The Meetup Dataset

What are we going to do?

- Introduce the dataset
- Modeling workflow
- Intro to the property graph model
- Load CSV data

The Meetup dataset

Meetup makes recommendations

People in this Meetup are also in:

Data & Analytics Innovation & Entrepreneurship
1,167 Members

SocDM Society of Data Miners
531 Analytics Practitioners

Cassandra London
1,635 Members

London New Tech
5,558 New technologists

Meteor London
1,596 Meteorites

London Ajax User Group
1,407 Software Engineers

ay: Join 190 R Users at "LondonR Meeting (and Workshop)" Inbox x

LondonR <info@meetup.com> Unsubscribe 15:40 (15 hou
to me

Events in this message

LondonR Meeting (and Workshop) Mon 30 Nov 2015 14:30 – 22:00 (WET)

Add to Google Calendar

Meetup

MONDAY

LondonR Meeting (and Workshop)

LondonR

Monday, November 30, 2015
2:30 PM

Balls Brothers
Minster Court, Mincing Lane, EC3R 7PP
London

190 R Users going, including:

John Van Praag
"R in my main coding language. A
looking to gain deeper understanc

THURSDAY, NOVEMBER 26

6:30 PM KNIME User Group UK
Anomaly Detection in Predictive Maintenance with KNIME
3 KNIMErs going

6:30 PM Spark London
11th Spark London Meetup - part of Big Data Week London
120 Members going

New Meetup Group: Agile without Borders Inbox x

Meetup <info@meetup.com> Unsubscribe 21 Nov (4 days a
to me

Meetup

Agile without Borders

Community of Agile practitioners with in interest in developing techniques to scale across geography, time and cultures.

Join us

Find out more

Organized by

+1

What recommendations can <u>we</u> make?

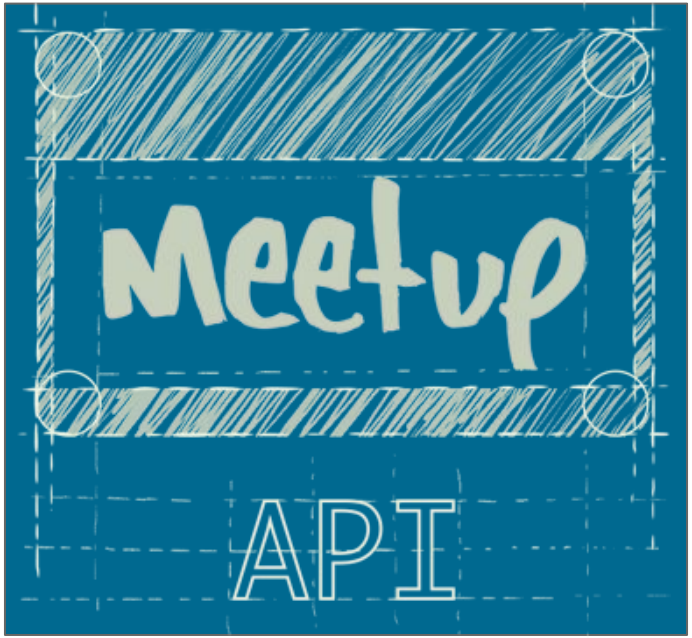
What recommendations can <u>we</u> make?

- Several different types
  - groups to join
  - topics to follow
  - events to attend

As a user of meetup.com trying to

find groups to join and events to attend

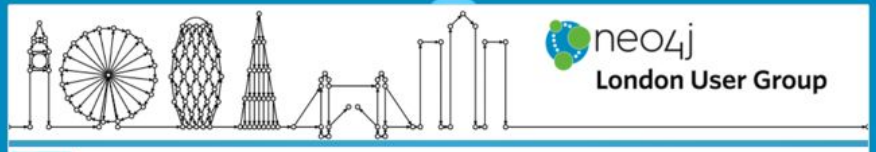
The data



meetup.com/meetup_api/

Find a Meetup Group | Start a Meetup Group | Invite Friends to Join

Messages | Notifications



neo4j London User Group

Home | Members | Sponsors | Photos | Pages | Discussions | More | Group tools | My profile

neo4j

London, United Kingdom

Founded Jun 1, 2011

About us...

Invite friends

People	2,823
Group reviews	39
Upcoming Meetups	7
Past Meetups	154
Our calendar	

edit

Organizers:
Neo4j, Mark Needham

Contact

We're about:
Data Mining · New Technology · Web Development · Data Visualization · Data Analytics · Open Source

Graphs are everywhere. Use them.

+ SCHEDULE A NEW MEETUP

Upcoming 7 | Past | Drafts 3 | Calendar

FEATURED MEETUP

Beyond SQL: Managing events & relationships in Social Care with Neo4j

Skillsmatter (CodeNode)
CodeNode, 10 South Place, EC2M 2RB, GB, London (map)

Wed Apr 13
6:30 PM
I'M GOING
41 going
1 comment

This month we'll have Andre[illegible]yson from Capita looking at how Ne[illegible]d other NoSQL technologies are [illegible]g used to build a new Social Care syste[illegible]based upon a modern cloud-ready software architecture. The talk will include a l[illegible]at the graph model used to enable the mapping of person-centric events and relationships, as well as the... LEARN MORE

Hosted by: Mark Needham (Co-Organizer)

Cypher for SQL Developers

Neo Technology
5 - 11 Lavington Street, London, SE10NZ, London (map)

Tue Apr 5
6:30 PM
I'M GOING
3 days left
48 going
0 comments

This session is an introduction to graphs for SQL developers. We'll start with an overview of Neo4j and its query language Cypher before moving on to a worked example... LEARN MORE

Hosted by: Mark Needham (Co-Organizer), and Rik

What's new

MORE

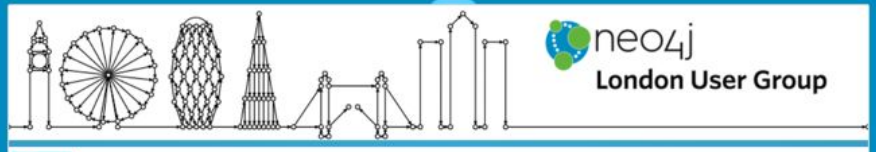
NEW MEMBER
Mads Jensen joined
2 days ago

NEW MEMBER
Laurie Clark-Michalek joined
2 days ago

NEW COMMENT
Mark Needham commented on Real-Time Fraud Detection with Graphs - Dr Jim Webber, Chief Scientist at Neo4j
2 days ago

NEW MEMBER
Gareth Chillingworth

Find a Meetup Group | Start a Meetup Group | Invite Friends to Join | Messages | Notifications



neo4j London User Group

Home | Members | Sponsors | Photos | Pages | Discussions | More | Group tools | My profile

neo4j

London, United Kingdom

Founded Jun 1, 2011

About us...

Invite friends

People	2,823
Group reviews	39
Upcoming Meetups	7
Past Meetups	154
Our calendar	

edit

Organizers:
Neo4j, Mark Needham

Contact

We're about:
Data Mining · New Technology · Web Development · Data Visualization · Data Analytics · Open Source

Graphs are everywhere. Use them.

