neo4j graphsummit

Building a Graph Solution in Neo4j

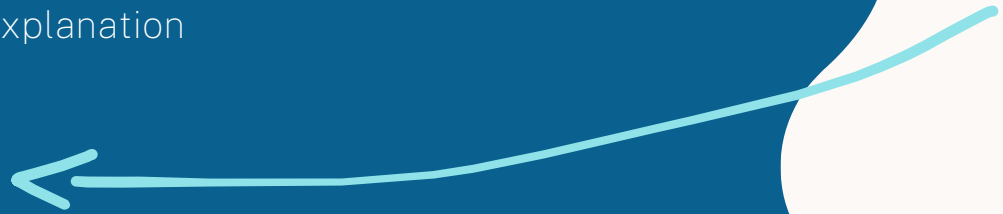
Agenda

- Logistics
- Introduction
- Use Case Explanation
- Modeling

- BREAK

- Building the solution
- Q & A

That's right!
Time for Tea/Coffee



Logistics

WIFI Access:	etcvenues / stayconnected
Restrooms:	In corridor to your right (my left)
Chargers/Sockets:	Under *some* of the desks
Material for the workshop:	https://github.com/cskardon/gsummit2023 
More training:	luke.masters@neo4j.com

neo4j
graphsummit

Introduction

What is a property graph?

Good news!

you only need to know

4

things

Graph components

Node (Vertex)

- The main data element from which graphs are constructed

Graph components

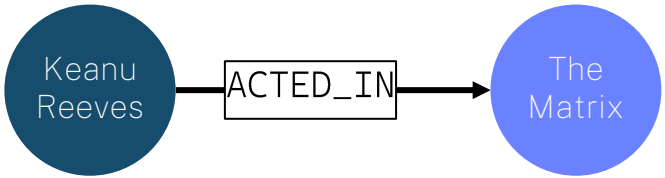
Node (Vertex)

- The main data element from which graphs are constructed

A node without relationships is permitted, a relationship without nodes is not

Relationship (Edge)

- A link between two nodes
  - Direction
  - Type



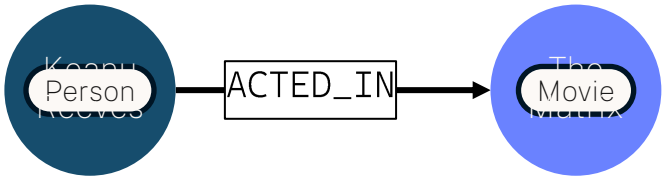
Property graph database

Node (Vertex)

Relationship (Edge)

Label

- Define node role (optional)



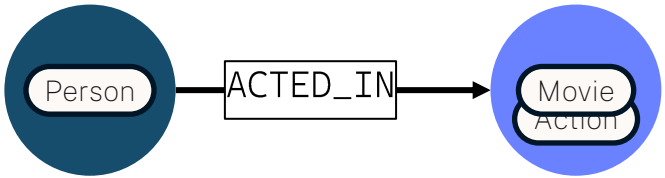
Property graph database

Node (Vertex)

Relationship (Edge)

Label

- Define node role (optional)
- Can have more than one



Property graph database

Node (Vertex)

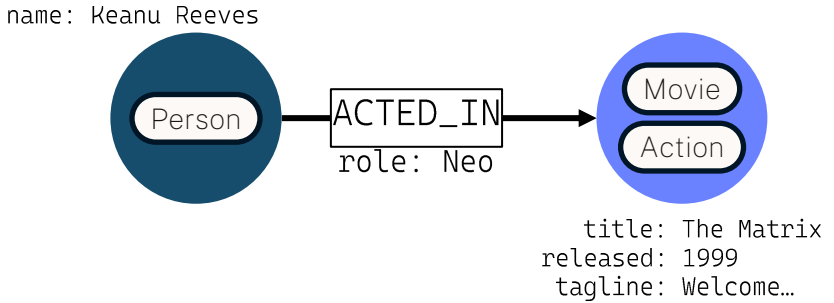
Relationship (Edge)

Label

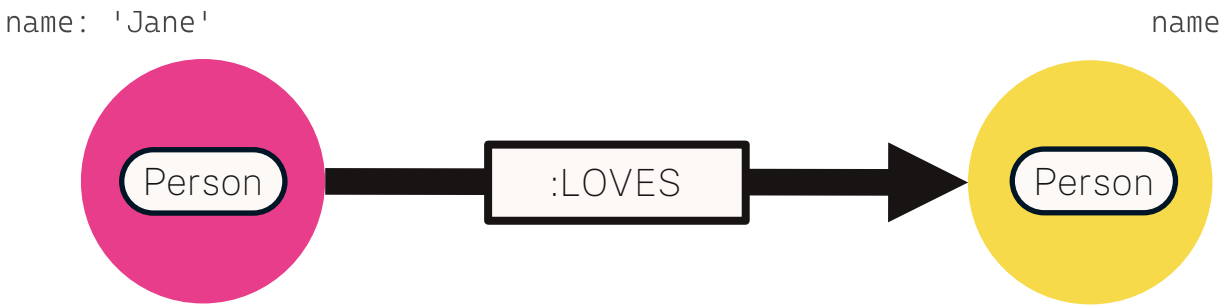
- Define node role (optional)
- Can have more than one

Properties

- Enrich
  - nodes
  - relationships



Cypher: powerful and expressive query language




```
                NODE                    RELATIONSHIP                NODE
CREATE (:Person {name: 'Jane'}) -[:LOVES]-> (:Person { name:  'Arlo' })
        Label    Property                    Label    Property
```

Cypher: Matching


```
MATCH (p:Person {name: 'Jane'}) -[:MARRIED_TO]-> (spouse:Person)
RETURN p, spouse
```

