:neo4j graphsummit

Create a Graph-Backed App from Scratch

March/April 2025

Goal of Today

Demonstrate (with YOUR HELP :-)), how to build a Graph “Application”

Agenda

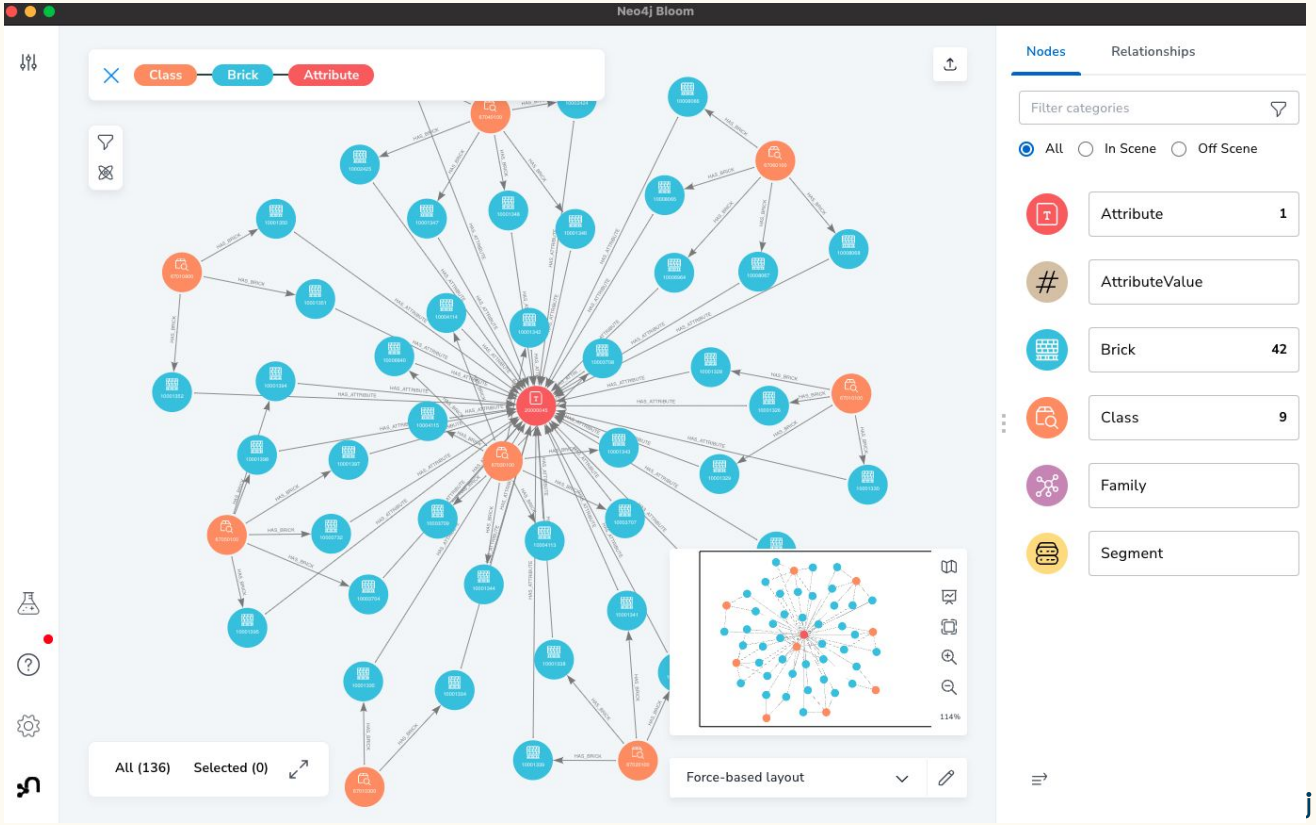
1. Logistics
2. Introduction
3. Use Case Explanation
4. Building the Solution

BREAK (15min) Around 15:30

5. Discussion: How to Improve the Use Case Further On
6. Final Q&A
7. Goodbye

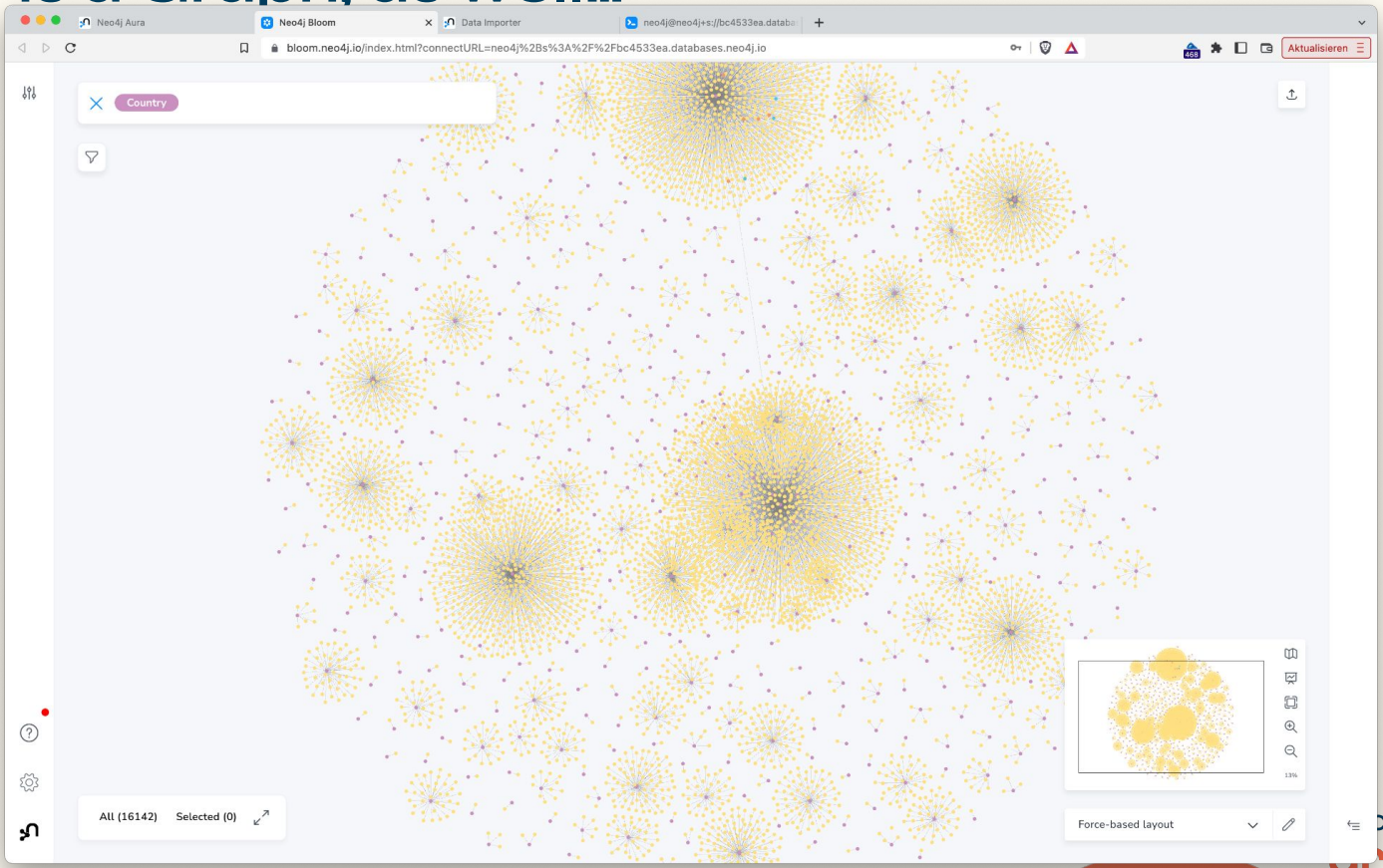
Introduction

This is a Graph …



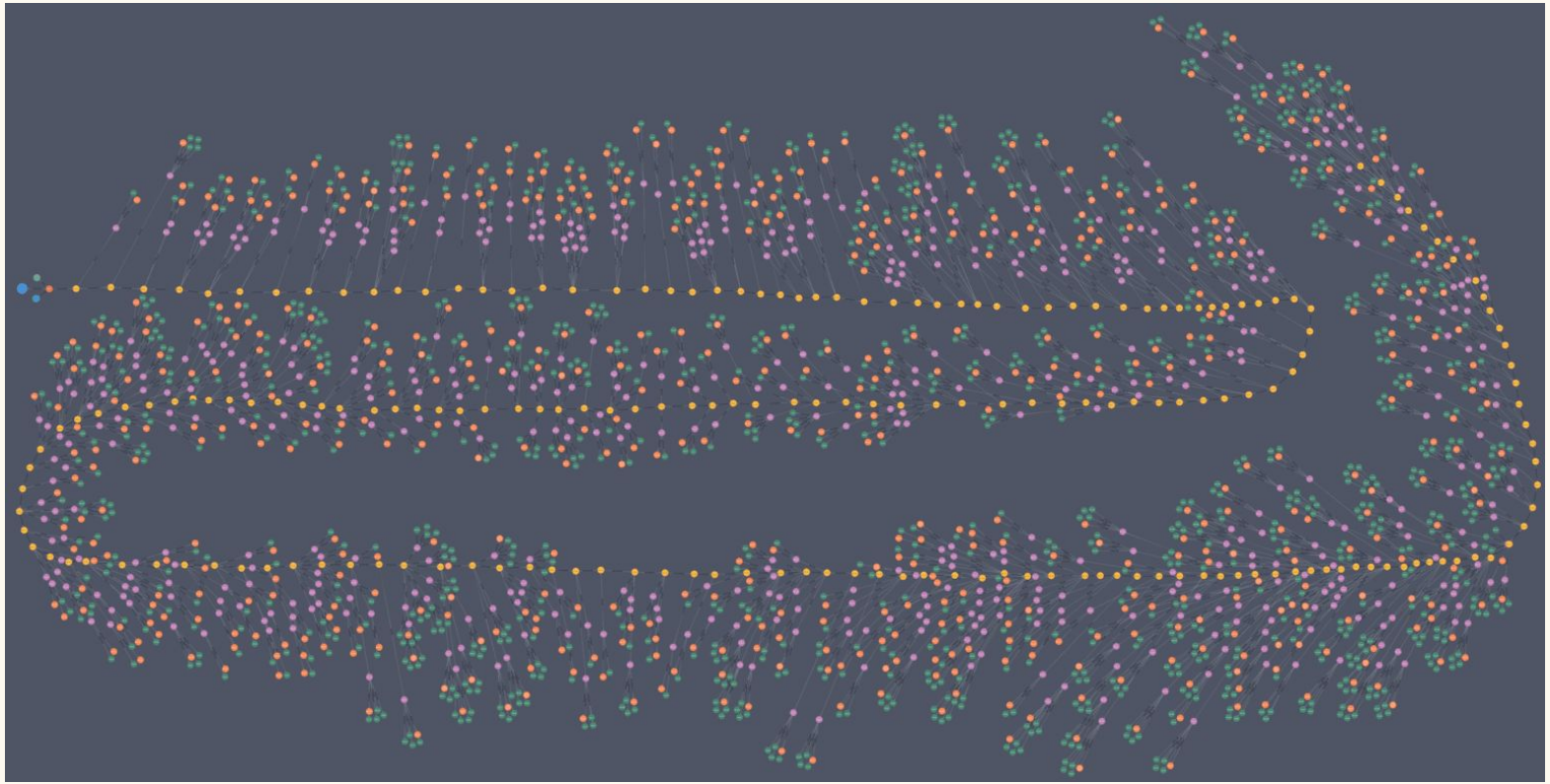
graphsummit

This is a Graph, as well...



graphsummit

...and this is also a Graph! Guess, what Graph could it be?