+ SCHEDULE A NEW MEETUP

Upcoming 7 | Past | Drafts 3 | Calendar

FEATURED MEETUP

Beyond SQL: Managing events & relationships in Social Care with Neo4j

Skillsmatter (CodeNode)
CodeNode, 10 South Place, London, EC2M 2RB, GB, London (map)

Wed Apr 13
6:30 PM
I'M GOING
41 going
1 comment

This month we'll have Andrew Bryson from Capita looking at how Neo4j and other NoSQL technologies are being used to build a new Social Care system based upon a modern cloud-ready software architecture. The talk will include a look at the graph model used to enable the mapping of person-centric events and relationships, as well as the... LEARN MORE

Hosted by: Mark Needham (Co-Organizer)

Cypher for SQL Developers

Neo Technology
5 - 11 Lavington Street, London, SE10NZ, London (map)

Tue Apr 5
6:30 PM
I'M GOING
3 days left
48 going
0 comments

This session is an introduction to graphs for SQL developers. We'll start with an overview of Neo4j and its query language Cypher before moving on to a worked example... LEARN MORE

Hosted by: Mark Needham (Co-Organizer), and Rik

What's new

MORE

NEW MEMBER
Mads Jensen joined
2 days ago

NEW MEMBER
Laurie Clark-Michalek joined
2 days ago

NEW COMMENT
Mark Needham commented on Real-Time Fraud Detection with Graphs - Dr Jim Webber, Chief Scientist at Neo4j
2 days ago

NEW MEMBER
Gareth Chillingworth

What data do we have?

- Groups
- Members
- Events
- Topics
- Time & Date
- Location

First Recommendation: Find similar groups

As a member of <my-favourite-group>

I want to find other similar meetup groups

So that I can join those groups

What makes groups similar?

neo4j

We're about:
Data Mining · New Technology · Web Development · Data Visualization · Data Analytics · Open Source · Cloud Computing · Graph Databases · Big Data · NoSQL · Neo4j · Database Development · Java · Computer programming

We're about:
Open Source · Technology · Web Development · Computer programming · Agile Project Management · Java · Software Development

Deep Learning London

We're about:
Big Data Analytics · Artificial Intelligence · Computer programming · Big Data · Computer Science · Natural Language Processing · Machine Learning · Data Analytics · Data Visualization · Data Mining · Data Science · Algorithms · Deep Learning · neural networks

Data Science London

We're about:
Data Science · Machine Learning · Predictive Analytics · Data Mining · Big Data · Artificial Intelligence · Statistical Computing · Applied Statistics · Data Analytics · Open Source · Web Analytics · Text Analytics · Natural Language Processing · Hadoop · NoSQL

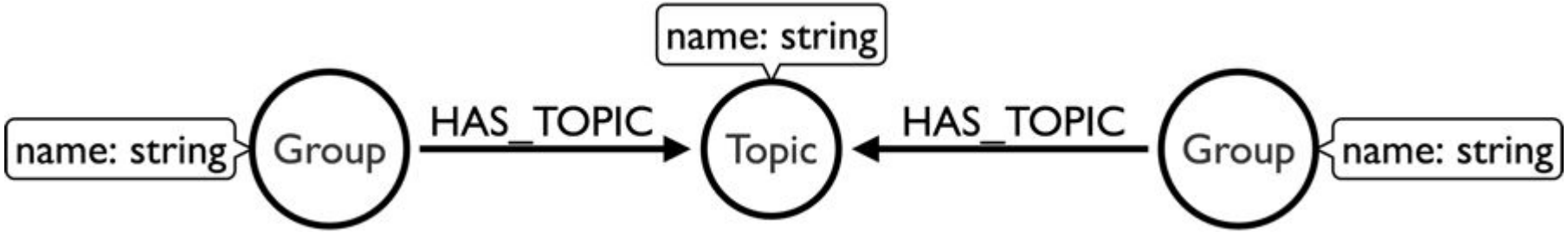
BIG DATA DEBATE

We're about:
BigData · Intellectual Discussion · Big Data · Debate · New Technology · Data Analytics · Data Visualization · Online Marketing · Database Development · Information Architecture · Information Science · Freedom · Cloud Computing · Business Strategy · Internet Professionals

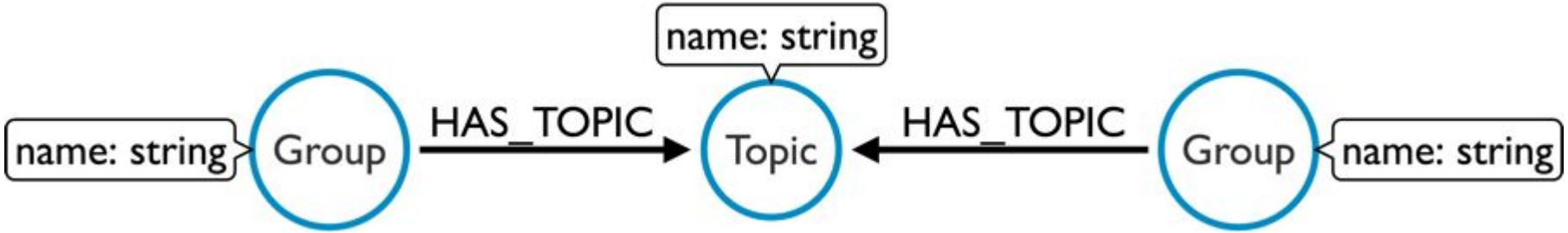
extract

We're about:
Marketing · Data · Software Development · New Technology · Web Technology · Business Intelligence · Cloud Computing · Business Strategy · Big Data · Machine Learning · Data Analytics · Data Visualization · Data Mining · Data Science · Big Data Analytics