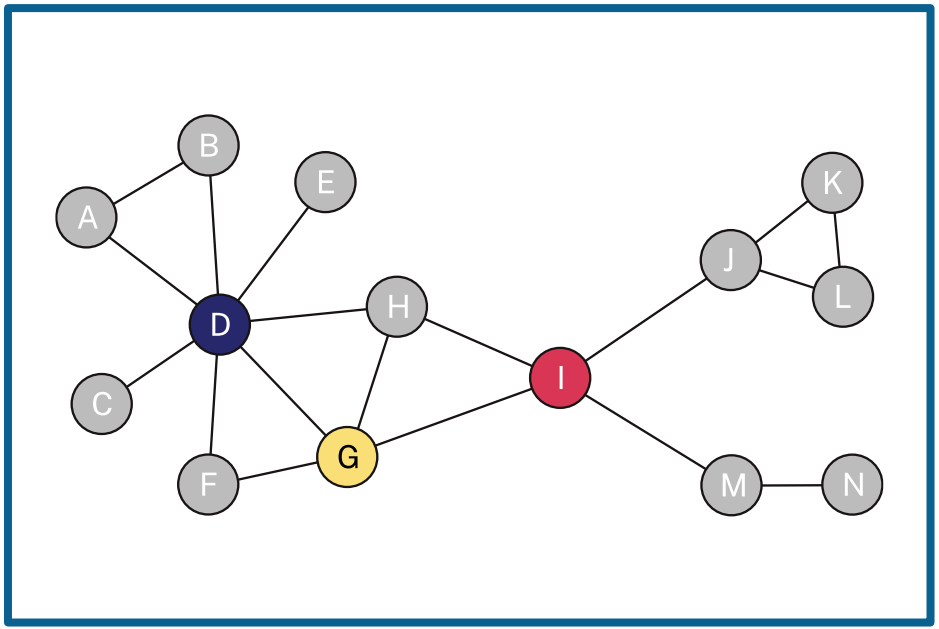
NODE — RELATIONSHIP — NODE

Variable — Label — Property — Variable

neo4j
graphsummit

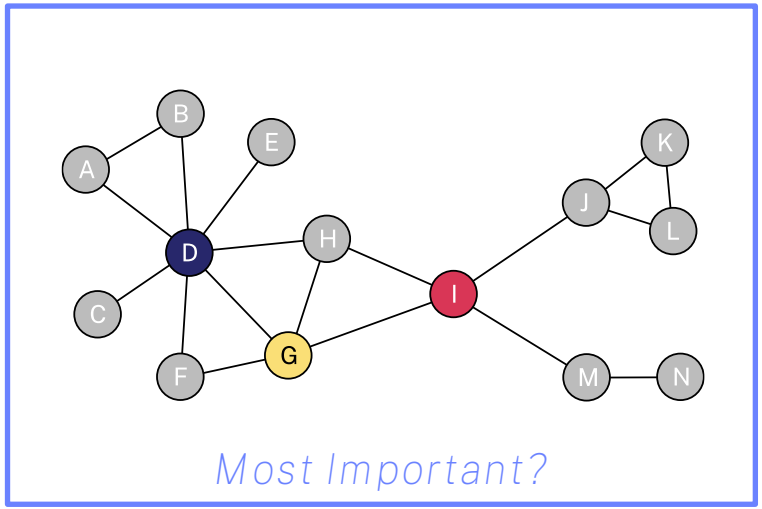
Neo4j Graph Data Science

Pop Quiz



Which of the coloured nodes would be considered the most "important"?

Graphs Contain Implicit Knowledge



D

D has the highest degree centrality (7)
This is the most connected individual in the network. If important is now well you are personally known, you pick D.

G

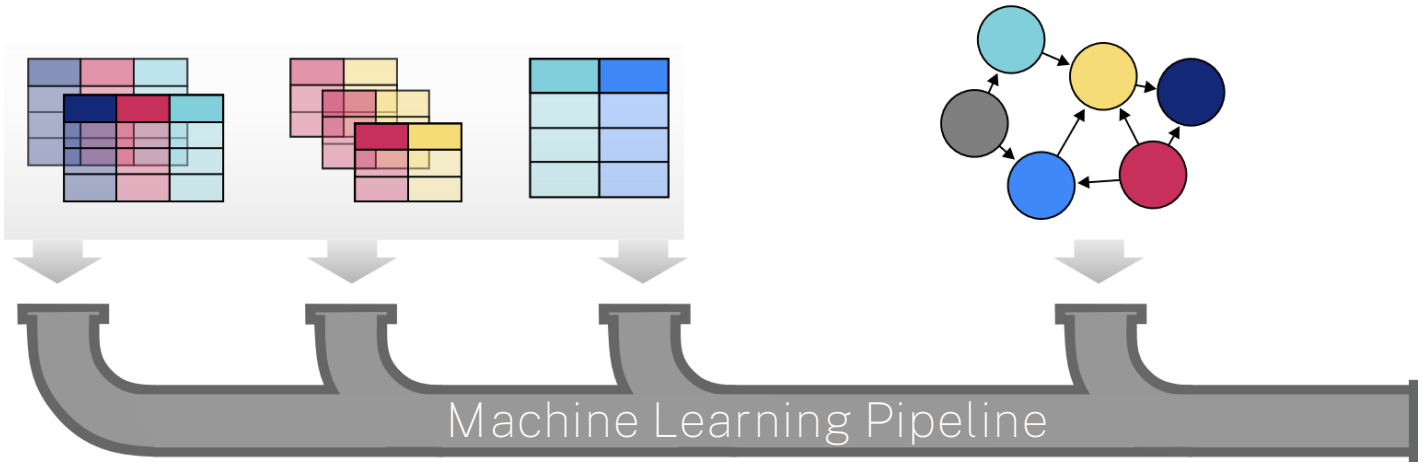
G has the highest closeness centrality (0.52)
Information will disperse through the network more quickly through this individual. If you need to get a message out rapidly, choose G.

I

I has the highest betweenness centrality (0.59)
This person is an efficient connector of other people. Risk of network disruption is higher if you lose this individual

Better Predictions With Data You Already Have

- Traditional ML ignores network structure because it's difficult to extract
- Graphs use relationships to unlock otherwise unattainable predictions
- Add graph data to existing ML pipelines to increase accuracy



Graph Algorithm Categories

Pathfinding & Search

Finds optimal paths or evaluates route availability and quality

Centrality / Importance

Determines the importance of distinct nodes in the network

Community Detection

Detects group clustering or partition

Link Prediction

Estimates the likelihood of nodes forming a future relationship

Similarity

Evaluates how alike nodes are by neighbours and relationships

Embeddings & ML

Compute low-dimensional vector representations of nodes in a graph, and allow you to train supervised machine learning models