* https://www.graphable.ai/blog/patient-journey-mapping/

Patient Journey Mapping - Digital Twin

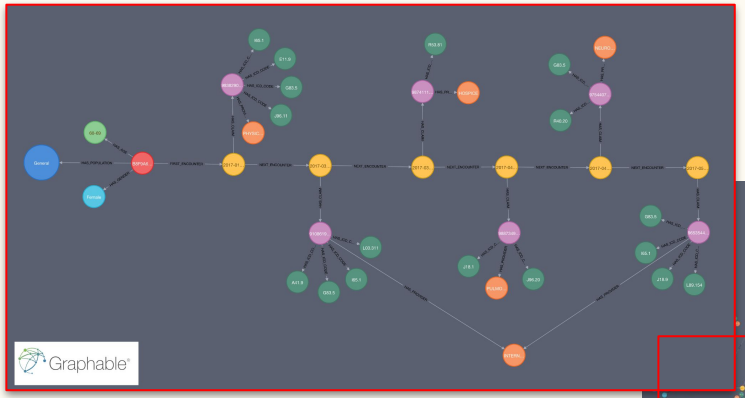


Figure 3: A patient graph schema supporting disease progression and cost exploration

Legend:

- **Patient**
- **Doctors Appointment**
- **Insurance Claim**
- **Results (ICD10 coded)**
- **Doctor**

6 Month Journey of a Breast Cancer Grade 1 Patient

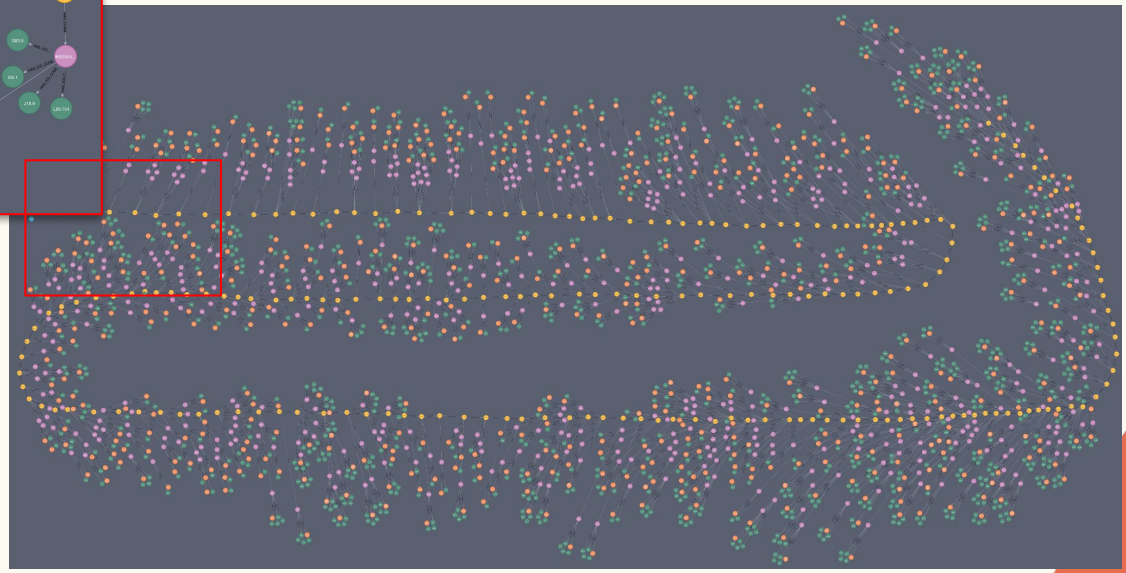


Figure 2: A single patient's clinical oncology journey

* https://www.graphable.ai/blog/patient-journey-mapping/

Labeled Property Graph Model

Nodes

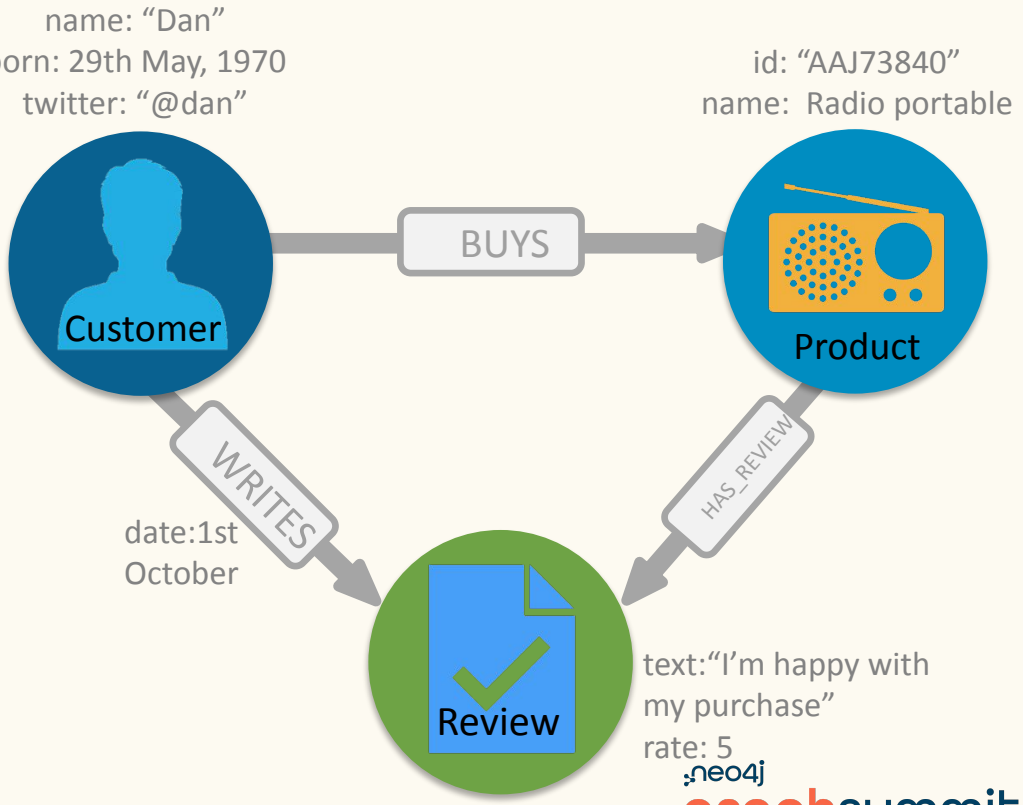
- Can have *Labels* to classify Nodes
- Labels have *native indexes*

Relationships

- Connect nodes by Type and Direction

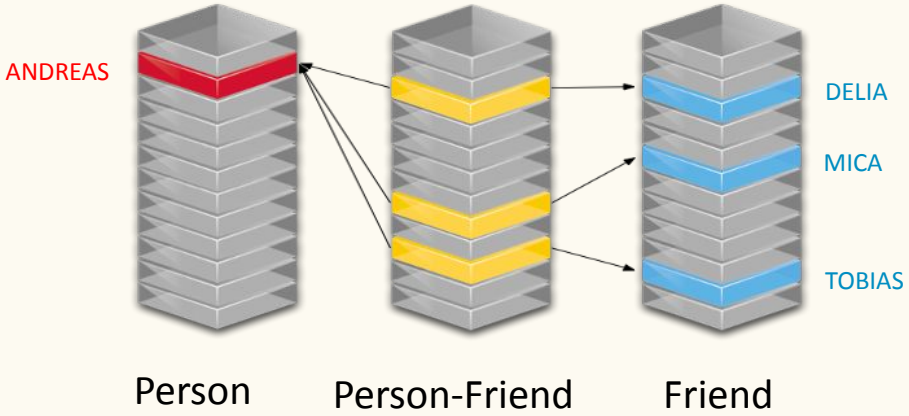
Properties

- Attributes of Nodes & Relationships
- Stored as Name/Value pairs
- Can have indexes and composite indexes