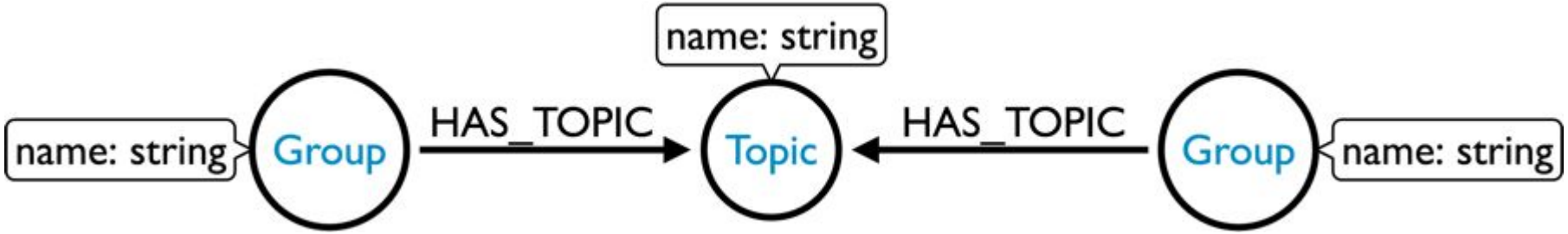
Labeled property graph



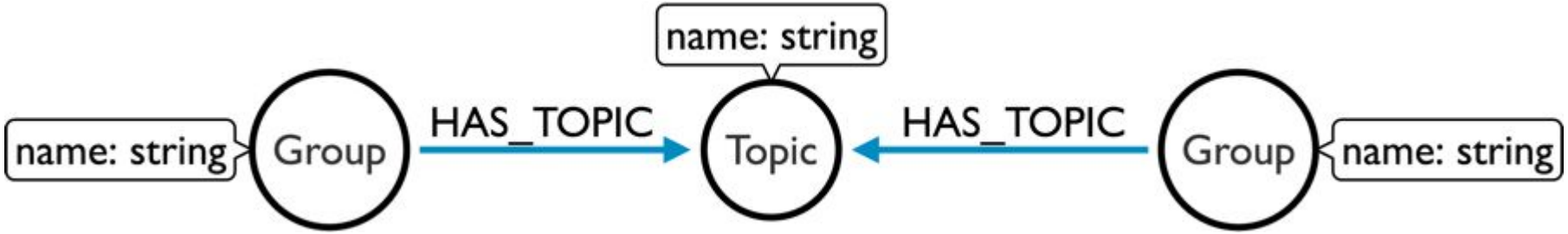
Nodes


name: string
name: string
Group
HAS_TOPIC
Topic
HAS_TOPIC
Group
name: string

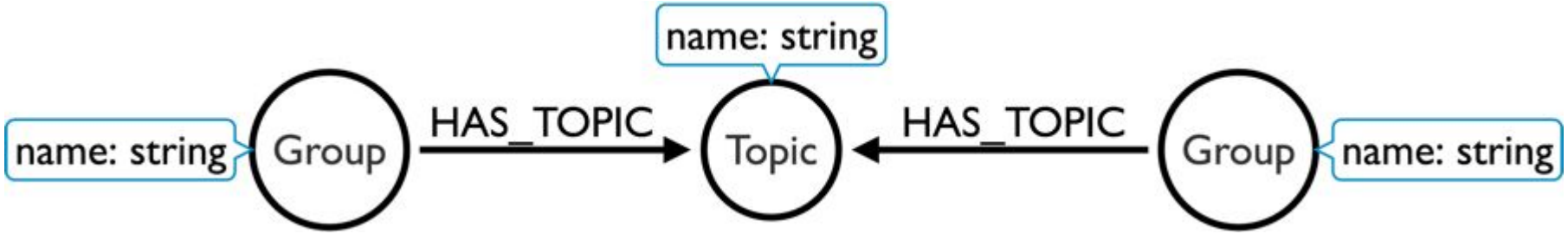
Labels


name: string
name: string
Group
HAS_TOPIC
Topic
HAS_TOPIC
Group
name: string

Relationships



Properties


name: string
name: string
Group
HAS_TOPIC
Topic
HAS_TOPIC
Group
name: string

Prepare your Neo4j graph.db directory

Copy folders

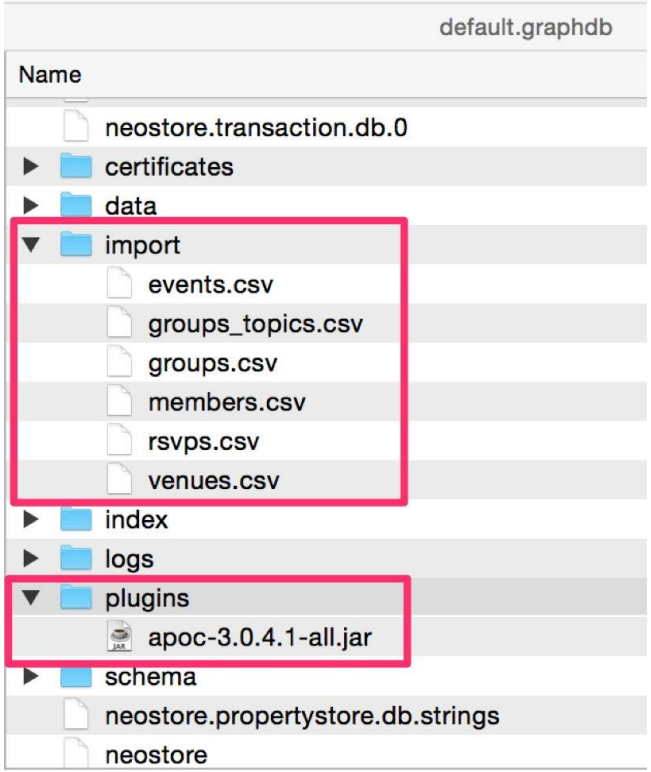
* `import`
* `plugins`

from **USB Stick**

to the **`default.graphdb`**

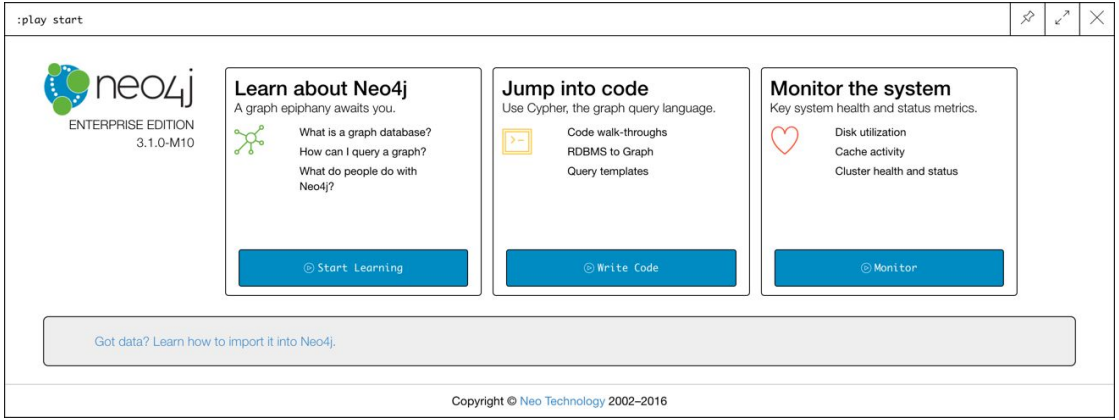
folder

(or **`$NEO4J_HOME`**)



Make sure you've got Neo4j running

1. Start the server.
2. It should be running on: http://localhost:7474
3. Log-in with default credentials
   - user: **neo4j** password: **neo4j**
4. Choose a **new** password



We're good to go!

Recommend groups by topic

Open your browser to http://localhost:7474

and execute the following command:

```
:play http://guides.neo4j.com/reco-nyc/file
```

Make sure to follow the setup instructions for the CSV files and import directory.

Then start the guide for this module Recommend Groups by Topic

Constraints and Indexes

Follow the guides in your browser until you see...

MERGE

MERGE is used to uniquely create graph structures.

It tries to find the provided pattern in the graph and creates it if not found.

For **unique nodes** it uses constraints for quick lookup and locking.

```
MERGE (node:Label {key:value})
ON CREATE SET node.property = 'some value'
```

It can also be used for **unique relationships** when both nodes are known (bound).

Unique Constraints

We create unique constraints to:

- ensure uniqueness
- allow fast lookup of nodes which match label-property pairs.