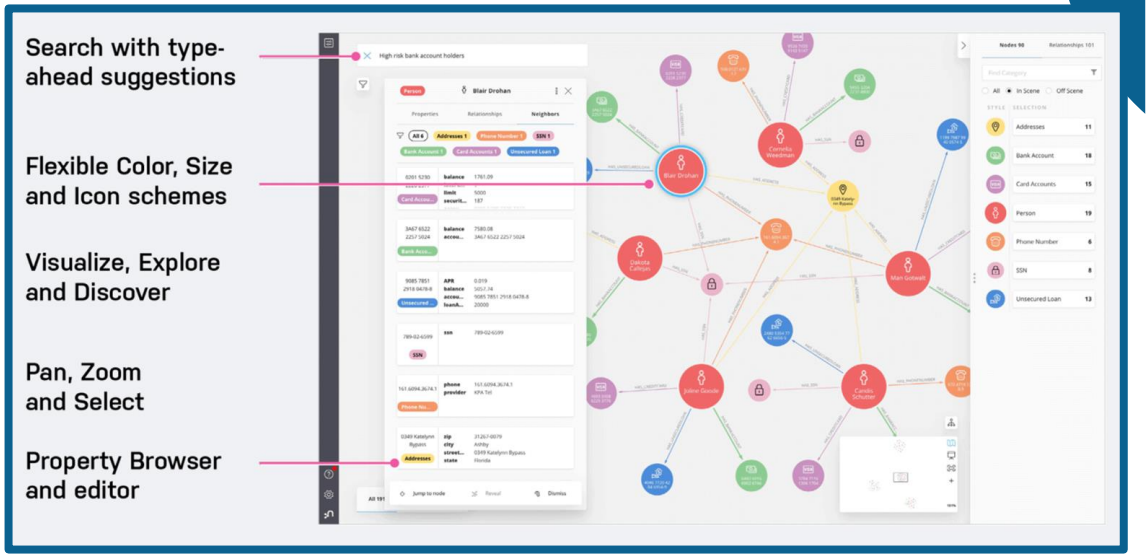
neo4j
graphsummit

Visualisation

Data Visualization with Neo4j Bloom

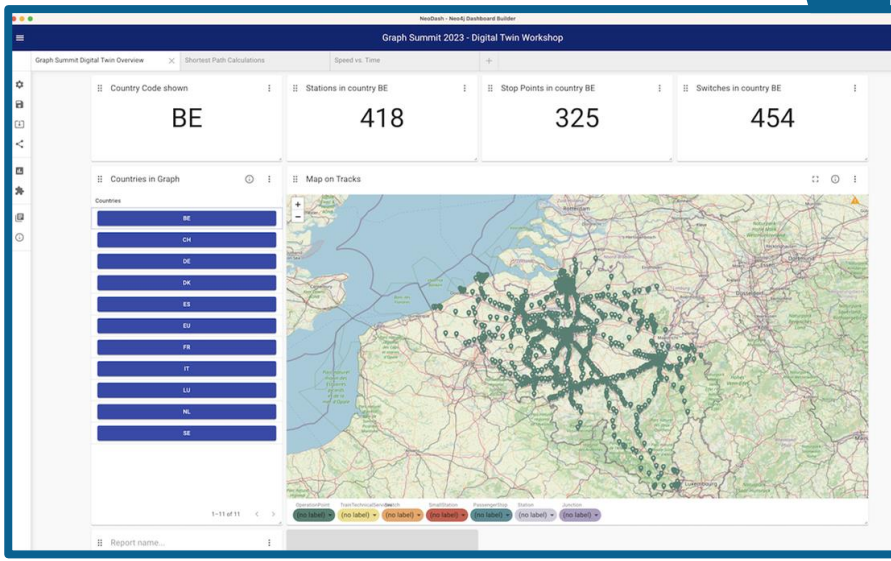
Neo4j's user-friendly graph database visualization, exploration and collaboration tool.

- Visually explore graphs
- Prototype faster
- Visualize and discover
- Easy for non-technical users



NeoDash

- Fully open source
  - https://github.com/neo4j-labs/neodash
  - Extensible
- Can be integrated in existing portal
- Supported through Neo4j's Professional Services team



neo4j
graphsummit

Use Case Explanation

Digital Twin - An Overview

What is a Digital Twin?

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

neo4j
graphsummit

Modeling a solution

What is graph data modeling?

A *collaborative* effort where the application domain is analysed by stakeholders and developers to come up with the optimal model for use with Neo4j.

Stakeholders include:

- Business analysts
- Architects
- Managers
- Project leaders
- Data Scientists

The Modeling Workflow

1. Derive the question

2. Obtain the data

3. Develop a model

4. Ingest the data

5. Query/Prove the model

no one

EVER

gets it right

FIRST

time

don’t worry

neo4j
graphsummit

Modeling – Step 1

Domain knowledge – High level requirements

We've already got this!

- (Normally)
- In this case we've talked about the Domain
- You will have the knowledge of your own Domain

neo4j
graphsummit

Modeling – Step 2

Sample Data

Get Sample Data

Static Rail Network*

- Sections of lines
  - Length,
  - Speed
- Operational Points (Stations etc)
  - Geolocation information,
- Points of Interest (POI) along lines

Source: Register of Infrastructure (RINF) - https://data-interop.era.europa.eu/

neo4j
graphsummit

Modeling – Step 3

Domain Questions

Data Modeling – Example Domain Questions

1. What is the route from Operational Point X to Operational Point Y?
   - What's the quickest way to get a repair crew from Technical Services to a given Switch?
2. What is an alternative route if an Operational Point on a route is closed?
   - A Switch *is* broken and we need to reroute Trains
3. How many routes are affected if I need to upgrade an Operational Point?
   - A Switch needs to be upgraded to support the network
4. What POIs are along a route?
   - Can we make revenue from referral commissions? Find busier routes during tourism season?

neo4j
graphsummit

Modeling – Step 4

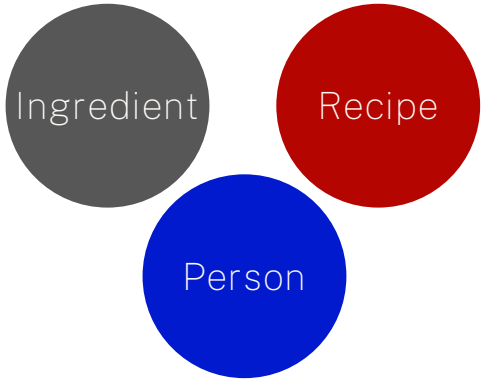
Identifying entities and connections

Identify Entities from Questions

Entities are the nouns in the domain questions:

1. What ingredients are used in a recipe?

2. Who is married to this person?

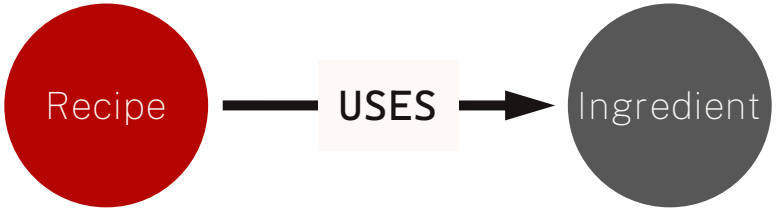


- The generic *nouns* often become labels in the model
- Use *domain knowledge* deciding how to further group or differentiate entities

Identify Connections between Entities

Connections are the verbs in the domain questions:

- What ingredients are used in a recipe?



