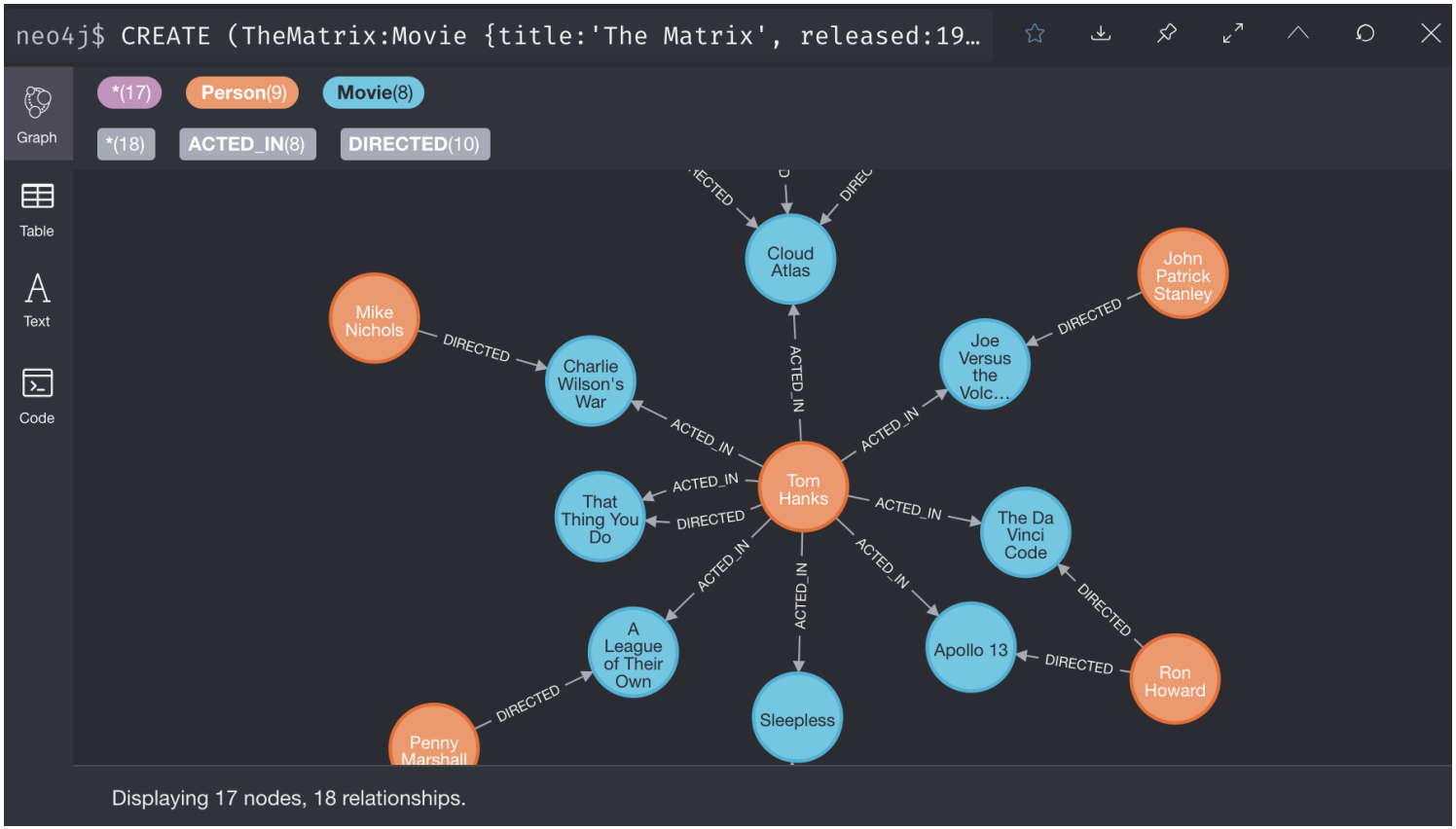
The Movie Graph is a mini graph application containing actors and directors that are related through the movies they've collaborated on.

This guide will show you how to:

1. Create: insert movie data into the graph
2. Find: retrieve individual movies and actors
3. Query: discover related actors and directors
4. Solve: the Bacon Path

- Let's follow the built-in tutorial of film data (i.e., Directors, Actors, Producers, etc.)

