



# Introduction to Database Systems: CS312

## Neo4J

Oliver BONHAM-CARTER

5<sup>th</sup> May2022

# 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

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this

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 ...

A Missed  
Discovery

Neo4J

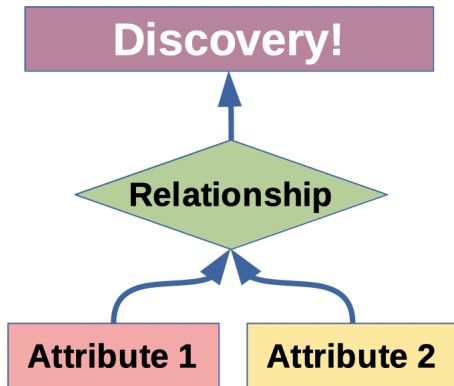
Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this



- 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 ...

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this

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*?

# Using Databases

## Data to Discovery

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this

Ideas and discovery

Knowledge!

Found Relationships

Queries

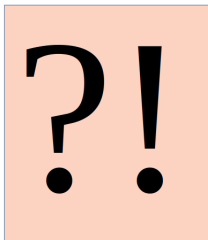
SQL Schema



humanGene	Ensembl	x00511204	x7d9d7119	x93904035
RMN1XA	ENSG00000153561.11	16.05406348886	15.6430361402	15.243.109382
RAD23A	ENSG00000179282.8	38.9356481105	21.5142980465	775745.036464
RAD17	ENSG00000152942.17	6.71326600079	5.55100617028	15.541.361155
TTDN1 (C7orf11)	ENSG00000169303.5	1.85218994128	3.36634373043	49253.8903263
URA5/L	ENSG00000385889.10	0.00970150767521	4.41325732573	15.29.8861733
UBE2N	ENSG00000177889.8	10.5177097015	8.83952862957	359788.007983
TNFR30A	ENSG00000112687.14	24.071953429	65.9105478055	702930.166466
POLG	ENSG00000140521.10	11.0086481904	14.6093304954	264802.854955
TIPIN	ENSG00000075131.8	1.0519040137	3.4787739239	46372.2369056
RECQL	ENSG00000004700.14	7.34079033224	13.8899052988	165062.413836
BRCA2 (FANCD1)	ENSG00000130618.13	0.0304680594309	2.60230876714	8123.47419519
RPA3	ENSG00000100389.10	2.73817648196	11.9905343474	98723.2206513
RNASEC123	ENSG00000136154.17	2.25140800487	2.16690519349	51635.1402182
RAD18	ENSG00000070950.8	1.03082443513	5.06228468473	48757.2494237
CAVKK1	ENSG00000004680.13	0.71550842686	1.95868467159	87031.7903047

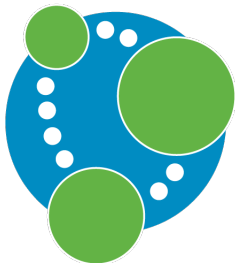
# Missing Discoveries?

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?!



# neo4j

- 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/>
- Graphists Projects: <https://neo4j.com/graphgists/>



# Networks Of Data

Relationships exist by connectivity

A Missed  
Discovery

Neo4J

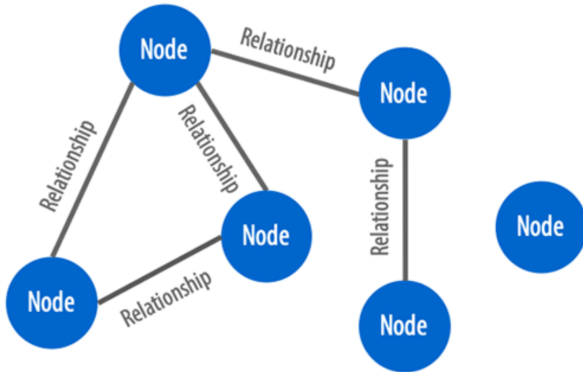
Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this