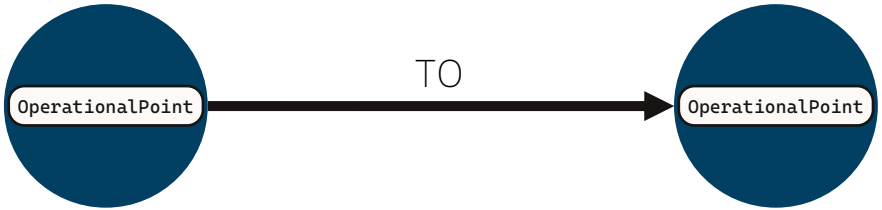
- Who is married to this person?

Using our Questions – Question 1

1. What is the route from Operational Point X to Operational Point Y?
   - What's the quickest way to get a repair crew from Technical Services to a given Switch?
2. What is an alternative route if an Operational Point on a Section is closed?
   - A Switch is broken and we need to reroute Trains
3. How many routes are affected if I need to upgrade an Operational Point?
   - A Switch needs to be upgraded to support the network
4. What POIs are near Station Operational Points on a Section?
   - Can we make revenue from referral commissions? Find busier routes during tourism season?

Using our Questions – Question 1 – Model

1. What is the route from Operational Point X to Operational Point Y?
   - What's the quickest way to get a repair crew from Technical Services to a given Switch?

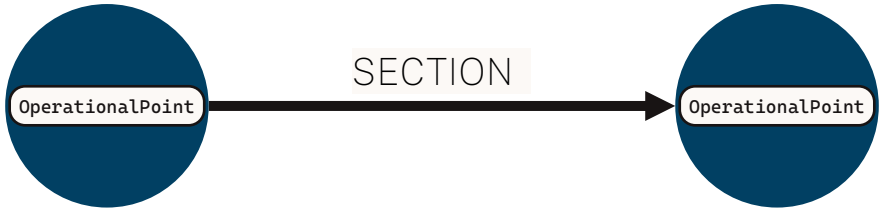


Using our Questions – Question 2

1. What is the route from Operational Point X to Operational Point Y?
   - What's the quickest way to get a repair crew from Technical Services to a given Switch?
2. What is an alternative route if an Operational Point on a Section is closed?
   - A Switch is broken and we need to reroute Trains
3. How many routes are affected if I need to upgrade an Operational Point?
   - A Switch needs to be upgraded to support the network
4. What POIs are near Station Operational Points on a Section?
   - Can we make revenue from referral commissions? Find busier routes during tourism season?

Using our Questions – Question 2 – Model

2. What is an alternative route if an Operational Point on a Section is closed?
   - A Switch is broken and we need to reroute Trains

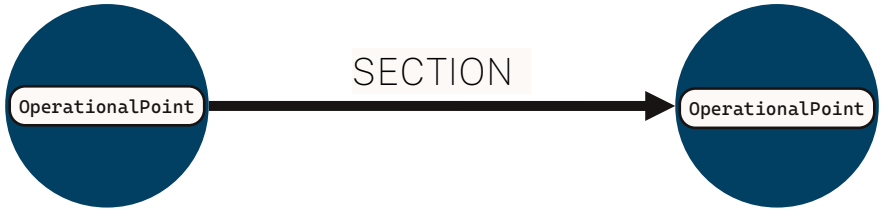


Using our Questions – Question 3

1. What is the route from Operational Point X to Operational Point Y?
   - What's the quickest way to get a repair crew from Technical Services to a given Switch?
2. What is an alternative route if an Operational Point on a Section is closed?
   - A Switch is broken and we need to reroute Trains
3. How many routes are affected if I need to upgrade an Operational Point?
   - A Switch needs to be upgraded to support the network
4. What POIs are near Station Operational Points on a Section?
   - Can we make revenue from referral commissions? Find busier routes during tourism season?

Using our Questions – Question 3 – Model

3. How many routes are affected if I need to upgrade an Operational Point?
   - A Switch needs to be upgraded to support the network

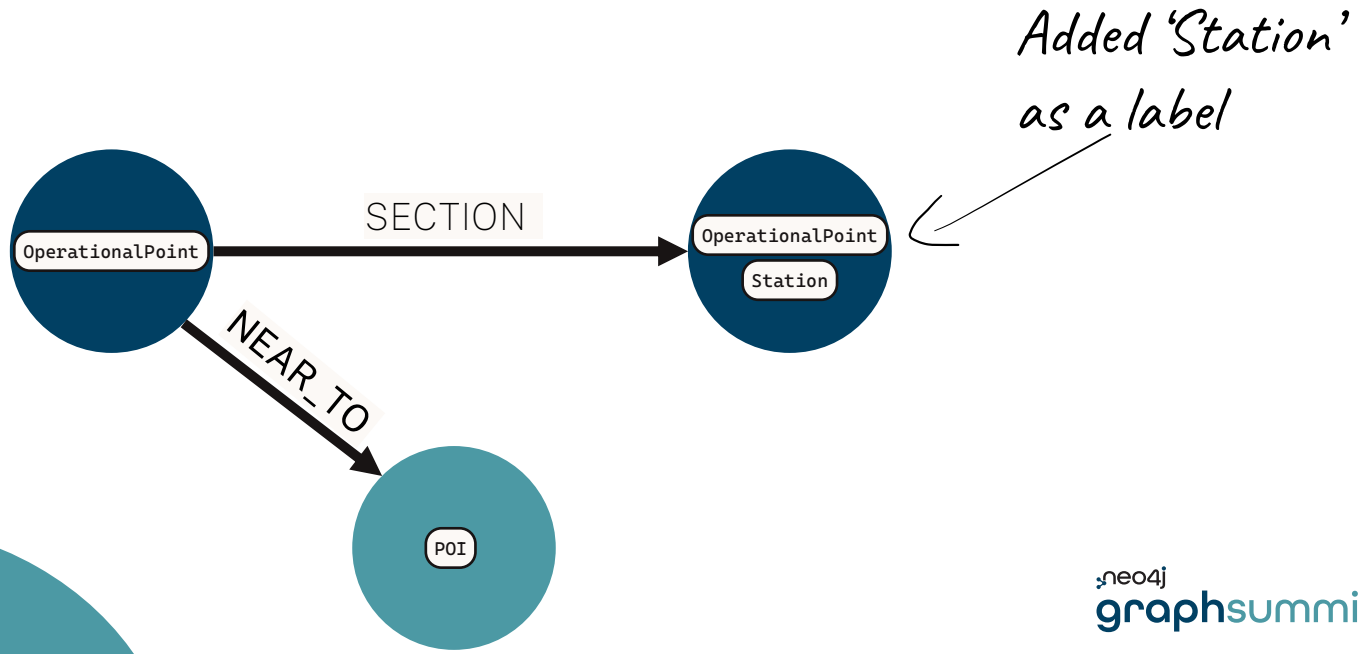


Using our Questions – Question 4

1. What is the route from Operational Point X to Operational Point Y?
   - What's the quickest way to get a repair crew from Technical Services to a given Switch?
2. What is an alternative route if an Operational Point on a Section is closed?
   - A Switch is broken and we need to reroute Trains
3. How many routes are affected if I need to upgrade an Operational Point?
   - A Switch needs to be upgraded to support the network
4. What POIs are near Station Operational Points on a Section?
   - Can we make revenue from referral commissions? Find busier routes during tourism season?