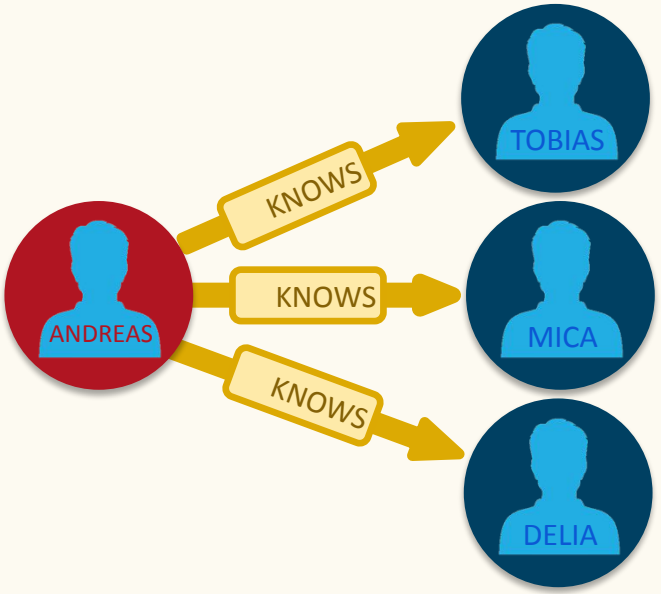


RDBMS vs Graph Model

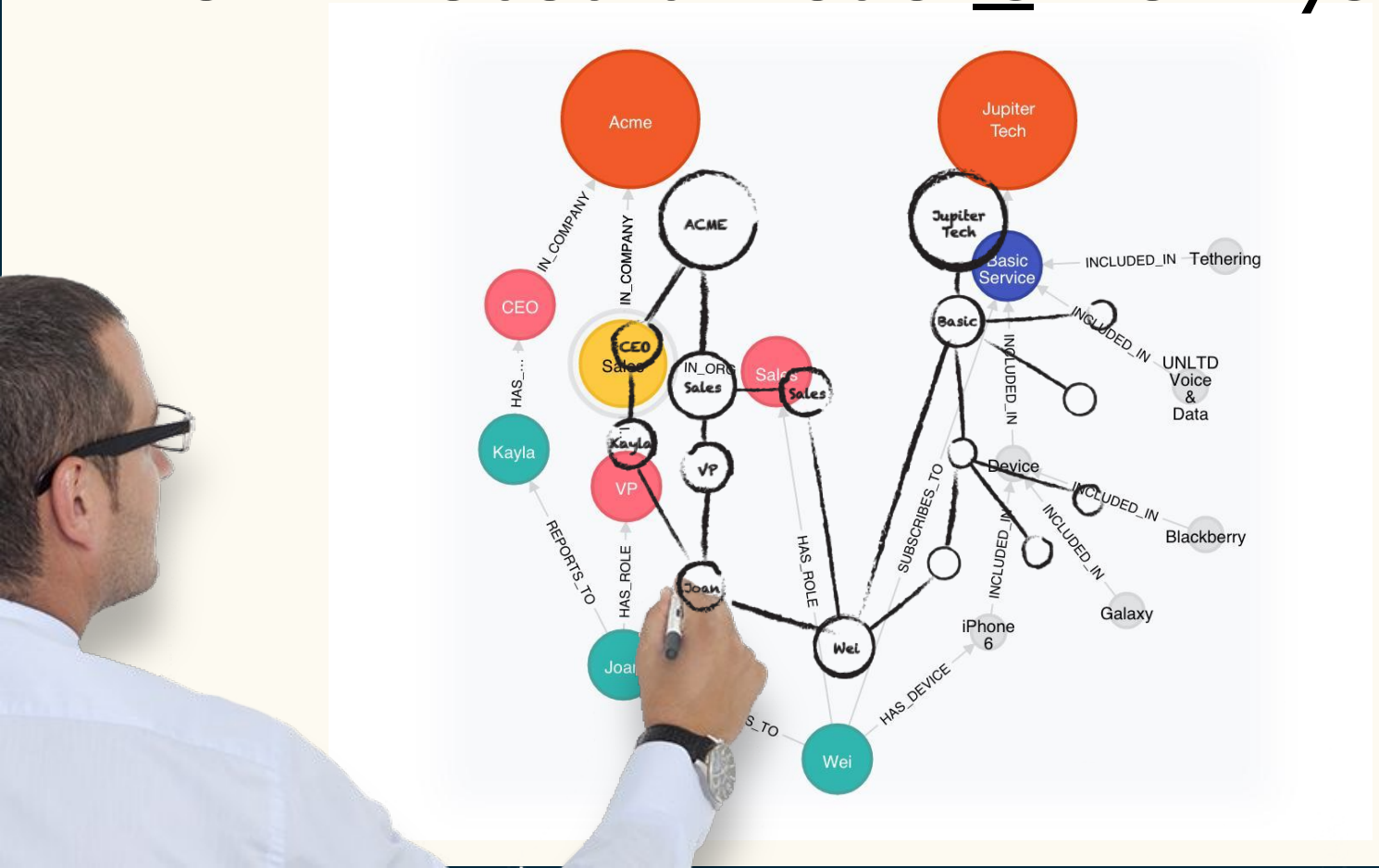
RDBMS Model



Graph Model



The Whiteboard Model Is the Physical Model



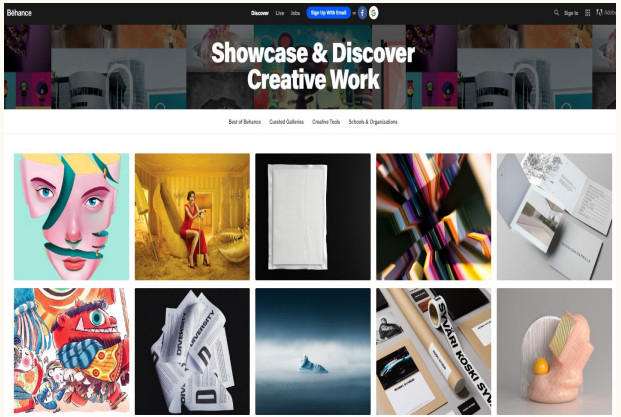
The Graph Problem Problem

Many organizations don’t realize that they have a graph problem

Adobe

Adobe Behance

Social Network of 10M
Graphic Artists



Background

- Social network of 10M graphic artists
- Peer-to-peer evaluation of art and works-in-progress
- Job sourcing site for creatives
- Massive, millions of updates (reads & writes) to Activity Feed
- *150 Mongos to 48 Cassandras to 3 Neo4j's!*

Business Problem

- Artists subscribe, appreciate and curate "galleries" of works of their own and from other artists
- Activities Feed is how everyone receives updates
- 1st implementation was 150 MongoDB instances
- 2nd implementation shrunk to 48 Cassandras, but it was still too slow and required heavy IT overhead

Solution and Benefits

- 3rd implementation shrunk to 3 Neo4j instances
- Saved over $500k in annual AWS fees
- Reduced data footprint from 50TB to 40GB
- Significantly easier to introduce new features like, "New projects in your Network"

On Aura (SaaS) up to

512GB

Vertical scaling

Unified DB for analytical & operational workloads

Guaranteed

99.95%

Uptime SLA

100s

TB Graphs with Sharding

Graph is enterprise-grade

Any cloud. Any workload

Encrypted Data

SOC2 Type 2

HIPAA Compliant

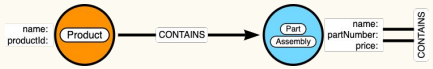
65+

Graph Algorithms

Integrations with

What is the Value Using a Native Graph Database?

How data is stored on disk



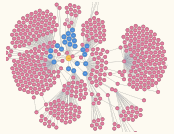
- Data is already stored as connected
- Data is efficiently placed on-disk using “Index-free Adjacency”
- A GDB stores **Nodes** and **Relations** instead of Rows & Columns
- **Semantics** may be built into the data!

How to query the data

```
(:Product)-[:CONTAINS]->(:Part)
```

- Cypher Query Language vs. SQL (ISO -> GQL)
- Simple, lesser lines of code, easier to read and maintain
- Queries possible, that span 100+ Hops in the Graph, which is comparable with SQL Joins over 100+ Tables!

Storing complex data networks & semantics

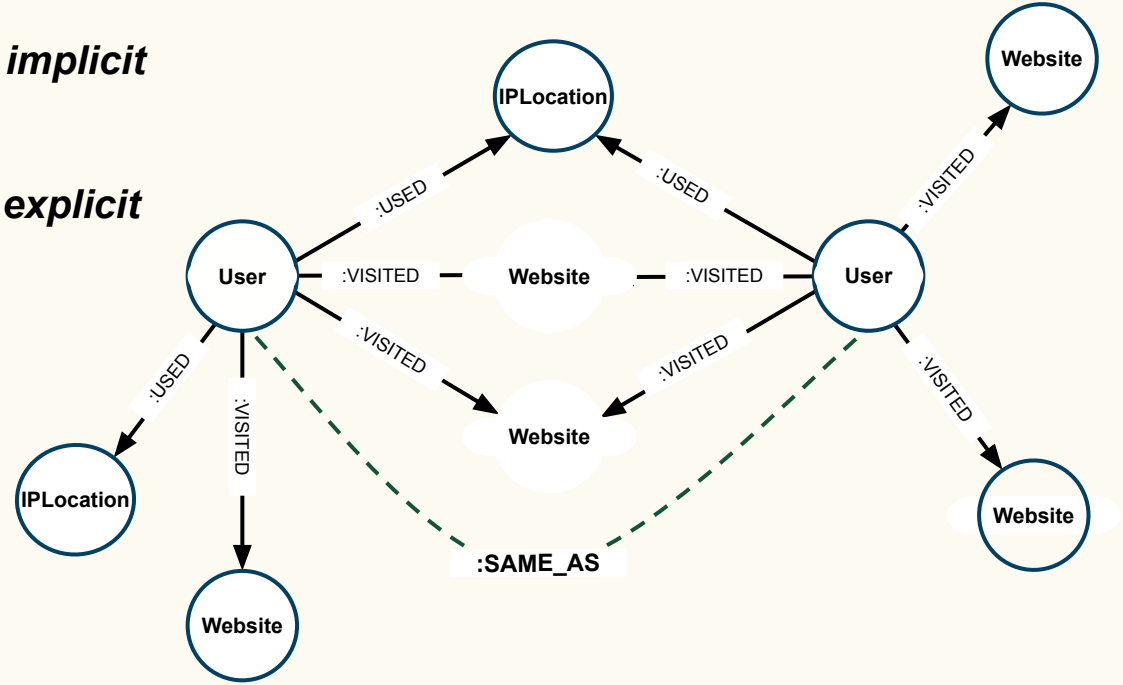


- Storing and analyzing complex relations between data
- Analyzing data from traditionally siloed data sources
- Extendable with data science algorithms and AI

Graphs....Grow!

Graphs allows you to make implicit relationships....