Unique Constraints

We create unique constraints to:

- ensure uniqueness
- allow fast lookup of nodes which match label-property pairs.

```
CREATE CONSTRAINT ON (label:Label)
ASSERT label.property IS UNIQUE
```

Indexes

We create indexes to:

- allow fast lookup of nodes which match label-property pairs.

Indexes

We create indexes to:

- allow **fast lookup** of nodes which match label-property pairs.

```
CREATE INDEX ON :Label(property)
```

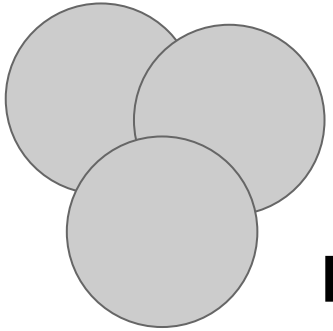
What are these fast lookups?

The following predicates use indexes:

- Equality
- `STARTS WITH`
- `CONTAINS`
- `ENDS WITH`
- Range searches
- (Non-) existence checks

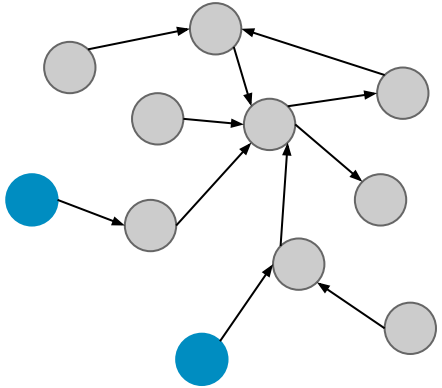
How are indexes used in neo4j?

Indexes are **only** used to find the **starting points** for queries.



Relational

Use index scans to look up rows in tables and join them with rows from other tables



Graph

Use indexes to find the starting points for a query.

Continue with the guide

Continue with the guide in your browser

Answers

Answers

Type the following command into the Neo4j browser to see the answers:

`:play` http://guides.neo4j.com/reco/answers/1.html

Continue with the guide

Continue with the guide in your browser

End of Module
The Meetup Dataset

Questions ?

LOAD CSV

```
[USING PERIODIC COMMIT] // optionally batch transactions
LOAD CSV      // load csv data
WITH HEADERS // optionally use first header row as keys in "row" map
FROM "url"    // file:// URL relative to $NEO4J_HOME/import or http://
AS row        // return each row of the CSV as list of strings or map
[FIELDTERMINATOR ";"] // optionally alt. delimiter
// ... rest of the Cypher statement ...
```

Group Membership

Exclude groups that I'm a member of

In our first recommendation query we suggested groups of which the user may already be a member. Let's address this redundancy.

What are we going to do?