Using our Questions – Question 4 – Model

4. What POIs are near Station Operational Points on a Section?
   - Can we make revenue from referral commissions? Find busier routes during tourism season?



Hang on...

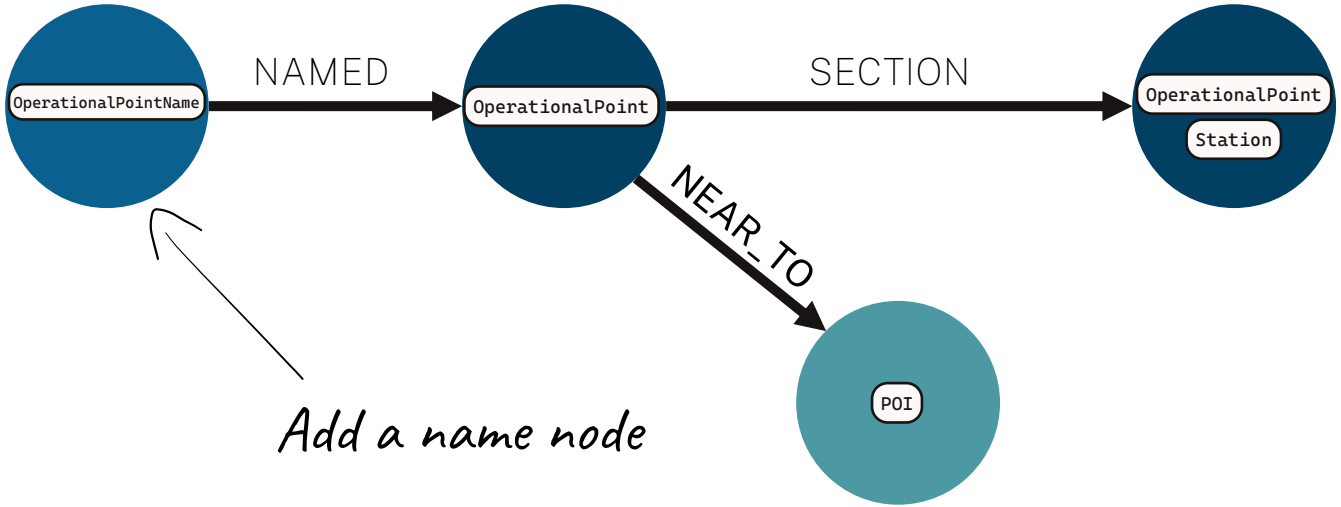
```
"EU00001","BorderPoint","Nieuweschans Grens","53.189","7.21112","Netherlands"
"EU00001","BorderPoint","Weener Grenze","53.189","7.21112","Germany"
```

We have duplicate entries

Wait, what?

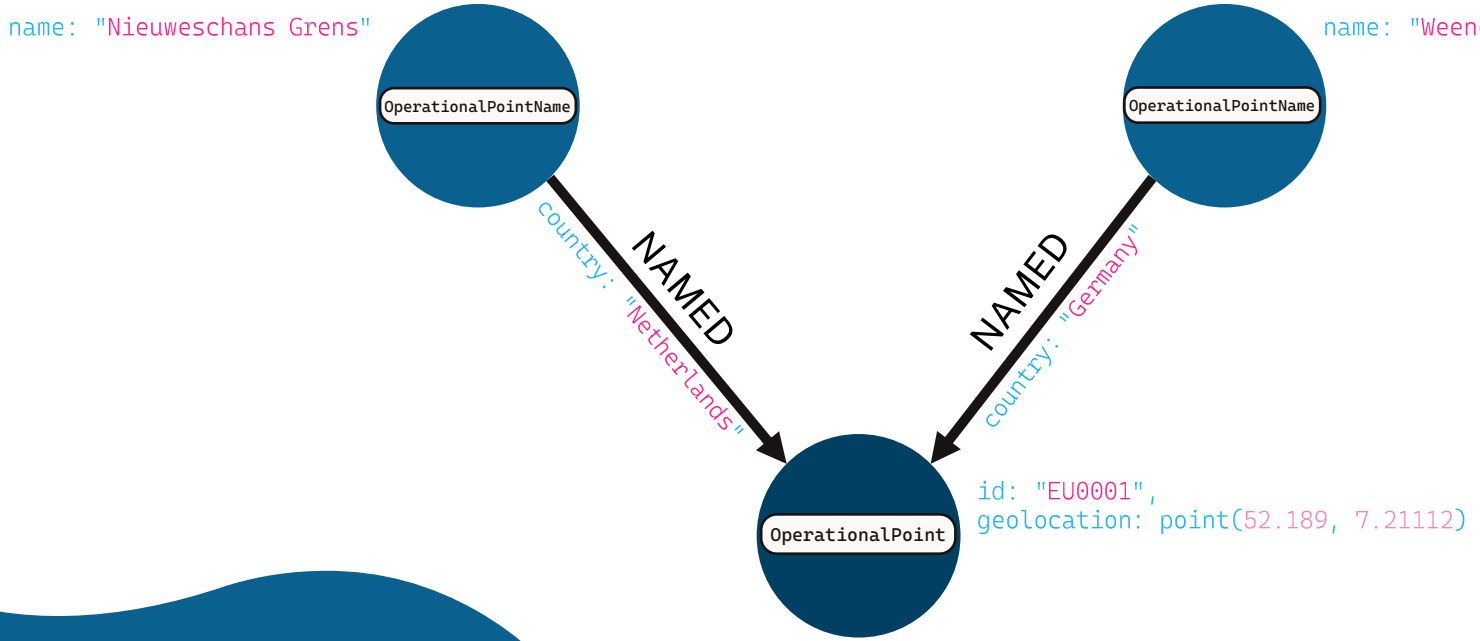
How do we deal with this?

Name nodes


OperationalPointName
NAMED
OperationalPoint
SECTION
OperationalPoint
Station
NEAR_TO
POI
Add a name node

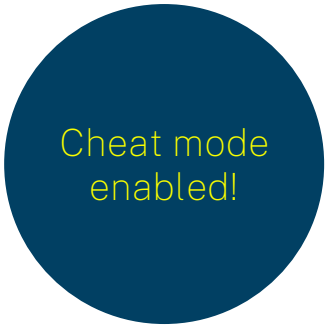
How does that look in reality?

```
"EU00001","BorderPoint","Nieuweschans Grens","53.189","7.21112","Netherlands"
"EU00001","BorderPoint","Weener Grenze","53.189","7.21112","Germany"
```



Whilst we're here, let's talk countries

- Multiple ways to model
  - Properties
  - Labels
- What should we choose?