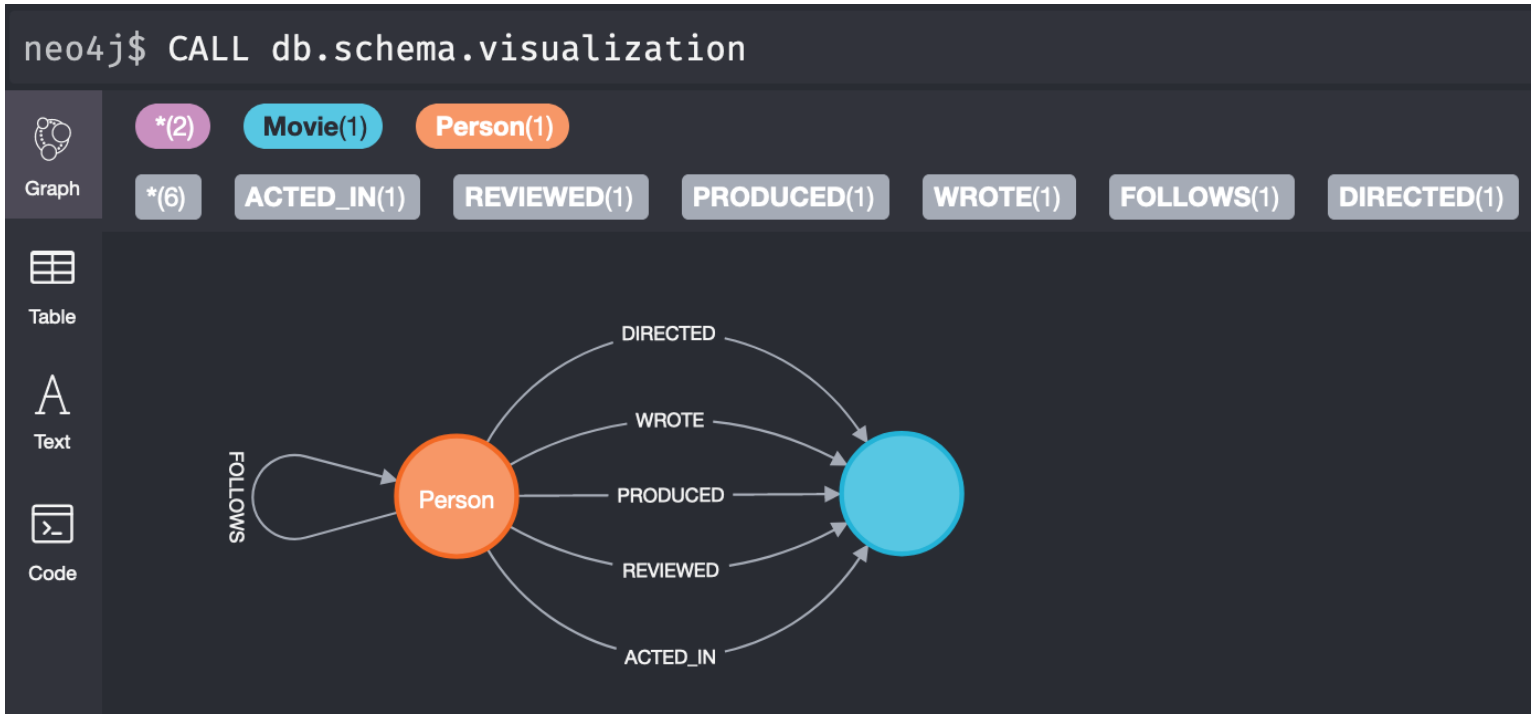
Play!



- Take a moment to play with the graph!
- Drag the nodes around!

Play!

Sample code in Cypher script



What is the Visual Schema?

```
CALL db.schema.visualization
```

More help?

Visit: `https://neo4j.com/developer/cypher/guide-cypher-basics/`

Play!

Sample code in Cypher script

What are the node types?

```
CALL db.schema.nodeTypeProperties
```

What are the relationship types?

```
CALL db.relationshipTypes()
```

Display all nodes

```
MATCH (n) RETURN n
```

Who acted in what?

```
MATCH p=()-[r:ACTED_IN]->() RETURN p
```

Who reviewed what?