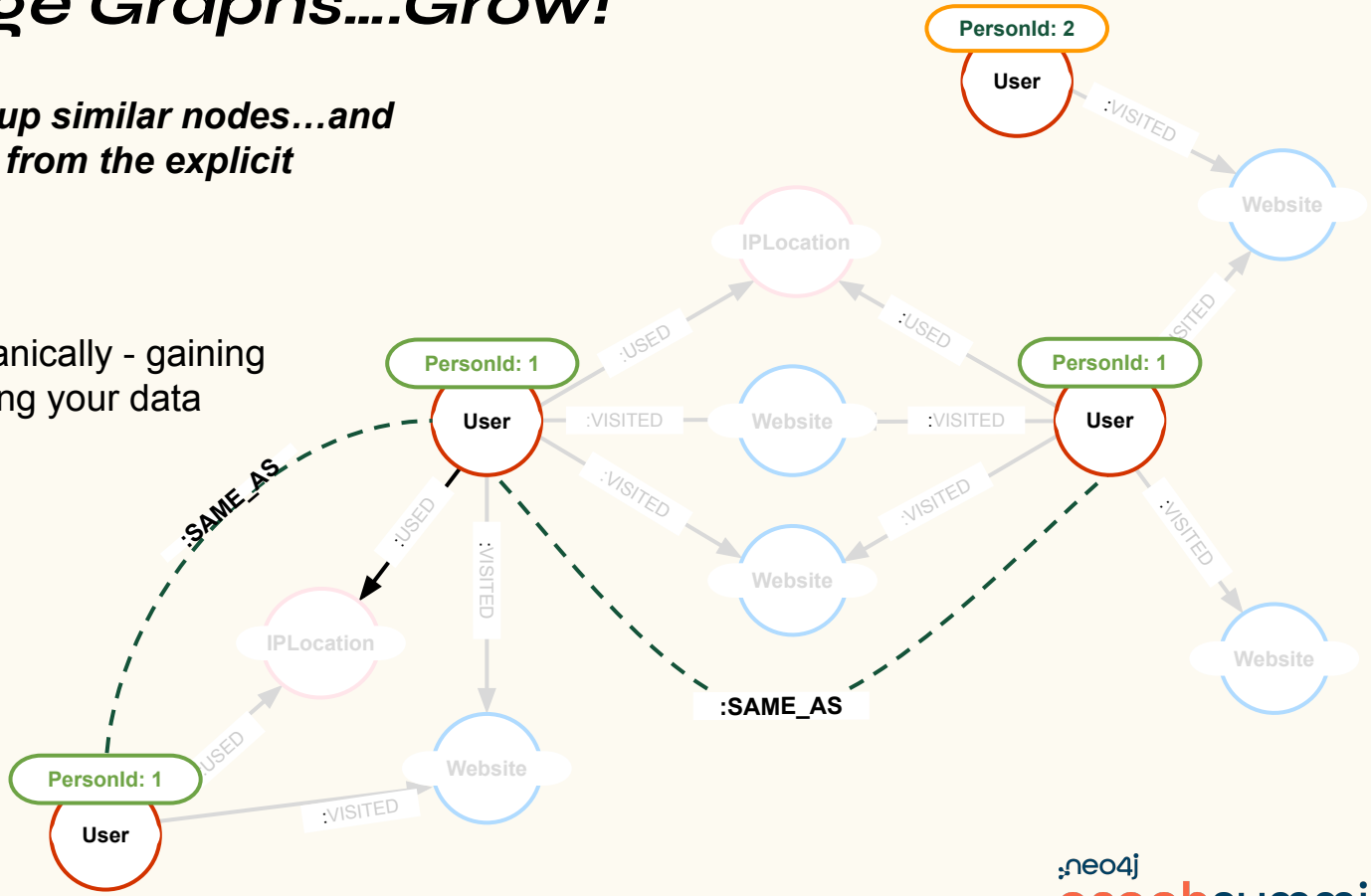
....explicit



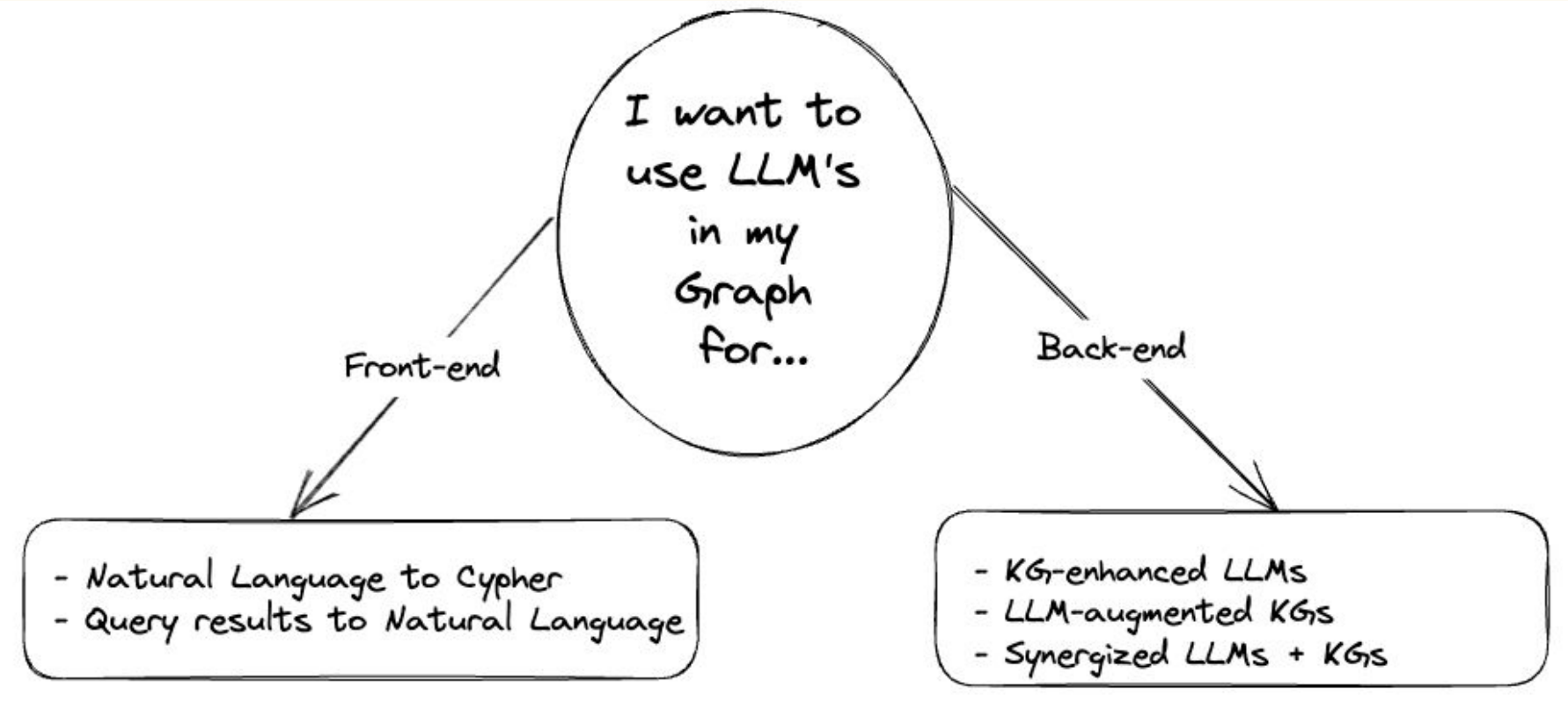
Knowledge Graphs….Grow!

…and can then group similar nodes…and create a new graph from the explicit relationships…

A graph **grows** organically - gaining insights and enriching your data

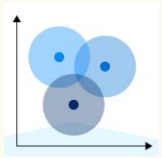


Setting the context - From a user point of view



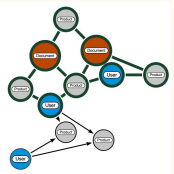
Semantic Search Journey

Vector Similarity Search



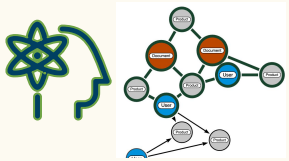
Find relevant documents and content for user queries

Graph Traversals & Pattern Matching



Find people, places, and things associated to content. Identify patterns in connected data.

Knowledge Graph Inference & ML



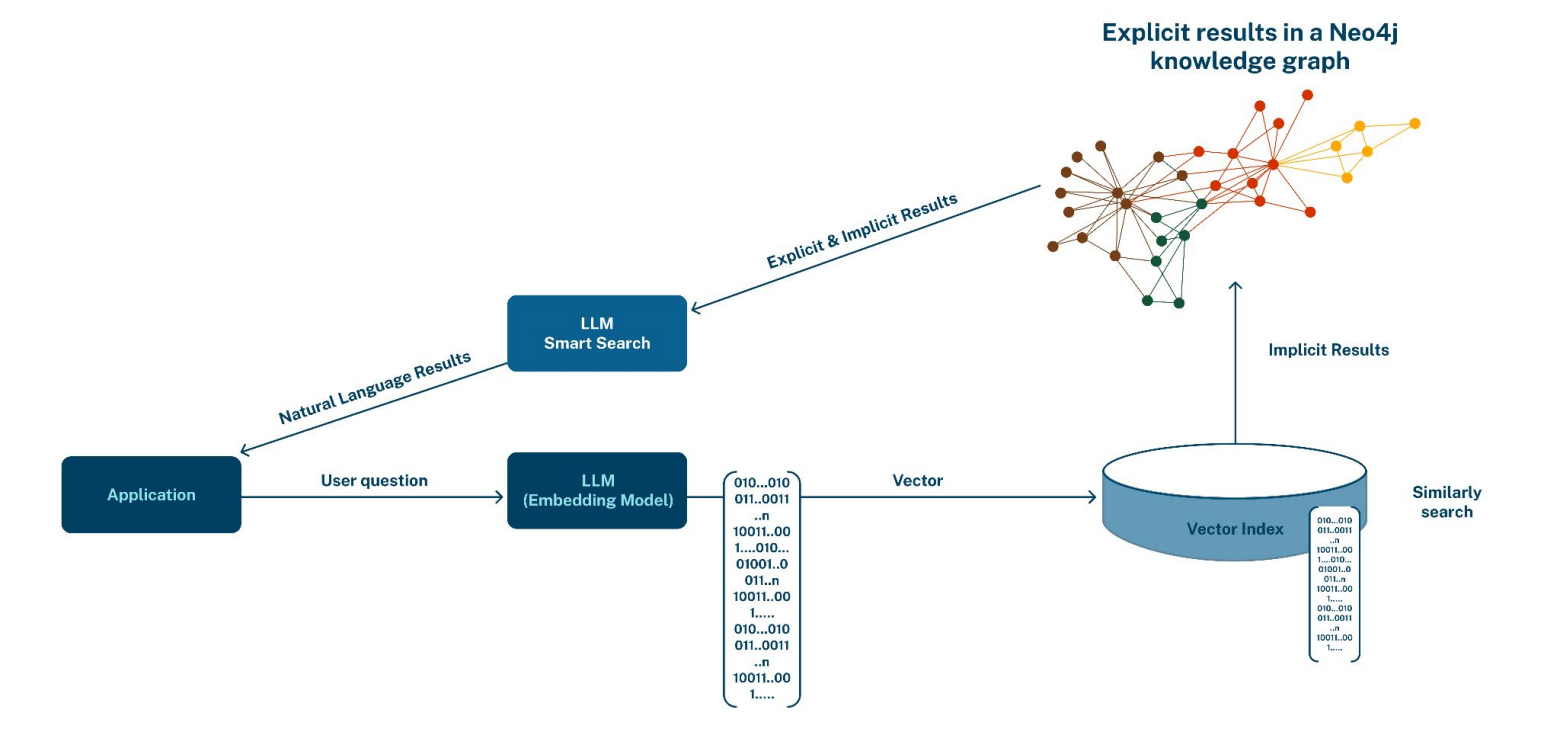
Further improve search relevance and insights by enhancing your Knowledge Graph.
Use graph algorithms and ML to discover new relationships, entities, and groups.

Vector Database

Graph Database

Neo4j

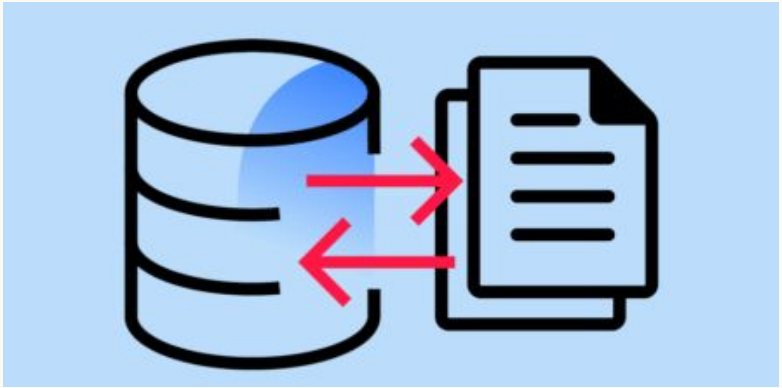
Knowledge Retrieval with Neo4j



Data Insertion & Loading

Data Insertion & Loading into Neo4j

- >>neo4j-admin import Command
- **Cypher CSV Loader**
- APOC
- ETL Tools (Apache Hop,…)
- Driver Connections
- 3rd Party Platforms (Kafka, Spark,Snowflake, BigQuery…)