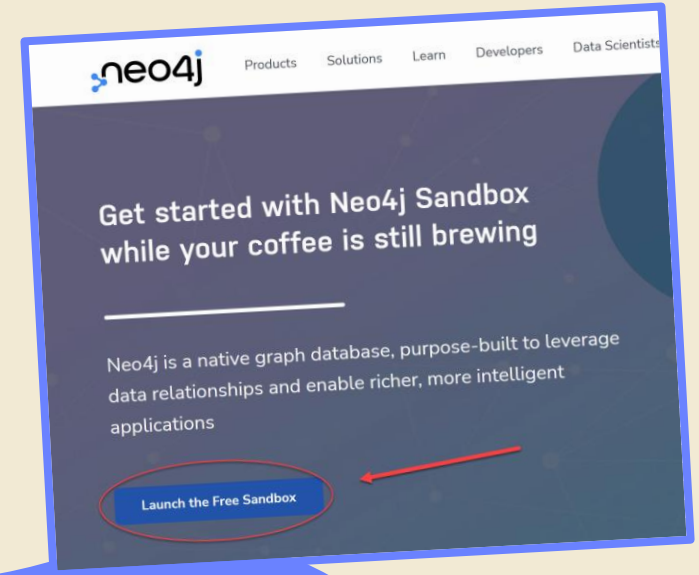
Before we go further

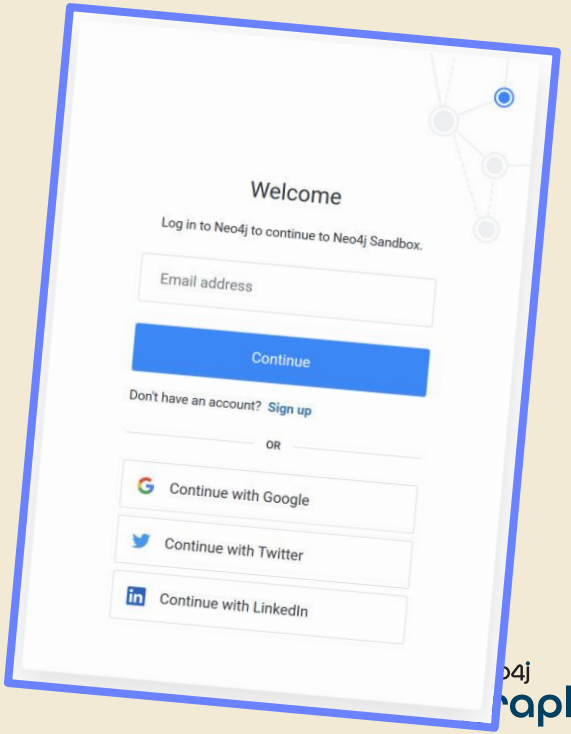
Environment for later

- Neo4j Desktop
- Neo4j Sandbox

Before we go further – Sandbox Login

- Create a sandbox for later
- https://sandbox.neo4j.com/

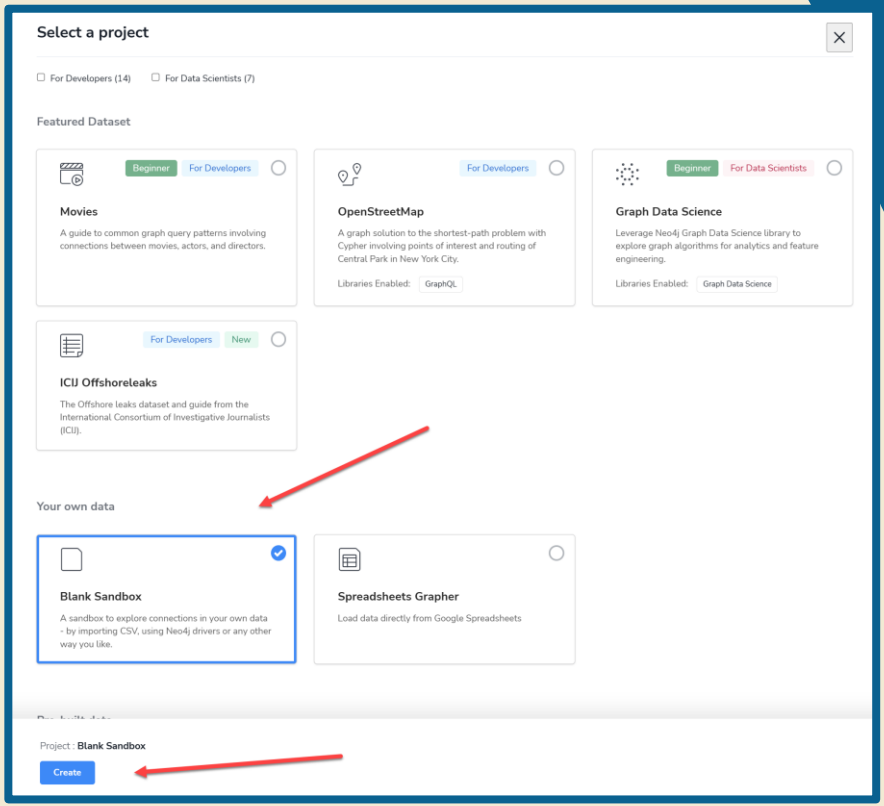




Before we go further – Blank Sandbox

- https://sandbox.neo4j.com/
- Create a ‘Blank Sandbox’

- If you’re using *Desktop*
  - Install APOC
  - Install GDS



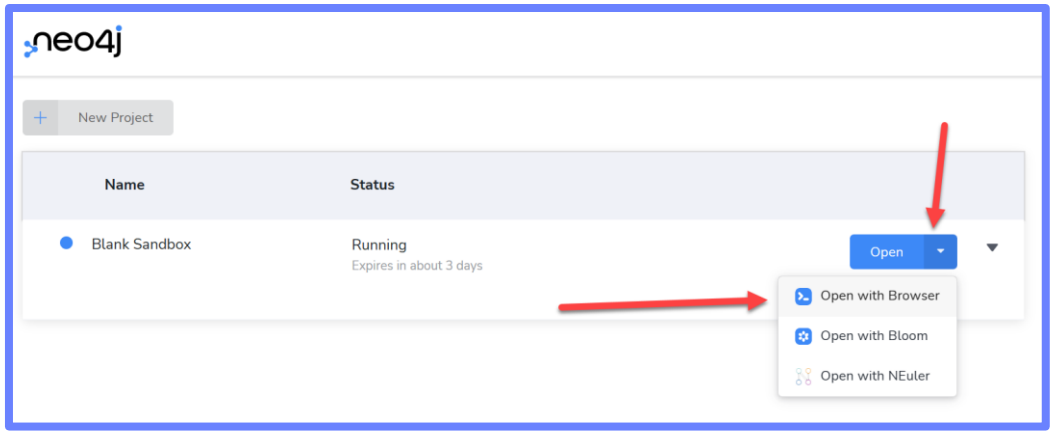
neo4j graphsummit

neo4j graphsummit

Workshop

Examples

- Go to the Neo4j Sandbox you created earlier (or Desktop!)
  - https://sandbox.neo4j.com/
- ‘Open with Browser’