- Add group memberships to the model
- Only recommend new groups to our user

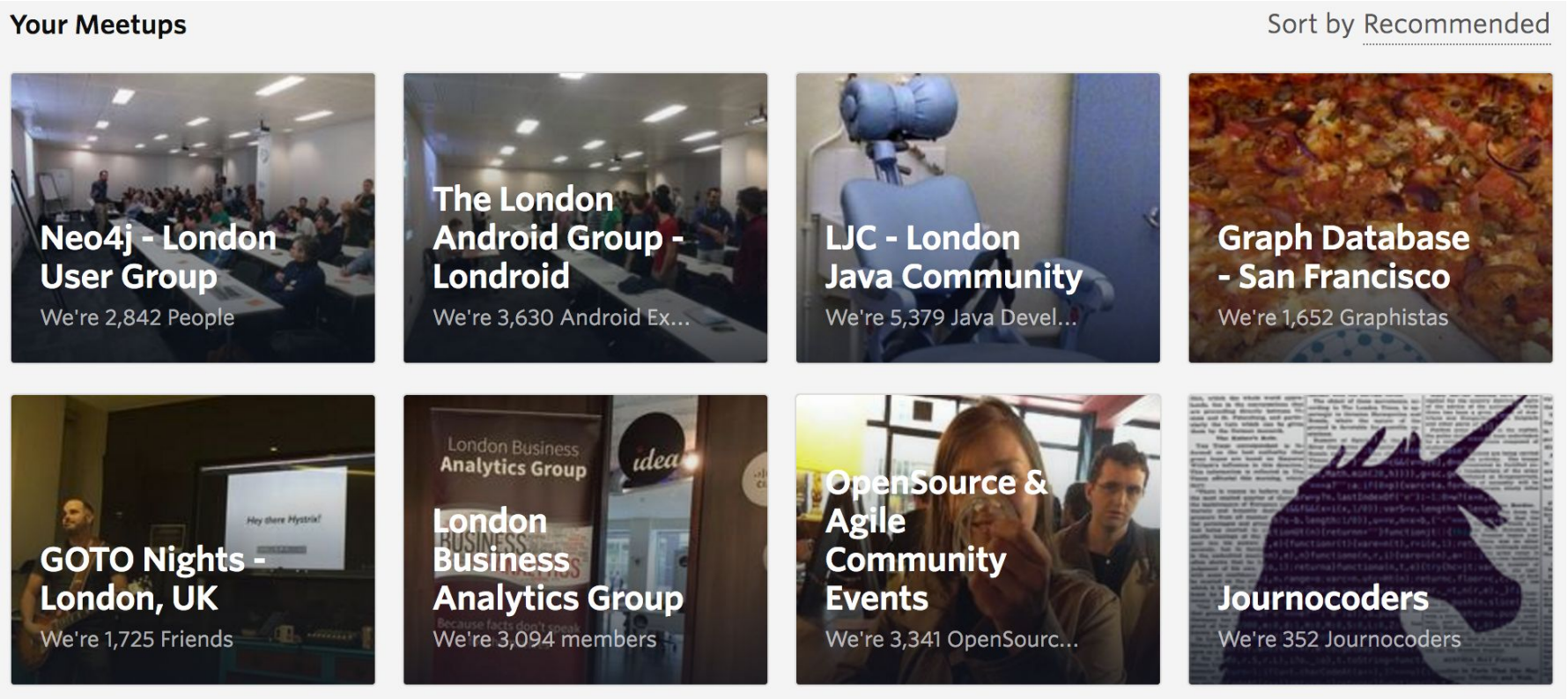
Exclude groups that I'm a member of

As a member of the Neo4j group

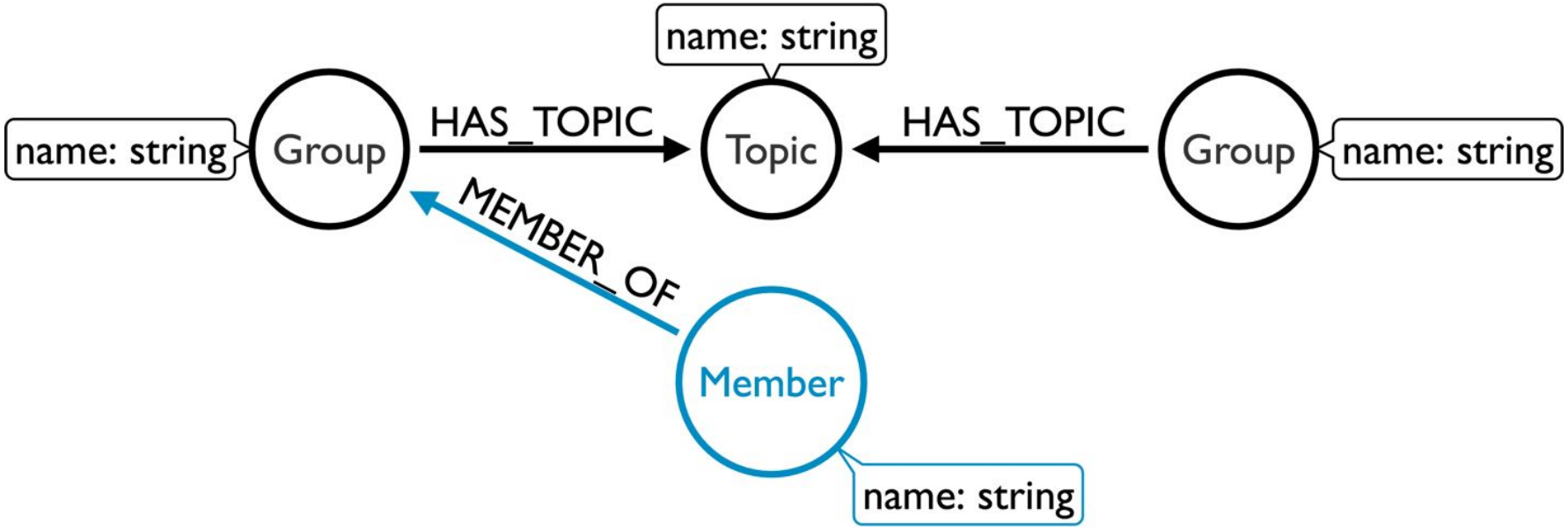
I want to find other similar meetup groups that I'm not already a member of

So that I can join those groups

Group Memberships



Adding members to the model



Start the next guide….

...if you aren’t looking at it already

Groups similar to mine

:play http://guides.neo4j.com/reco/file/02_my_similar_groups.html

Transaction State

Periodic Commit

Cypher keeps **all transaction state in memory** while running a query.

This is fine most of the time.

Periodic Commit

Cypher keeps **all transaction state in memory** while running a query. This is fine most of the time.

But when using `LOAD CSV` importing millions of rows, this state can get **very large** and may result in an `OutOfMemory` error.

Periodic Commit

```
// defaults to 1000
USING PERIODIC COMMIT
LOAD CSV
...
```

Periodic Commit

```
// defaults to 1000
USING PERIODIC COMMIT 10000
LOAD CSV
...
```

Continue with the guide

Continue with the guide in your browser

Answers

Answers

Type the following command into the Neo4j browser to see the answers:

`:play` http://guides.neo4j.com/reco/answers/2.html

The WITH Clause

WITH

The `WITH` clause allows query parts to be chained together, piping the results from one to be used as starting points or criteria in the next.

WITH

WITH is used to:

- limit the number of entries that are then passed on to other **MATCH** clauses.
- filter on aggregated values
- separate reading from updating of the graph

End of Module
Group Membership

Questions ?

Member Interests

What are we going to do?

- Add people’s interests to the model
- Infer interests for people who don’t specify them
- Use those interests to update our group recommendation query

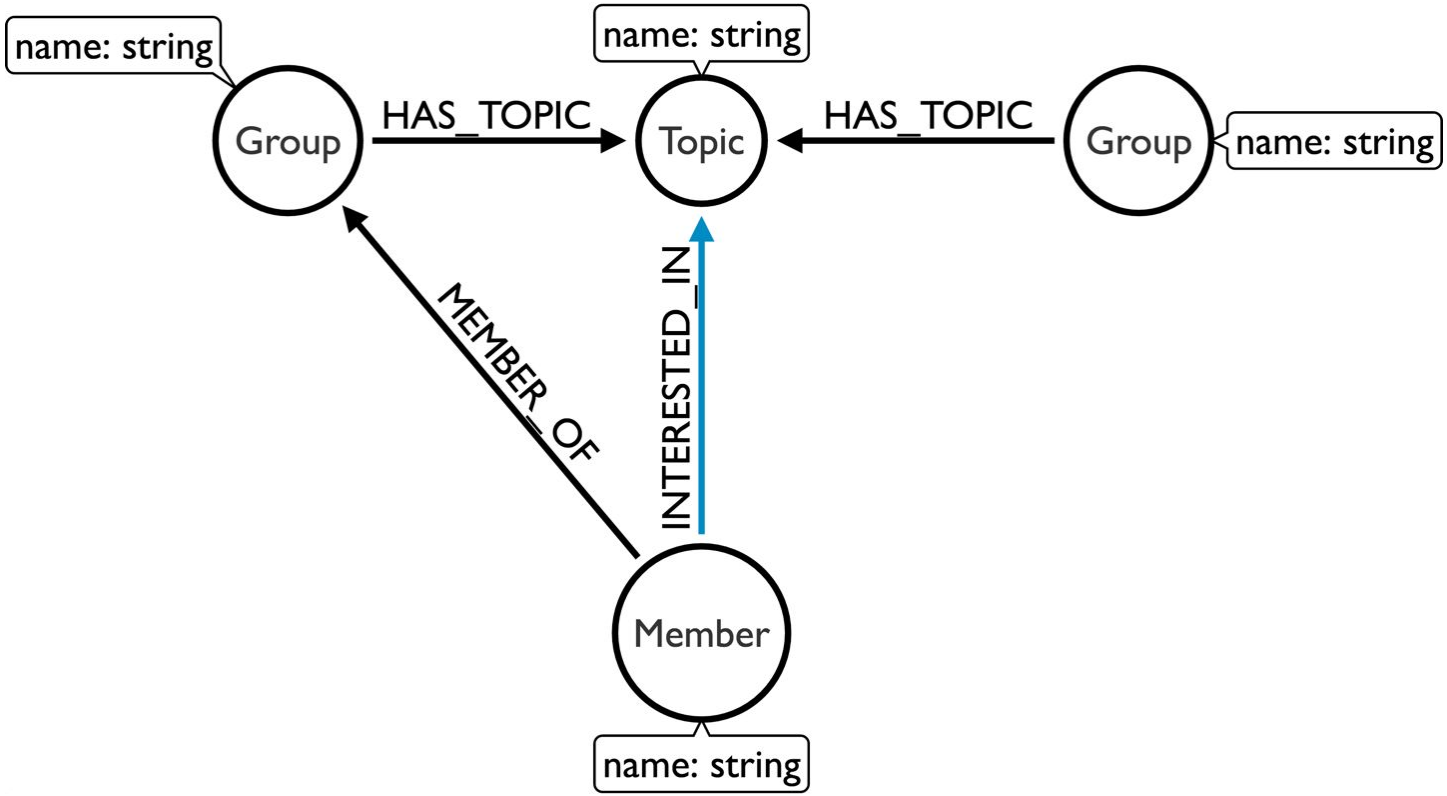
Adding interests

Interests

Hide interests on profile · Edit

Software Development · Test Driven Development · Open Source · Data Visualization · Natural Language Processing · Data Mining · Data Science · Machine Learning · Data Analytics · Neo4j · Graph Databases · NoSQL

Adding interests



Start playing the next guide….

neo4j

…if you aren’t viewing it already

My Interests

:play http://guides.neo4j.com/reco/file/03_my_interests.html

Lists with `split` and `UNWIND`

The split function

The `split` function lets us convert a string into a string list based on a delimiting character.