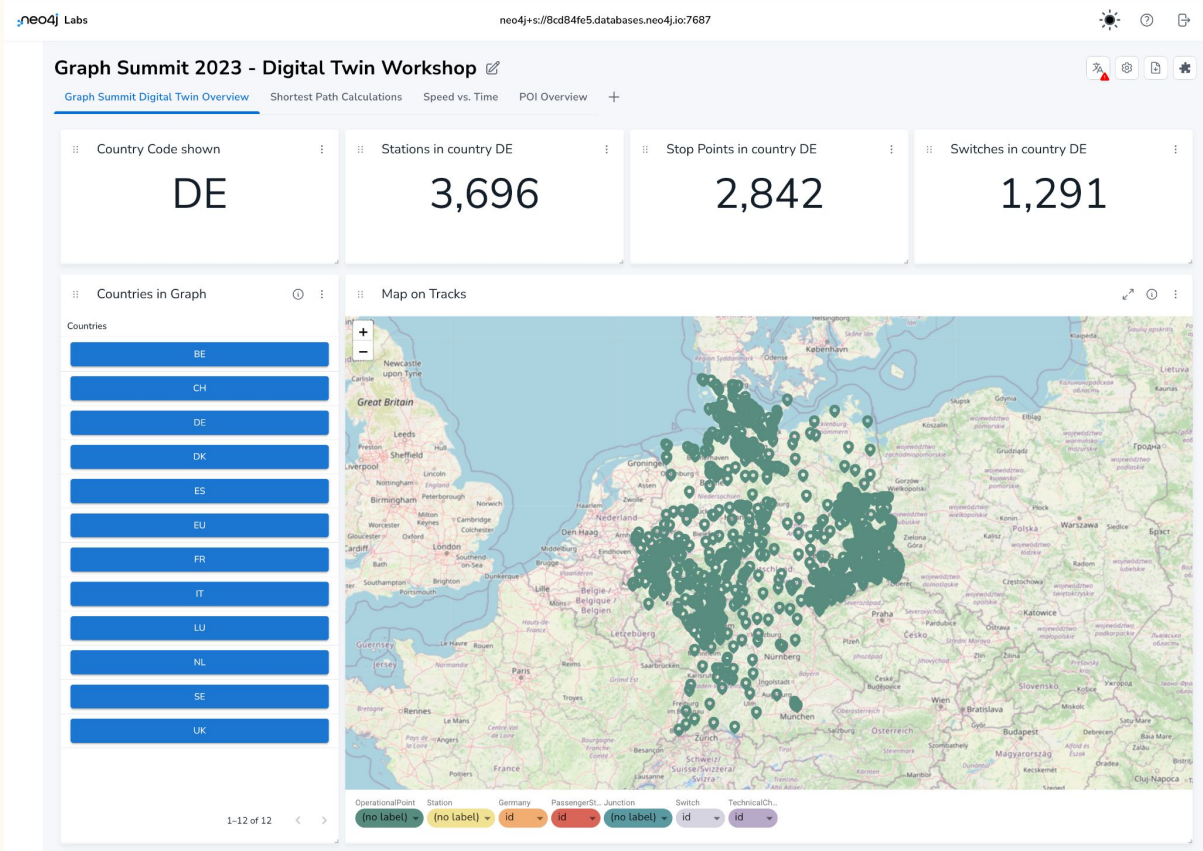


More about **Data loading, etc.** on the Neo4j Documentation

Data Visualization

Neo4j Explore - Neo4j Visualisation Platform (Bloom)

NeoDash - Dashboarding with Your Graph Data



Use Case Explanation

Digital Twin - An Overview

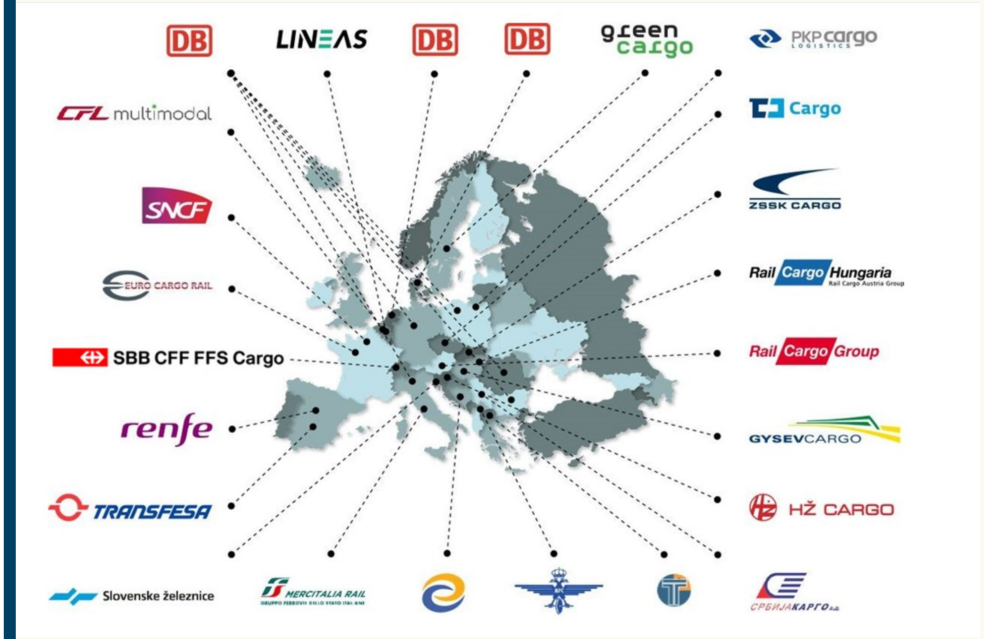
What is a Digital Twin?

A Digital Twin is a digital representation of an intended or actual real-world physical product, system, or process (a *physical twin*) that serves as the effectively indistinguishable digital counterpart of it for practical purposes, such as simulation, integration, testing, monitoring and maintenance.

* It has been done before

EU Railway Network - Track & Trace

- Challenge: **Legacy technology could not** track and analyze train journeys
- Solution: **Neo4j Knowledge Graph**
- **Identify and avoid bottlenecks (DB)**



Why do we need a Digital Twin? (few reasons)

Improved efficiency

It can optimize its operations and reduce costs by simulating different scenarios and making data-driven decisions.

Enhanced safety

It can identify potential hazards and test safety measures to improve safety for passengers and employees.

Predictive maintenance

It can monitor asset condition in real-time, predict maintenance needs, and increase asset lifespan.

Improved customer experience

A digital twin can simulate disruptions and help proactively address issues to enhance the customer experience and increase satisfaction.

Another reason why a digital twin might be helpful

https://www.euronews.com/travel/2023/02/21/unspeakable-botch-spain-spends-258-million-on-trains-that-are-too-big-for-its-tunnels

TRAVEL NEWS

‘Unspeakable botch’: Spain spends €258 million on trains that are too big for its tunnels

Two senior officials from Renfe and Adif have been dismissed over the error. – Copyright Canva

By **Angela Symons** • Updated: 21/02/2023

Spain has spent €258 million on **trains** that are too big to fit in its rail network’s tunnels.

After the blunder was exposed by local newspaper El Comercio in late January, two **transport** bosses were fired.

POI's Included

Unleash the power of the Graph

Abstracting from this Use Case

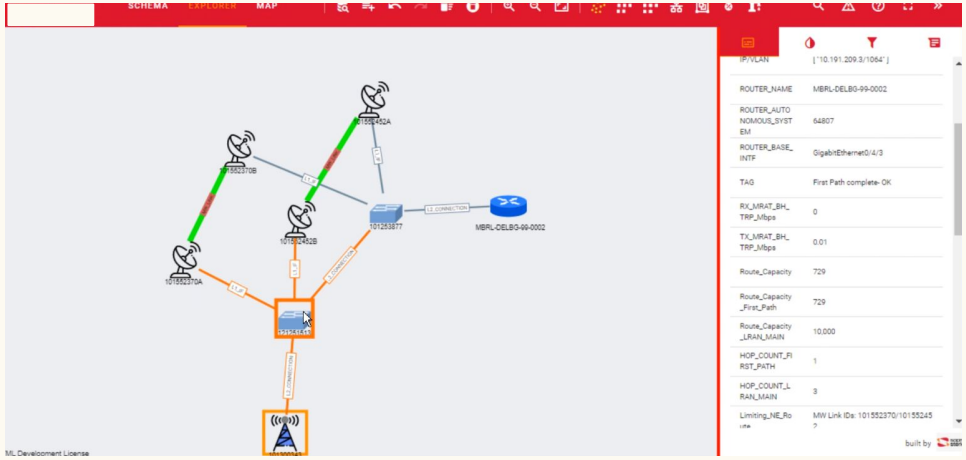
Digital Twin and Networks are everywhere

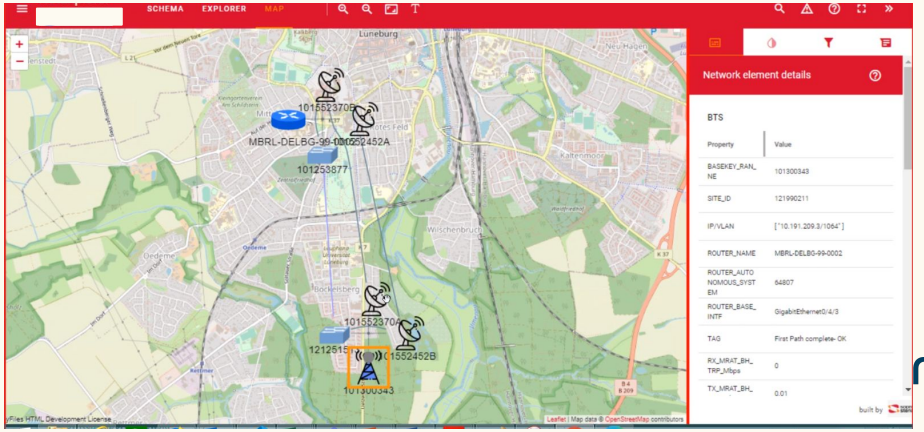
Many other use cases are networks or Digital Twins, too!

- Supply Chain
- Mobile Phones Networks
- Power Grids / Gas Grids / Fibre Networks
- Social Networks
- Patient Journeys
- Bill of Materials (BoM)
- Variants Management Graphs (Hard- & Software BoM combined)
- etc.

Mobile Network Operations - Large Mobile Provider

- **Digital Twin of Mobile Network**
- **Planning and operation**
  - Low signal strength of tower
  - Moving equipment around
- **Simulation of possible failure scenarios**
- **Repair & maintenance support**
  - Supply Chain Impact





"We wanted to create a solution that exploits artificial intelligence, integrates current and historical AIS positions as well as multiple data feeds and APIs. Such an effort would require a world-class infrastructure. That's why we selected Neo4j."

David Levy
Chief Marketing Officer
OrbitMI

Customer Case Study: Logistics and Supply Chain

Plan maritime routes based on distances, costs, and internal logic.

Results:

- **Subsecond** maritime routes planning
- Reduce global carbon emissions

  60,000 tons
- **12-16M ROI** for OrbitMI customers

OrbitMI

neo4j graphsummit

The Sample Dataset

Sample Data: Railway Network Digital Twin Dataset

Provided Data:

- **Consists of a Railway Track Network + Operation Point (OP)**
- **It includes attributes like:**
  - Tracks called Sections,
  - Stations and other OPs,
  - Geo location information,
  - Section length and speed,
  - other parameters

- **Add on from us: Point of Interest (POI) along tracks**

- **Format: CSV → Generated from XML**
- **Origin of Data: https://data-interop.era.europa.eu/search**

The Data Explained - In an Easy Way

“The data can be seen as a highway that has ramps throughout its entire path.