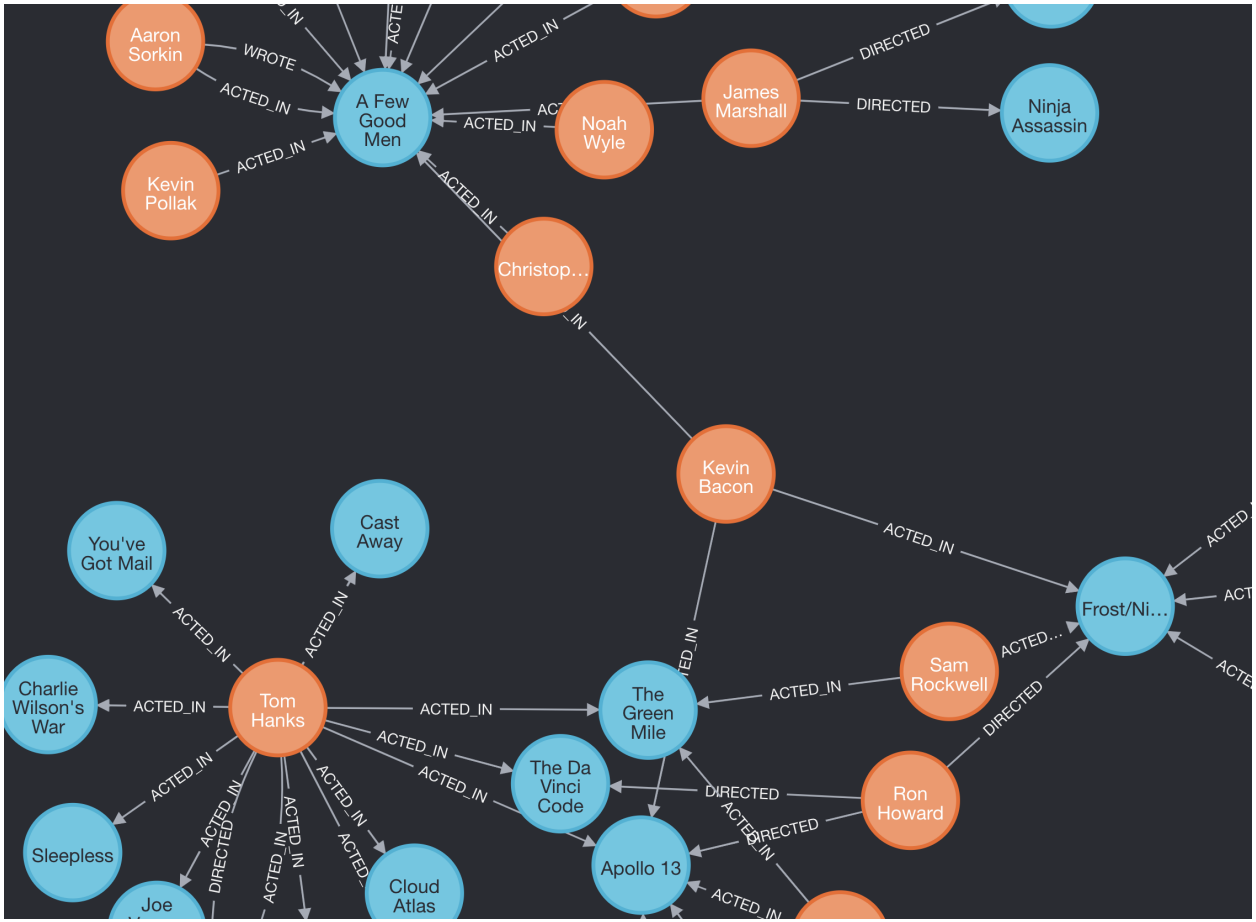
```
MATCH p=()-[r:REVIEWED]->() RETURN p LIMIT 25
```

Who produced what?

```
MATCH p=()-[r:PRODUCED]->() RETURN p LIMIT 25
```

Play!

Sample code in Cypher script



Where is Kevin Bacon?

```
MATCH (bacon:Person {name:"Kevin Bacon"})-[*1..3]-(hollywood)
RETURN DISTINCT bacon, hollywood
```

How To Shut Down a Session

Stop Neo4j container

```
docker stop testneo4j # Windows
sudo docker stop testneo4j # MacOS and Linux
```

Remove Neo4j container (if necessary!)

```
sudo docker image rm neo4j # MacOS and Linux
docker image rm neo4j # Windows
```

Consider This...

- Can you work with data as nodes and edges in the `movie` network?
- Can you discover new relationships between the nodes?