The split function

The `split` function lets us convert a string into a string list based on a delimiting character.

```
RETURN split("1;2;3", ";") AS topicIds
```

```
[1, 2, 3]
```

UNWIND

We can use `UNWIND` to transform any list into individual rows.

UNWIND

We can use `UNWIND` to transform any list into individual rows.

```
UNWIND [1,2,3] AS value
RETURN value
```

```
1
2
3
```

Combining split and UNWIND

```
UNWIND split("1;2;3", ";") AS topicId
RETURN topicId
```

1

2

3

Continue with the guide

Continue with the guide in your browser

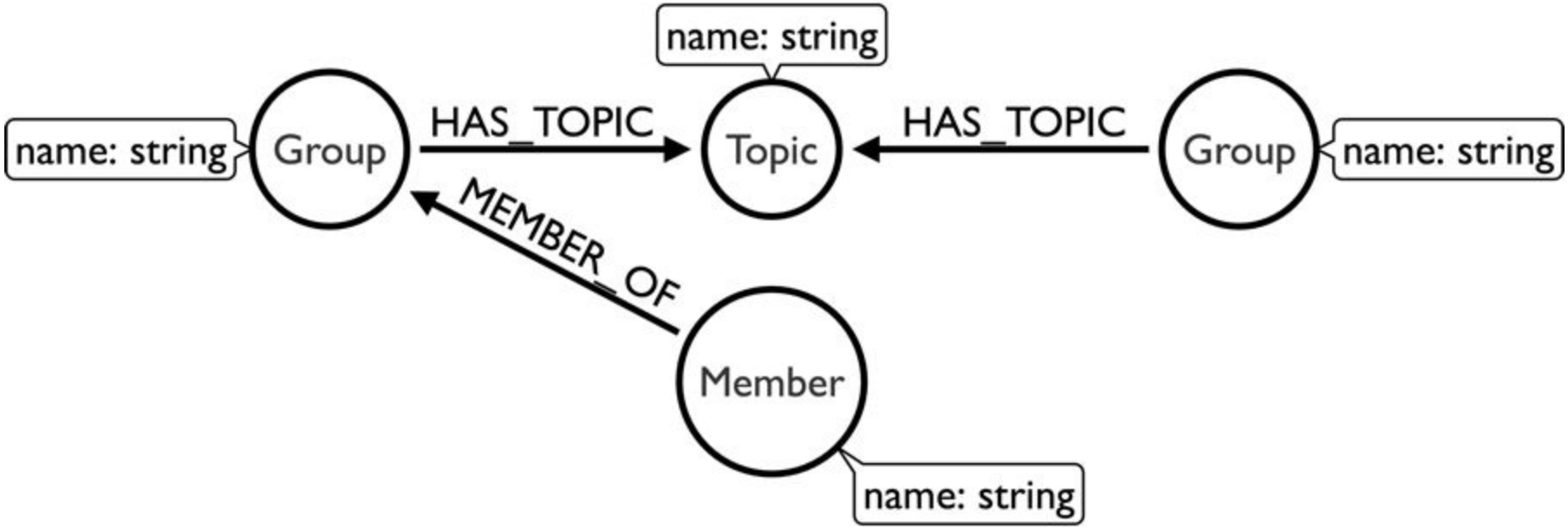
What am I actually interested in?

The `INTERESTED_IN` relationships we created earlier are based on interests that a user declared when he signed up for meetup.com.

People typically don't change those as their interests change but they do join new groups frequently. This allows us to know their interests better than they do!

We can do this by inferring `INTERESTED_IN` relationships based on user's groups.

What am I actually interested in?



Exercise: Inferring interests

1. Find topics that I'm interested in via the groups I'm a member of that I don't have an `INTERESTED_IN` relationship to.

2. Which topic shows up most frequently?

Answers

Answers

Type the following command into the Neo4j browser query pane to see the answers:

```
:play http://guides.neo4j.com/reco/answers/3.html
```

Continue with the guide

Continue with the guide in your browser

End of Module
Member Interests

Questions ?

Event Recommendations

Recommending events

Thus far we've only been recommending groups that people should join, but the whole point of meetup.com is to get people to attend events. In this section we'll do that.

What are we going to do?

- Add events to the model
- Layered Recommendations
  - events in my groups
  - events based on my interests

Events

Upcoming 3 Past Calendar

Simulating APIs for Effective Testing: (Micro)Service Virtualisation

UCL - Anatomy G29 JZ Young Lecture Theatre, Anatomy Building,
Gower Street, London, WC1E 6BT., London (map)

Wed Sep 28
6:15 PM
120 going
2 comments

Simulating APIs for Effective Testing: (Micro)Service Virtualisation in the 21st Century As we work more with distributed systems, microservices and legacy services, we... Learn more

Hosted by: Harry Hilditch

Meetup with James Gosling – the Father of Java

Excel, Royal Victoria Dock, 1 Western Gateway
E16 1XL, London (map)

Thu Oct 6
2:00 PM
Waitlist
7 days left
225 going
62 waiting
0 comments

Java: past present and future We are delighted to be able to announce that on the 6th October our friends at IP EXPO Europe are hosting a special LJC Meetup with the... Learn more

Hosted by: Barry Cranford (Founder)

WOMEN WHO CODE

London, United Kingdom
Founded Jan 6, 2014

About us...

Invite friends

Coders	3,308
Group reviews	27
Upcoming Meetups	19
Past Meetups	90
Our calendar	
Worldwide	71

Organizers:

Women Who Code, Claire Tran, Gen Ashley, hali, Sheree, Vinita Rathi

Featured Meetup

NEO4J - Week 2 of 3 - Hands on Cypher Workshop

Export Tell a friend Share

Thursday, September 29, 2016
6:30 PM to 9:00 PM

Neo Technology UK
5-11 Lavington Street, London, SE1 0NZ, London (map)

This is Part 2 of a 3-Part Series. This is a pilot program we've designed with the good folks at Neo4j.

This session will be a mixture of theory and hands-on practice sessions, and you will quickly learn how easy it is to work with a powerful graph database using Cypher as the query language.

We'll start by loading a movies dataset into neo4j, briefly covering an introduction to graph based modeling.
We'll then spend the majority of the session learning Cypher's syntax and writing increasingly powerful queries with it.
By the end of the session we'll be able to calculate the Bacon number for any of the actors/actresses in the database.

We'll end the session with a brief overview of the options for loading data into Neo4j in preparation for you to build your own graph over the next week/in next week's session.

Are you going?