Get the code

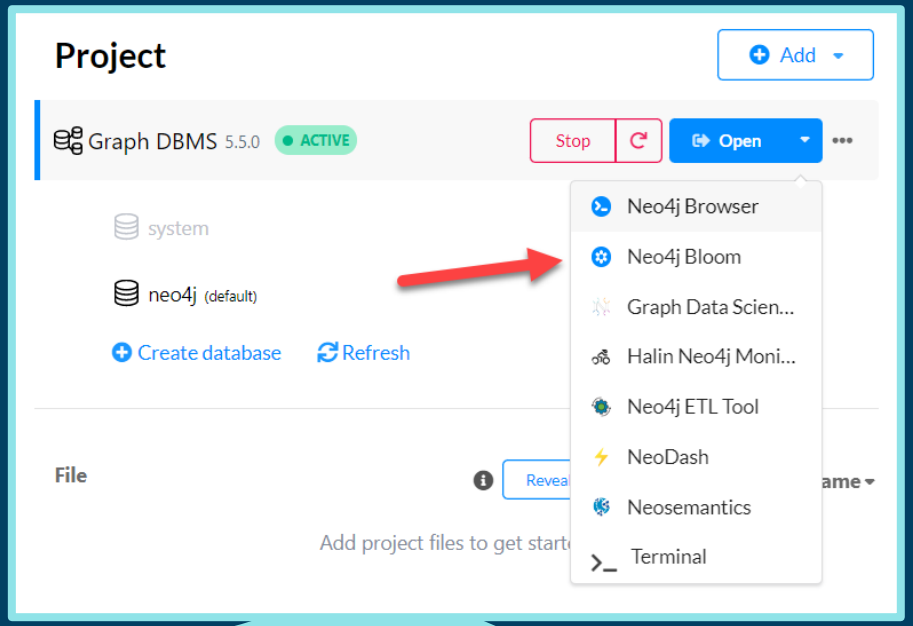
- Open the Github page:
- https://github.com/cskardon/gsummit2023