Each ramp can be an Operation Point (OP) like a Station, a Switch, etc.

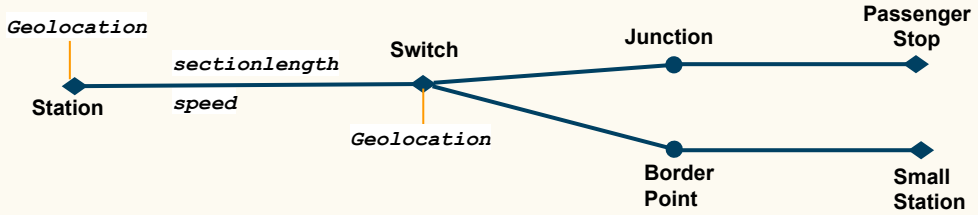
Tracks have a length and a speed associated. Tracks have a start point and an endpoint and will be inter-connecting Operation Points.

An Operation Point has a name and is referred to by a relationship with a country code.

Operation Points (OP) - Data Explanation (Nodes)

CSV Header titles:

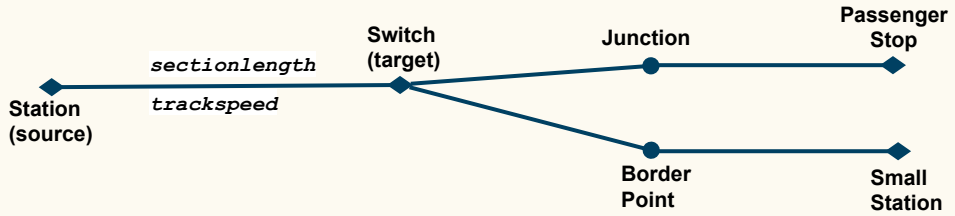
- `id: the internal number of the OP`
- `extralabel: the kind of OP we deal with, e.g. Station, Junction, Switch, etc.`
- `name: the name of a OP`
- `latitude: of the OP`
- `longtitude: of the OP`



Section Connection Data Explanation (Relationships)

CSV Header titles:

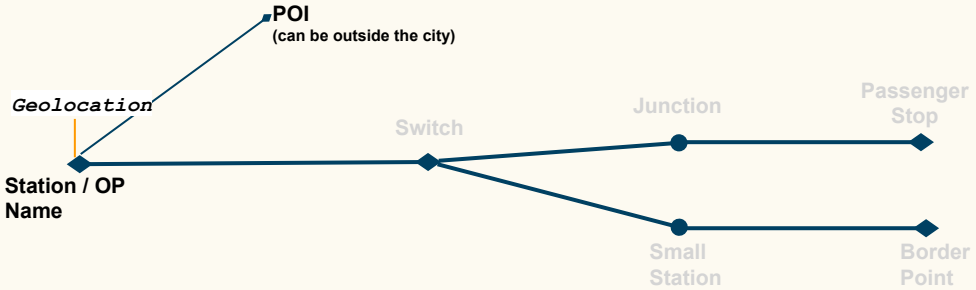
- `source: start OP for this section`
- `target: end OP for this section`
- `sectionlength: the length in km of that section`
- `trackspeed: max speed allowed on that section`



Point of Interest (POI) Data Explanation

CSV Header titles:

- `CITY: City the POI is in or close by`
- `POI_DESCRIPTION: A short description of of the POI`
- `LINK_FOTO: a short name`
- `LINK_WEBSITE: is the name of the track corresponding to the shortcut`
- `LAT: Latitude of the POI`
- `LONG: Longitude of the POI`
- `SECRET: True if not well know, False if well known (e.g. Berlin)`



Data Modeling - Writing down Questions & Answers

Data Modeling - The Questions are the Input

1. **How long is the journey in KM from station X to station Y?**
2. **How long in KM is an alternative route if a station on my journey is closed?**
3. **Do I have alternative routes if a station in my journey is down?**
4. **What is the *shortest distance* in KM between stations X and Y?**
5. **Can I geolocate stations in my static rail network?**
6. **What are the stations expecting the most traffic?**
7. **What POIs are along my route?**

Note: Aggregations are also possible, but we focus on the “graphy” queries here!

Expected Answers

1. Your journey is 30 stations and 239km long
2. Your alternative route is 36 stations and 271km long
3. The shortest route between X and Y is XXX km
4. The geo location of your station is Lat / Long
5. The station with the most traffic is … ?
6. The following POIs are along your route …

Data Modeling
Building the Model

Modeling workflow example

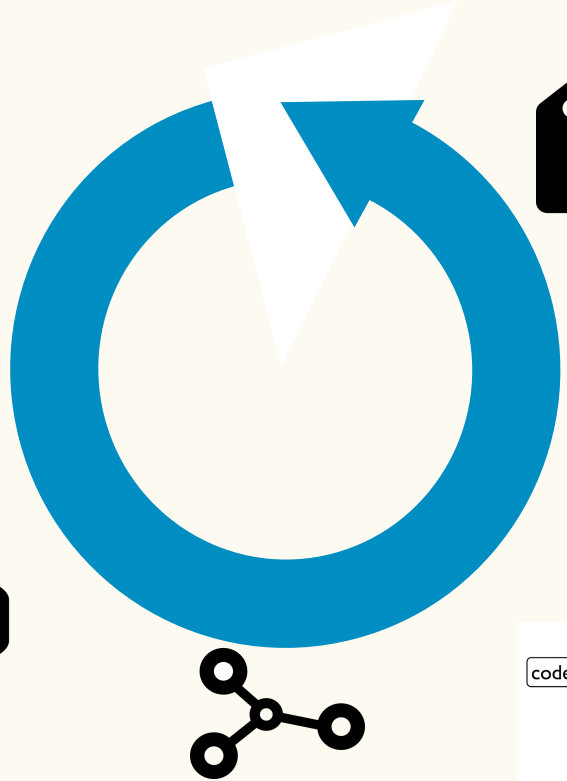
As an air travel enthusiast

I want to know how airports are connected

So that I can find the busiest ones

Is **Airport** A connected to **Airport** B?

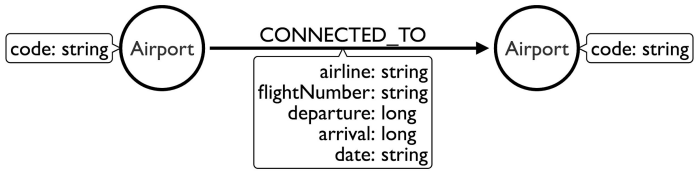
```
|---------+------+------------|
|  Origin | Dest | FlightNum  |
|---------+------+------------|
|  IAD    | TPA  | 335        |
|  IAD    | TPA  | 3231       |
|  IND    | BWI  | 448        |
|  IND    | BWI  | 3920       |
|---------+------+------------|
```



```
(airport)-[:CONNECTED_TO]->(airport)
```

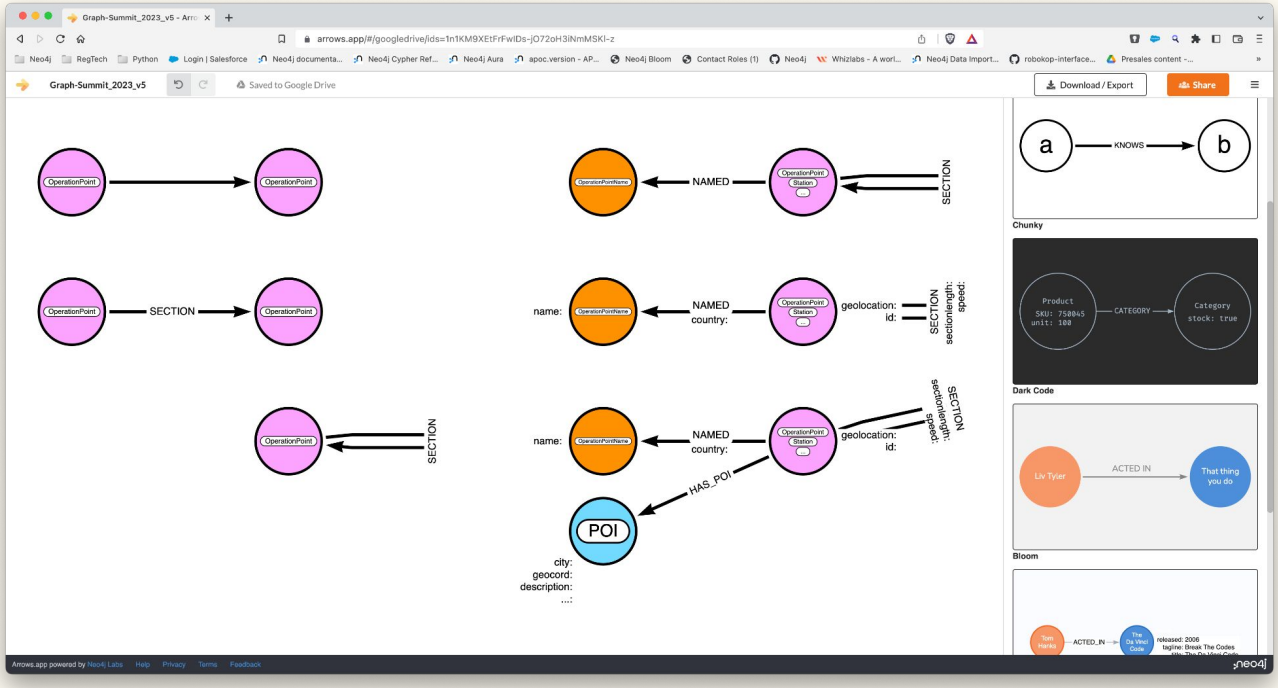
```
MATCH (origin:Airport {code: "LAX"})
      -[flight:CONNECTED_TO]->
      (destination:Airport {code: "LAS"})
RETURN origin, destination, flight
LIMIT 10
```

```
CREATE (lax:Airport {code: "LAX"})
CREATE (las:Airport {code: "LAS"})
CREATE (las)-[connection:CONNECTED_TO {
    airline: "WN",
    flightNumber: "82",
    date: "2008-1-3",
    departure: 1715,
    arrival: 1820}]->(lax)
```



Tools for Graph Data Modeling

Arrows Tool → https://arrows.app



Data Modeling Using the Questions as Input

Remember those questions?

1. How long is the journey in KM from station X to station Y?
2. How long in KM is an alternative route if a station on my journey is closed?
3. Do I have alternative routes if a station in my journey is down?
4. What is the *shortest distance* in KM between stations X and Y?
5. Can I geolocate stations in my static rail network?
6. What are the stations expecting the most traffic?
7. What POIs are along my route?