Yes No

48 going

2 spots available

Gen Ashley
London Director, Co-Organizer, Event Host
Twitter @coderinheels | Ex-VP Citigroup | Head of Developer Outreach, Skills Matter | Chairman,... more

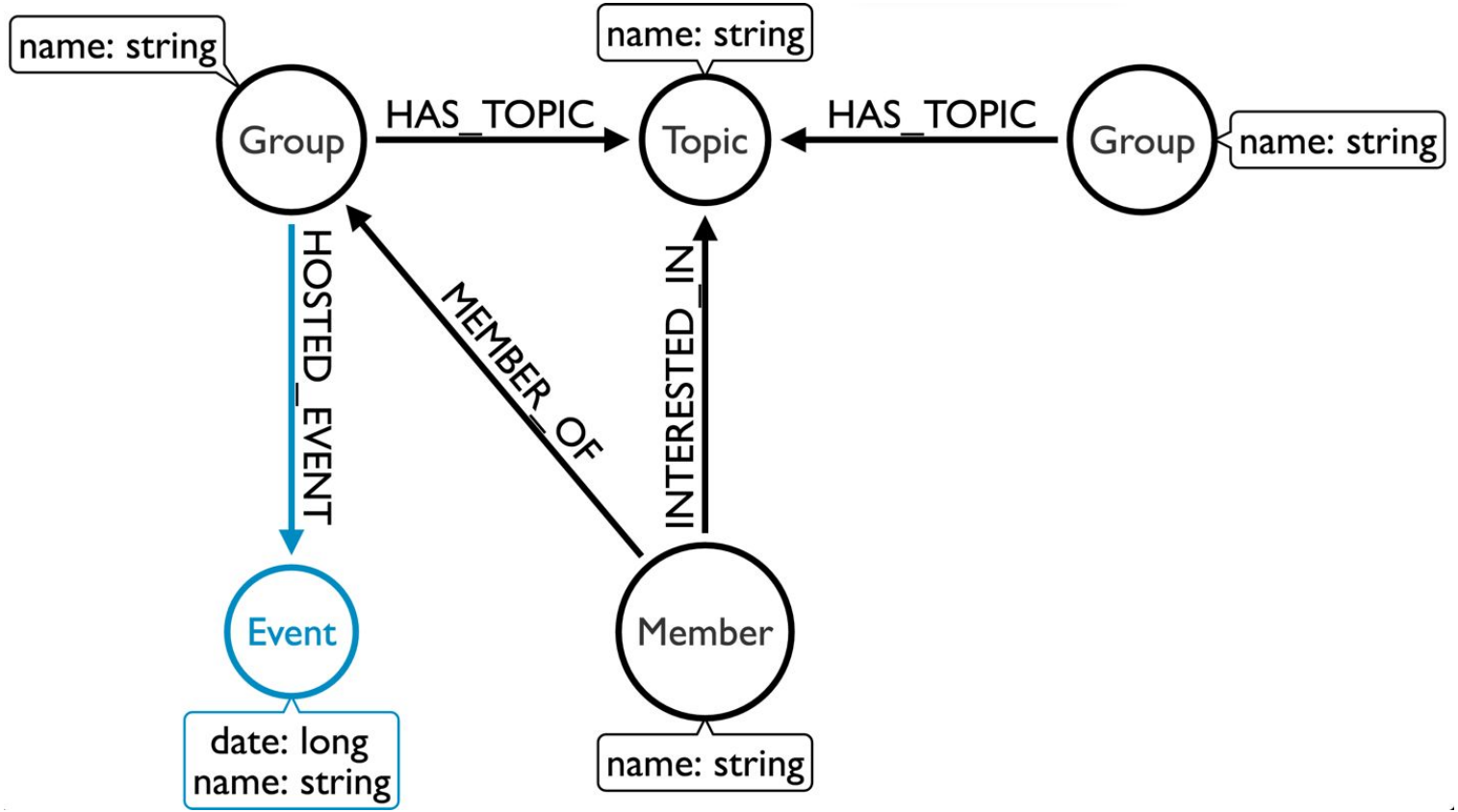
Claudine Letsae
Interested in tech and coding(novice)

IElizaveta Voron

Cathy White

Jo Kennedy
Developer and

Adding events to the model



Start playing the next guide….

...if you aren’t playing it already

Event Recommendations

:play http://guides.neo4j.com/reco/file/04_events.html

Layered Recommendations

Layered recommendations

We want to make a recommendation based on multiple features, but we want to make some recommendation even if all those features aren't present.

We can achieve this by combining `WITH` and `OPTIONAL MATCH`.

OPTIONAL MATCH

The `OPTIONAL MATCH` clause is used to search for a pattern but if it doesn't find all of it then it will return `NULL` for the missing part.

OPTIONAL MATCH

The **`OPTIONAL MATCH`** clause is used to search for a pattern but if it doesn't find all of it then it will return **`NULL`** for the missing part.

`OPTIONAL MATCH` is similar to SQL's **`LEFT JOIN.`**

OPTIONAL MATCH

```
MATCH (member:Member {name: "Mark Needham"})
OPTIONAL MATCH (member)-[:MEMBER_OF]->(group)
RETURN member, COLLECT(group) AS groups
```

Continue with the guide

Continue playing the guide in your browser

Answers

Answers

Type the following command into the Neo4j browser query pane to see the answers:

```
:play http://guides.neo4j.com/reco/answers/4.html
```

Continue with the guide

Continue playing the guide in your browser

End of Module Event Recommendations

Questions ?