- 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

A Missed  
Discovery

Neo4J

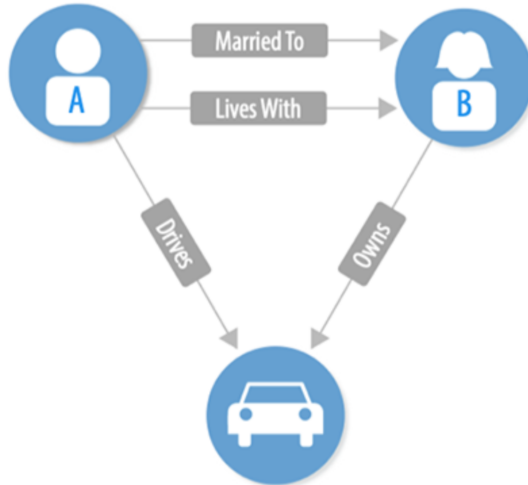
Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this



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

# Networks In Neo4J

A Missed  
Discovery

Neo4J

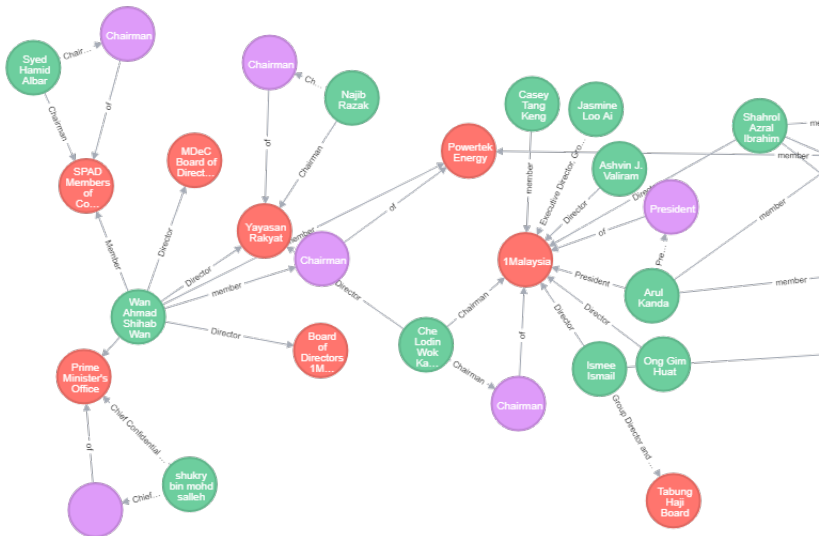
Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this





# Networks Of Data

## Relationships exist by connectivity

## A Missed Discovery

## Neo4J

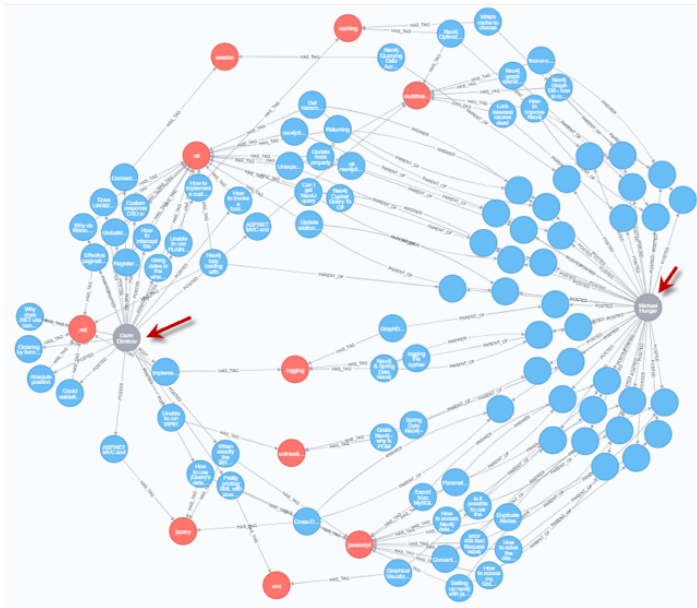
## Start Neo4j in Docker

## Open a Project

## Play Time

## Shutting Down

Consider this



# Getting started with Neo4j in Docker

These files are located in `sandbox/`

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this



Run batch file to create a Docker container from DockerHub for Neo4j: Windows.

```
run_neo4j_windows.bat
```

Run sh file to create a Docker container from DockerHub for Neo4j: MacOS and Linux.

```
sh run_neo4j_macOSAndLinux.sh
```

# Getting started with Neo4j in Docker

## Specific Terminal commands

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

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
```

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this