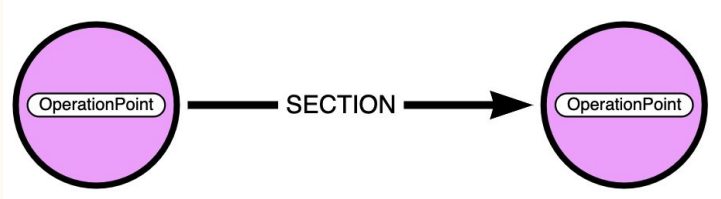
Blue → Nodes / Properties? **Orange** → Relationship Types

Building a First Data Model 1/4

- Nouns in your questions are the nodes, verbs are the relationships
- Nodes have a Label, Relationships MUST have a Type
- Properties should help to uniquely identify Nodes and/or answer questions

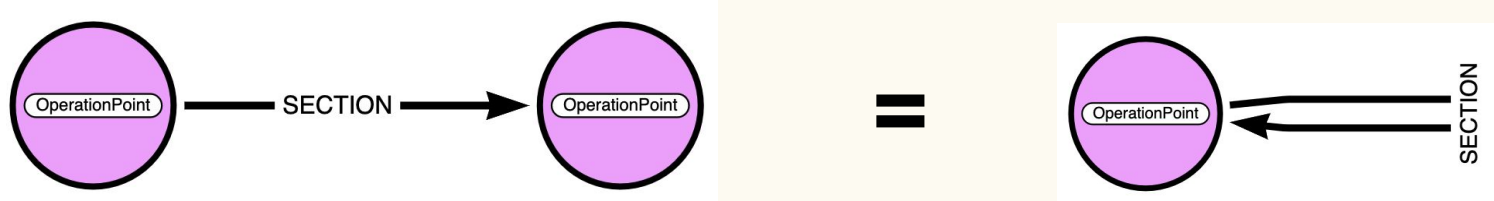
Building a First Data Model 2/4

- Nouns in your questions are the nodes, verbs are the relationships
- Nodes have a Label, Relationships MUST have a Type
- Properties should help to uniquely identify Nodes and/or answer questions



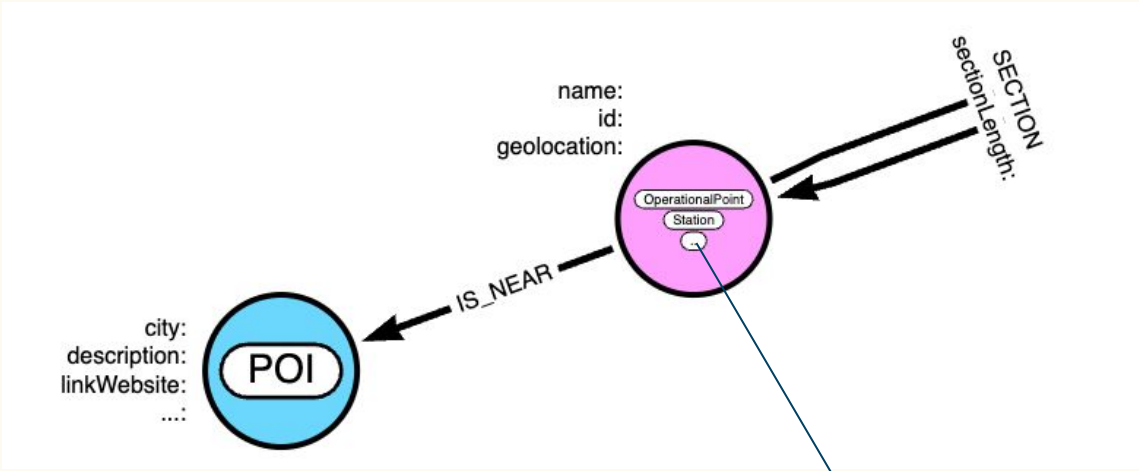
Building a First Data Model 3/4

- Nouns in your questions are the nodes, verbs are the relationships
- Nodes have a Label, Relationships MUST have a Type
- Properties should help to uniquely identify Nodes and/or answer questions



Building a First Data Model 4/4

- Nouns in your questions are the nodes, verbs are the relationships
- Nodes have a Label, Relationships MUST have a Type
- Properties should help to uniquely identify Nodes and/or answer questions

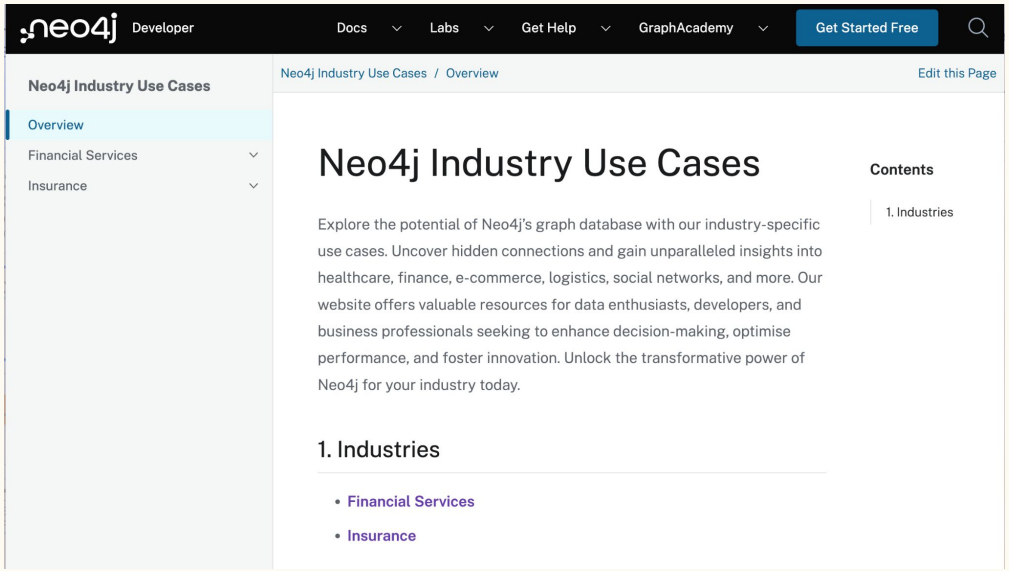


We will add more Labels to this node, eg. Switch, BorderPoint, etc.

You DON'T have to do it again and again!

Use Case Template repository*, that is growing

https://neo4j.com/developer/industry-use-cases/finserv/



*** And we have more templates to contribute**

Workshop Time

Let’s build the solution together ...

Let's Do it Together …

1. Create a Neo4j Graph instance via any of:
   a. Training db:
      https://workspace.neo4j.io
      neo4j+s:// gsams25.graphdatabase.ninja:443
      Ask for a username/login

   b. **Neo4j Aura:**
      https://console.neo4j.io/
      i. You can spin up a **AuraDB Free** aura instance here. If you already have one Aura Free db in use and don't want to clear it use an other option for this workshop
      ii. You can spin up a AuraDB Professional instance in a **free** trial
      Use trial, 4gb, and enable Graph Analytics

2. Open the Github page at: https://github.com/kvegter/gsummit2025
   a. **Open up the file:** `cypher/load-all-data.cypher`
   b. **Copy and paste the file contents into your Neo4j browser/Neo4j Query.**

.

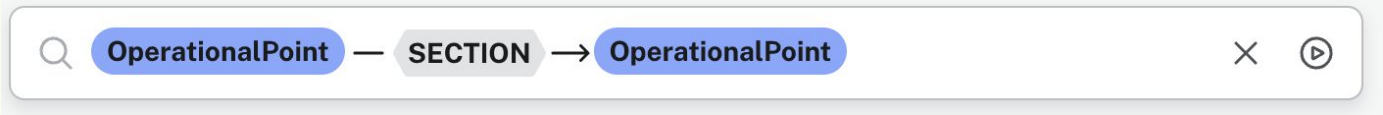
Explore the data

Open the Neo4j Browser/ Neo4j Query and check the data:

- Click on the SECTION relationship
- Set the limit on 1000

Open if available Neo4j Bloom/Neo4j Explore

- Search for (OperationalPoint)-[SECTION]->(OperationalPoint)



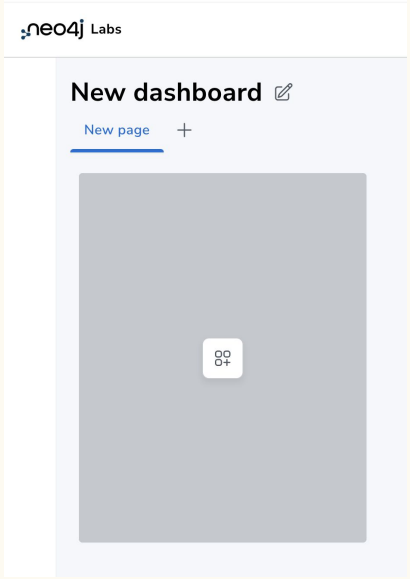
Bonus Use the Betweenness Algorithm

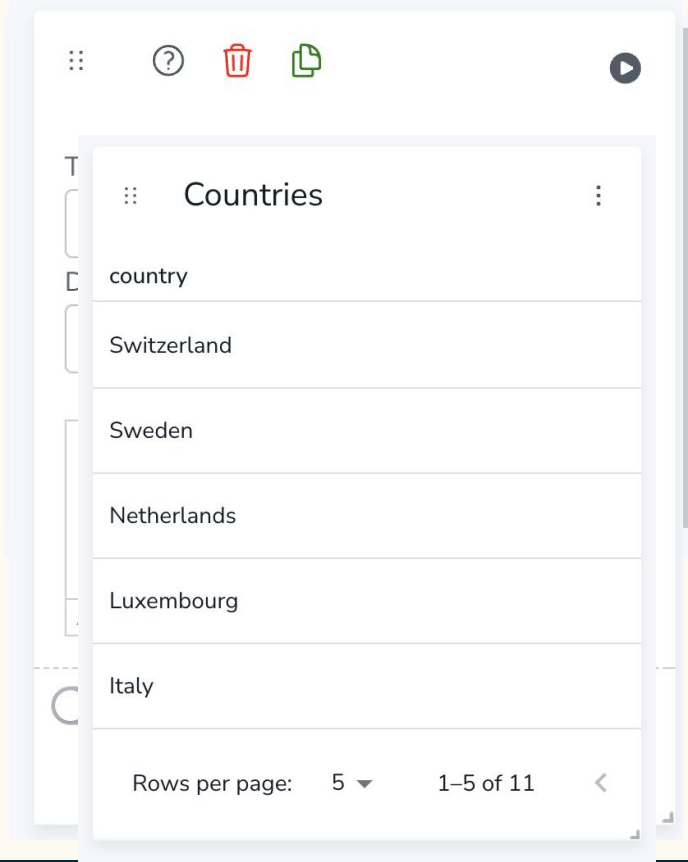
NeoDash

Open https://neodash.graphapp.io/ when you have a sandbox or aura db

Import the dashboard from GitHub

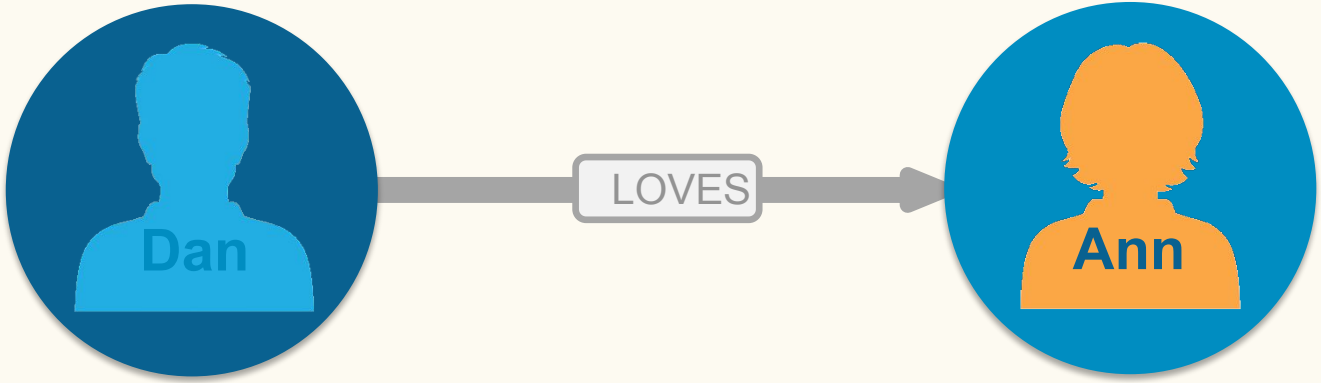
NeoDash First Component






```
WITH ["Switzerland"
,"Sweden"
,"Netherlands"
,"Luxembourg"
,"Italy"
,"France"
,"Spain"
,"Denmark"
,"Germany"
,"Belgium"
,"UK"] as countryList
UNWIND countryList as country
RETURN country
```