neo4j
graphsummit

Visualisation - Bloom

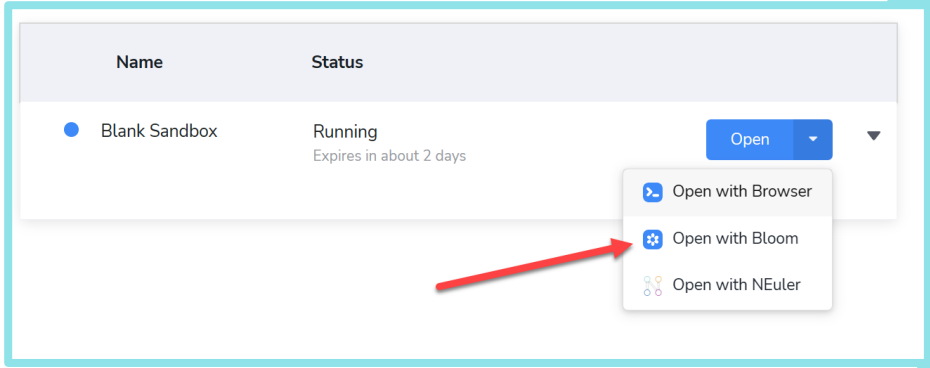
Bloom - Desktop

- Open Bloom

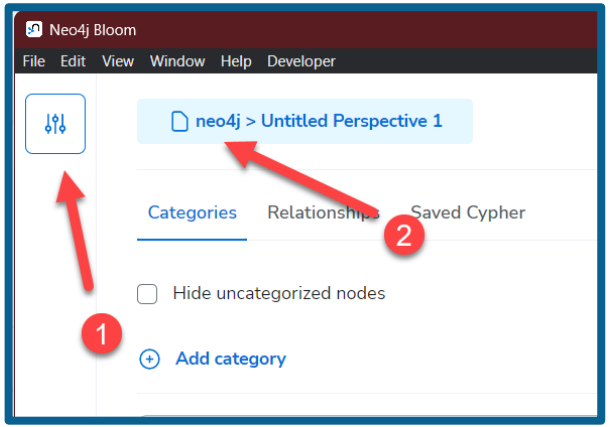


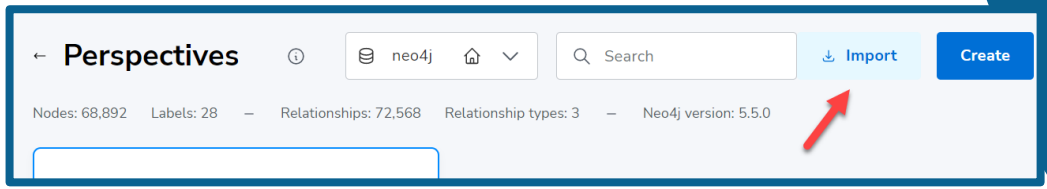
Bloom - Sandbox

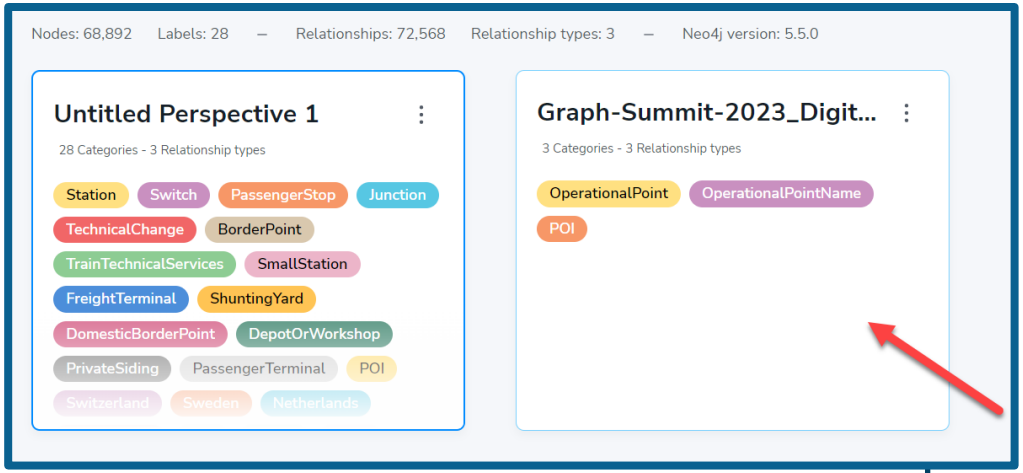
- Open Bloom



Bloom – Import Perspective







graphsummit

neo4j
graphsummit

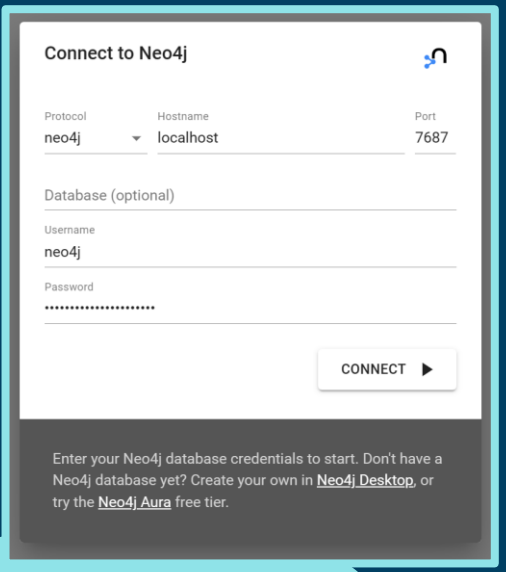
Visualisation - NeoDash

NeoDash

- Open NeoDash:
  - http://neodash.graphapp.io/

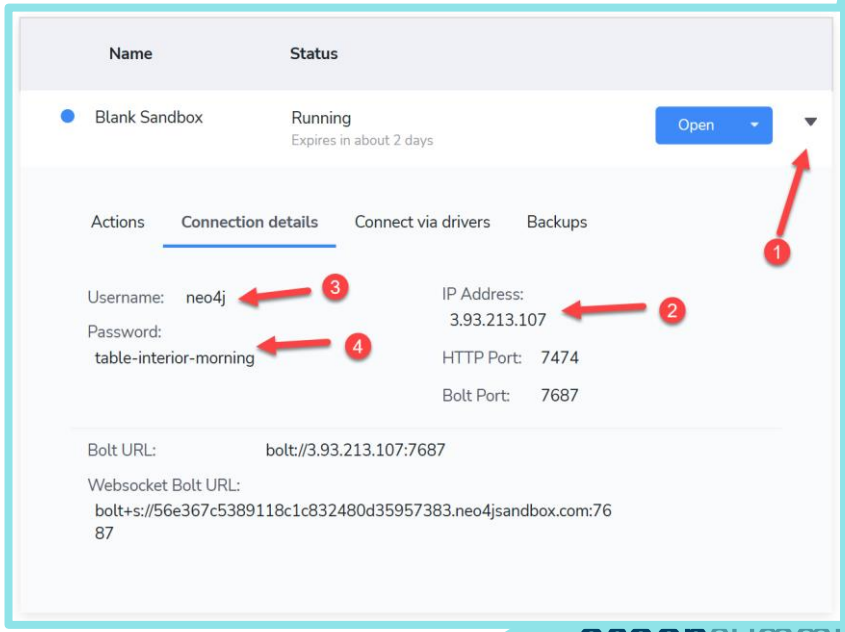
NeoDash - Desktop

- Host: localhost
- User: neo4j
- Password: YOURS!



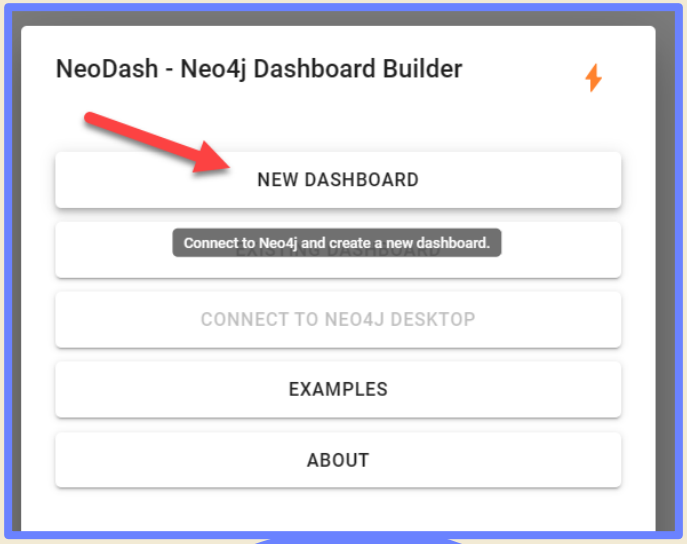
NeoDash - Sandbox

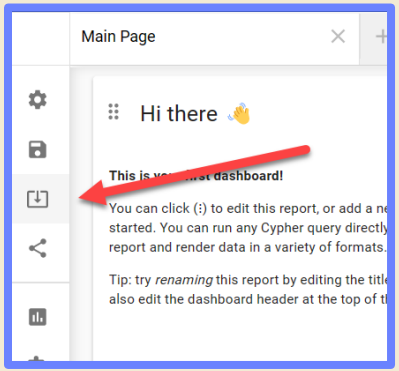
- Open Sandbox
- Get connection details

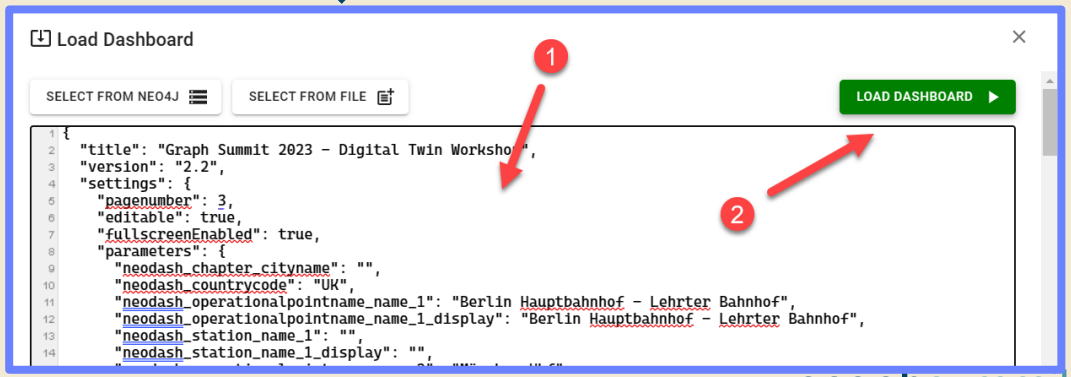


NeoDash – Import Dashboard

- Create New Dashboard
- Import from Source files







graphsummit

neo4j
graphsummit

Q & A

neo4j graphsummit

Thank you

Contact us at
sales@neo4j.com