```
$ :server connect
```

## Connect to Neo4j

Database access requires an  
authenticated connection.

Connect URL

Username

Password

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

- Your first login

- **User:** neo4j
- **Password:** test

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker


Open a  
Project

Play Time

Shutting  
Down


Consider this

\$ :play start



### Learn about Neo4j

A graph epiphany awaits you.




What is a graph database?  
How can I query a graph?  
What do people do with Neo4j?

Start Learning

### Jump into code

Use Cypher, the graph query language.




Code walk-throughs  
RDBMS to Graph

Write Code

### Monitor the system

Key system health and status metrics.



Disk utilization  
Cache activity  
Cluster health and status

Monitor

Copyright © Neo4j, Inc 2002 - 2019

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



\$ :play movie graph



- 1 Type **:play movie graph** in the editor at the top.
- 2 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



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

A Missed  
Discovery

Neo4J

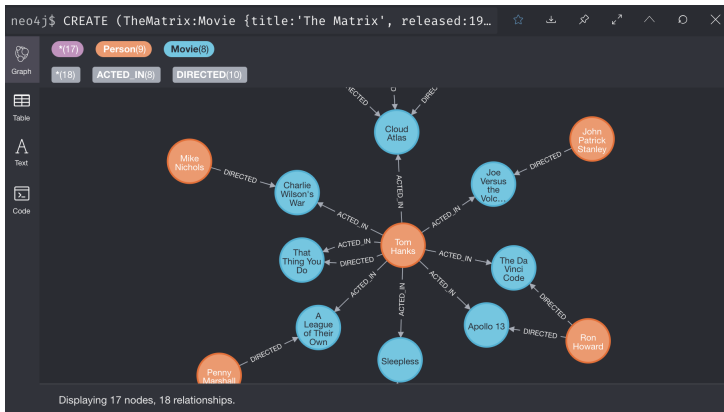
Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this



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

# Play!

Sample code in Cypher script

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

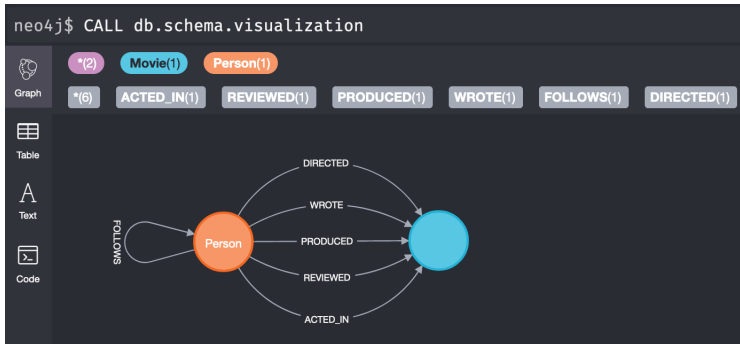
Open a  
Project

Play Time

Finding Kevin  
Bacon

Shutting  
Down

Consider this



## 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

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

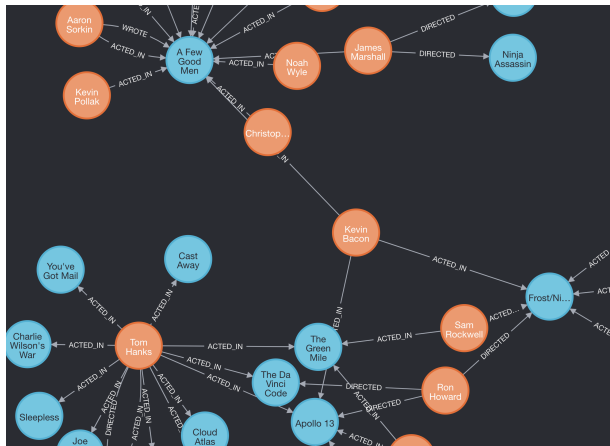
Open a  
Project

Play Time

Finding Kevin  
Bacon

Shutting  
Down

Consider this



## 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!)

```
docker rm containerID # Windows
```

```
sudo docker rm containerID # MacOS and Linux
```

# Consider This...

A Missed  
Discovery

Neo4J

Start Neo4j  
in Docker

Open a  
Project

Play Time

Shutting  
Down

Consider this



# THINK

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