Cypher: Powerful and Expressive Query Language



NODE RELATIONSHIP NODE

```
CREATE (:Person { name:“Dan”} ) -[:LOVES]-> (:Person { name:“Ann”} )
```

LABEL PROPERTY LABEL PROPERTY

MATCH <PATTERN> RETURN <DATA>

Cypher

```
MATCH (u:Customer
{customer_id:'customer-one'})-[:BOUGHT]->(p:Product)<-
[:BOUGHT]-(peer:Customer)-[:BOUGHT]->(reco:Product)
WHERE not (u)-[:BOUGHT]->(reco)
RETURN reco as Recommendation, count(*) as Frequency
ORDER BY Frequency DESC LIMIT 5;
```

SQL

```
SELECT product.product_name AS Recommendation,
count(1) AS Frequency
FROM product, customer_product_mapping, (SELECT
cpm3.product_id, cpm3.customer_id
FROM Customer_product_mapping cpm,
Customer_product_mapping cpm2,
Customer_product_mapping cpm3
WHERE cpm.customer_id = 'customer-one'
AND cpm.product_id = cpm2.product_id
AND cpm2.customer_id != 'customer-one'
AND cpm3.customer_id = cpm2.customer_id
AND cpm3.product_id NOT IN (SELECT DISTINCT
product_id
FROM Customer_product_mapping cpm
WHERE cpm.customer_id = 'customer-one')
) recommended_products
WHERE customer_product_mapping.product_id =
product.product_id
AND customer_product_mapping.product_id IN
recommended_products.product_id
AND customer_product_mapping.customer_id =
recommended_products.customer_id
GROUP BY product.product_name
ORDER BY Frequency DESC
```

Cypher Check Up

```
// What is this Query Doing?
MATCH (a:Person)-[:FRIEND_OF]->(b:Person)<-[:FRIEND_OF]-(c:Person)
WHERE a.name = “Niels de Jong”
AND c.name = “Kees Vegter“
RETURN b.name as Name, b.id as FriendID
```

Cypher Hands-on Questions

```
// query for all the nodes with the label "Netherlands"
```

Cypher Hands-on Questions

```
// query for all the nodes with the label "Netherlands"
// In the Neo4j Browser/Query

MATCH (n:Netherlands)
RETURN n
```

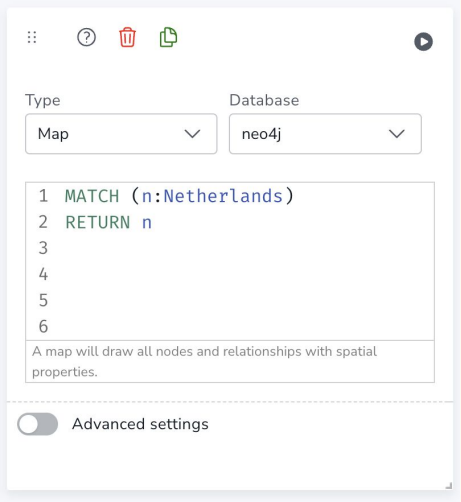
Cypher Hands-on : NeoDash

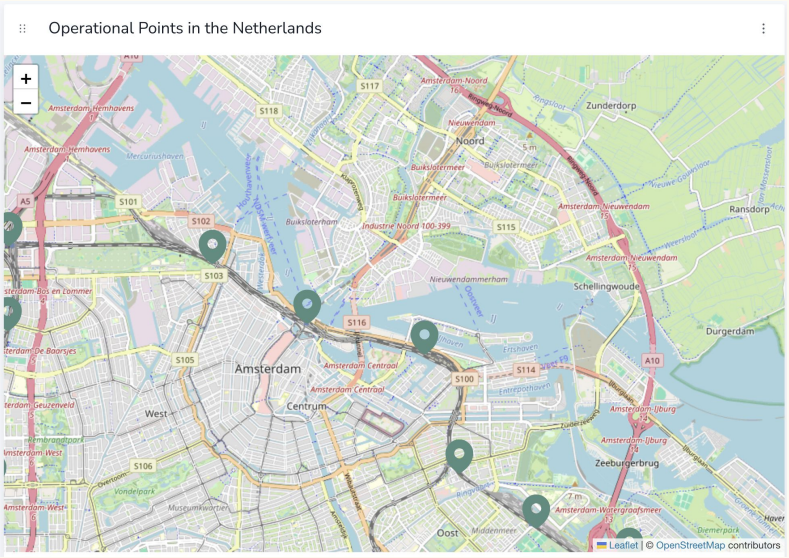
```
// query for all the nodes with the label "Netherlands"
// In NeoDash, use Type "Map"

MATCH (n:Netherlands)
RETURN n
```

Cypher Hands-on : NeoDash

- query for all the nodes with the label "Netherlands"
- In NeoDash create a new Component and use Type "Map"
- Resize the component





NeoDash parameters

It is possible to link Components together via NeoDash parameters:

- Enable Extensions via:  enable: Report Actions
- On the country List Component click on the three dots and click on 
- Add a new action and fill it in like this

Now a parameter $neodash_country can be used in other components

NeoDash parameters

The Operational Point Component can now use the **$neodash_country** parameter:

- Operational Point Component:
  click on the title and replace "the Netherlands" with $neodash_country

- Operational Point Component:
  On the country List Component click on the three dots and change the cypher query with a dynamic label reference

  ```
  MATCH (n:$($neodash_country))
  RETURN n
  ```

Cypher Hands-on Questions

```
// Find the 50 Operation Points, Ordered By Neo4j internal id's
```

```
// Find All Operation Points Names in the city of Amsterdam
```

Cypher Hands-on Questions

```
// Find the 50 Operation Points, Ordered By Neo4j internal id's
MATCH (op:OperationalPoint)
RETURN op
ORDER BY id(op)
LIMIT 50;

// Find All Operation Points Names in the city of Amsterdam
```

Cypher Hands-on Questions

```
// Find the 50 Operation Points, Ordered By Neo4j internal id's
MATCH (op:OperationalPoint)
RETURN op
ORDER BY id(op)
LIMIT 50;

// Find All Operation Points in the city of Amsterdam
MATCH (op:OperationalPoint)
WHERE op.name CONTAINS 'Amsterdam'
RETURN op.name;
```

NeoDash

Make two components in Neodash:

- 1 Type "Parameter Select"-> Free Text -> parameter name '**opname**' ($neodash_opname)

- 2 Type "table" or "Graph" or Map

```
MATCH (op:OperationalPoint)
WHERE op.name CONTAINS $neodash_opname
RETURN op.name;
```

Cypher Hands-on Questions

```
// Find the Station id's for Berlin and Amsterdam
// HINT: Use the pattern (OperationalPoint)
```

Cypher Hands-on Questions

```
// Find the Station id's for Berlin and Amsterdam
// HINT: Use the pattern (OperationalPoint)


MATCH (n:OperationalPoint)
WHERE n.name STARTS WITH 'Berlin'
RETURN n.name, n.id

MATCH (n:OperationalPoint)
WHERE n.name STARTS WITH 'Amsterdam'
RETURN n.name, n.id
```

Cypher Hands-on Questions

```
// Find the Shortest path between Berlin and Amsterdam
// (Berlin-id: DE000BL, Amsterdam-id: NLASD)
TIP: use shortestPath
```

Cypher Hands-on Questions

```
// Find the Shortest path between Berlin and Amsterdam
// (Berlin-id: DE000BL, Amsterdam-id: NLASD)
TIP: use shortestPath

MATCH (op1:OperationalPoint WHERE op1.id = 'NLASD')
MATCH (op2:OperationalPoint WHERE op2.id = 'DE000BL')
MATCH sp=shortestPath((op1)-[:SECTION*]-(op2))
RETURN sp;
```

NeoDash

Make 5 components in Neodash:

- 1 Type "Parameter Select"-> Free Text -> parameter name '**fromname**' ($neodash_fromname)
- 2 Type "table"

```
MATCH (op:OperationalPoint)
WHERE lower(op.name) CONTAINS $neodash_fromname
RETURN op.name as From, op.id as ___fromid;
```

- 3 as 1 but now parameter 'toname'
- 4 as 2 but now parameter $neodash_toname and ___toid
- For 2 and 4 use one Report action each to fill the variables $neodash_fromid and $neodash_toid with the '___' hidden values

NeoDash

Make 5 components in Neodash:

- 5 Query component with Graph output

```
MATCH (a:OperationalPoint { id: $neodash_fromid})
,     (b:OperationalPoint { id: $neodash_toid})
RETURN shortestPath((a)-[*]-(b))
```

GraphSummit Amsterdam 2025

Home | Shortest Path | +

From

fromName

amsterdam centr

From Station

from

AMSTERDAM CENTRAAL

Rows per page: 5 1–1 of 1

To

toName

berlin hau

To Station

to

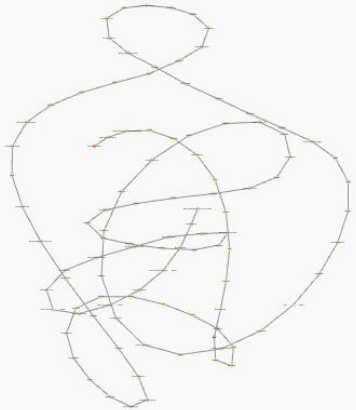
BERLIN HAUPTBAHNHOF - LEHRTER BF S-BAHN

BERLIN HAUPTBAHNHOF TUNNEL (S-BAHN)

BERLIN HAUPTBAHNHOF-LEHRTER BF (STADTB)

Rows per page: 5 1–4 of 4

Shortest Path



OperationalP... Netherlands PassengerTe... Station Junction SmallStation Switch BorderPoint Germany PassengerSt...

summit