Location Based Recommendations

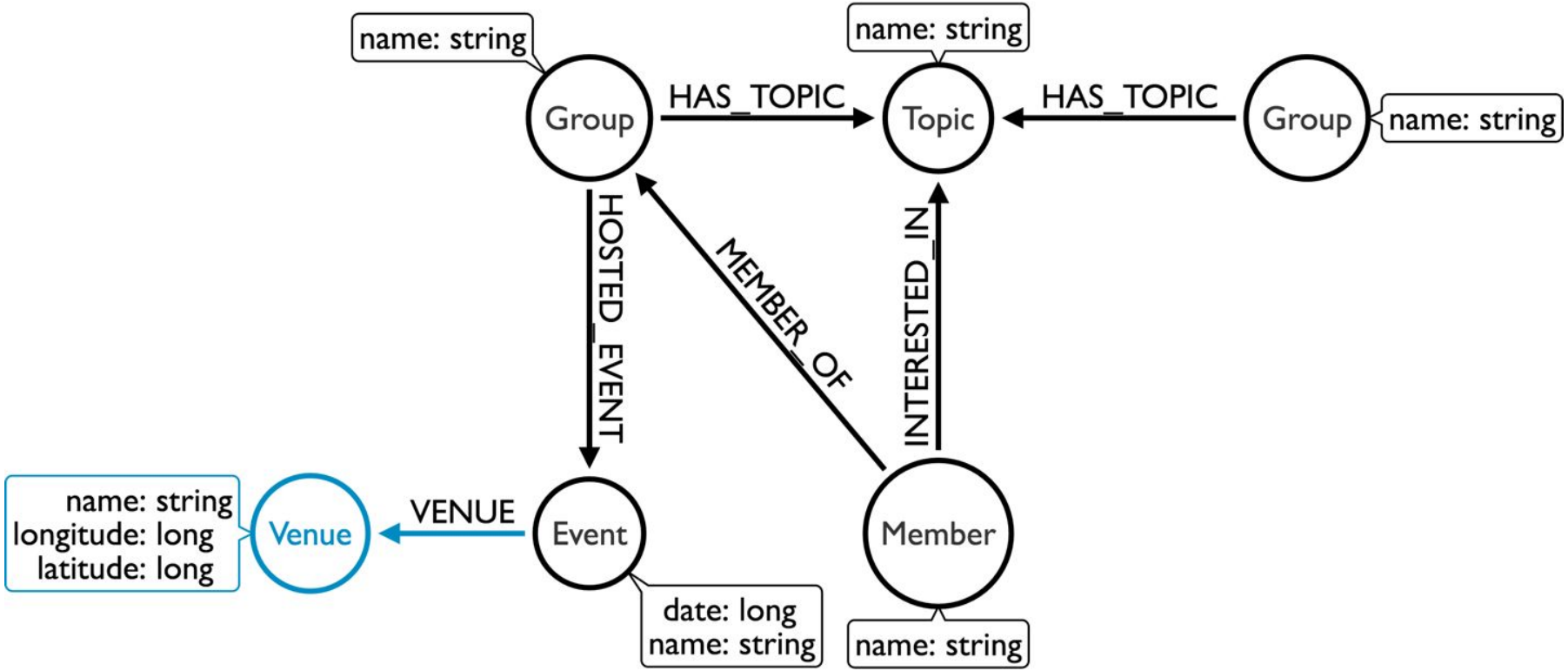
Location based recommendations

Location is very important when making recommendations, especially for meetups. When going to events after work people want to go somewhere nearby rather than trekking across the city.

What are we going to do?

- Add venues to the model
  - This time you're going to do the import!
- Recommend events that are close by

Adding venues to the model



Start playing the next guide….

...if you aren’t playing it already

Venues

:play http://guides.neo4j.com/reco/file/05_venues.html

Answers

Answers

Type the following command into the Neo4j browser query pane to see the answers:

`:play` http://guides.neo4j.com/reco/answers/5a.html

Distances

Calculating distances

The `distance` function lets us find the distance in metres between two points

Calculating distances

The `distance` function lets us find the distance in metres between two points

Points can be maps or nodes which contain the properties 'latitude' and 'longitude'.

```
RETURN distance(point1, point2)
```

Calculating distances

```
WITH point({latitude: 37.7952919, longitude: -122.4036433}) AS p1,
     point({latitude: 51.486323, longitude: -0.106042}) AS p2
RETURN distance(p1, p2)
```

```
8626170.522263126
```

Continue with the guide

Continue playing the guide in your browser

Answers

Answers

Type the following command into the Neo4j browser query pane to see the answers:

`:play` http://guides.neo4j.com/reco/answers/5b.html

Continue with the guide

Continue playing the guide in your browser

RSVPS

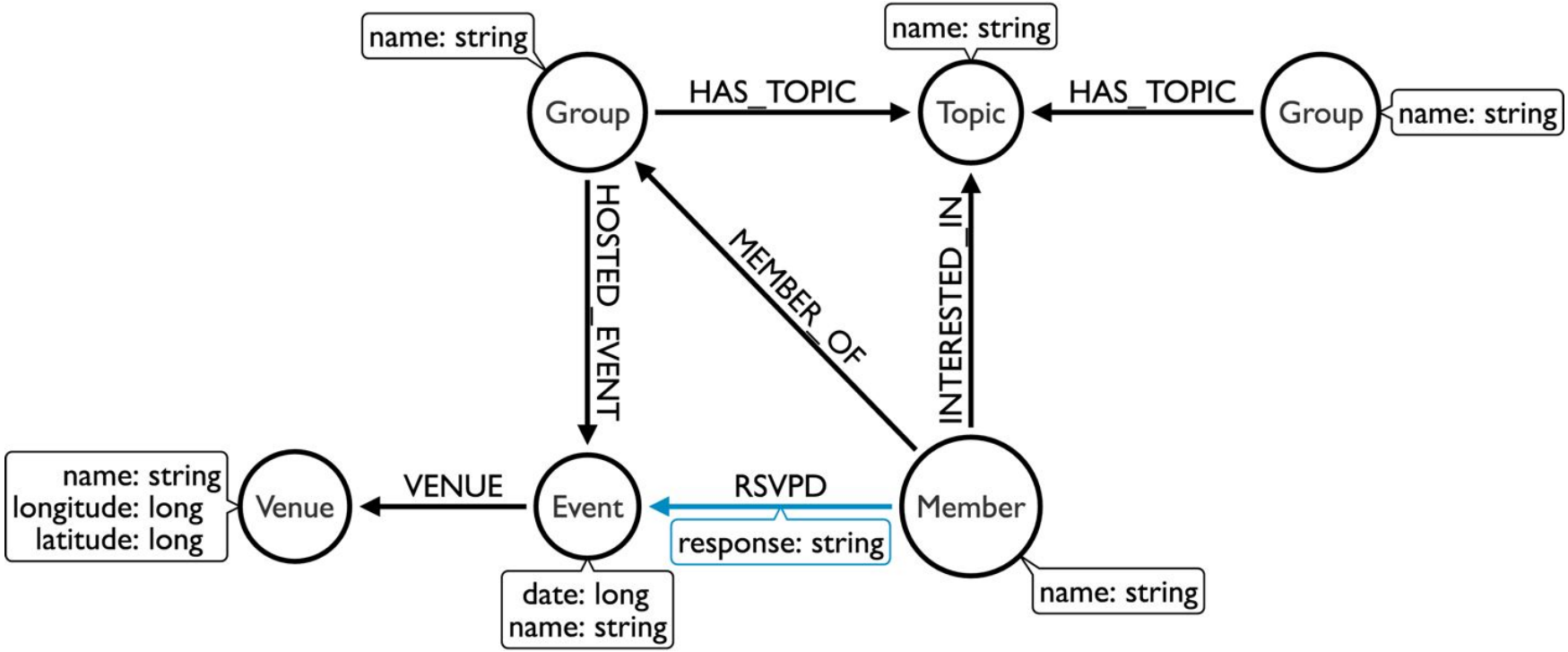
RSVPs

In this section we're going to look at the events that people have previously attended to help us make recommendations of what they might like to attend in future.

What are we going to do?

- Add RSVPs to the model
- Recommend events at groups whose events I've previously attended
- Recommend events at venues that I've previously attended

Adding RSVPs to the model



Start playing the next guide….

...if you aren’t playing it already

RSVPs

:play http://guides.neo4j.com/reco/file/06_rsvps.html

Answers

Answers

Type the following command into the Neo4j browser query pane to see the answers:

```
:play http://guides.neo4j.com/reco/answers/6.html
```

Continue with the guide

Continue playing the guide in your browser

Procedures

What are we going to do?

- Intro to Neo4j procedures
- Install the apoc (Awesome Procedures) library
- Explore the available procedures and use them in our queries

Procedures

Procedures were introduced in Neo4j 3.0 and allow us to write custom Java code, deploy it to the server, and then call it from Cypher queries.

Start playing the next guide….

neo4j

…if you aren’t playing it already

Procedures

:play http://guides.neo4j.com/reco/file/07_procedures.html

Answers

Exercise 7 Answers

Type the following command into the Neo4j browser query pane to see the answers:

```
:play http://guides.neo4j.com/reco/answers/7.html
```

Latent social graph

The latent social graph

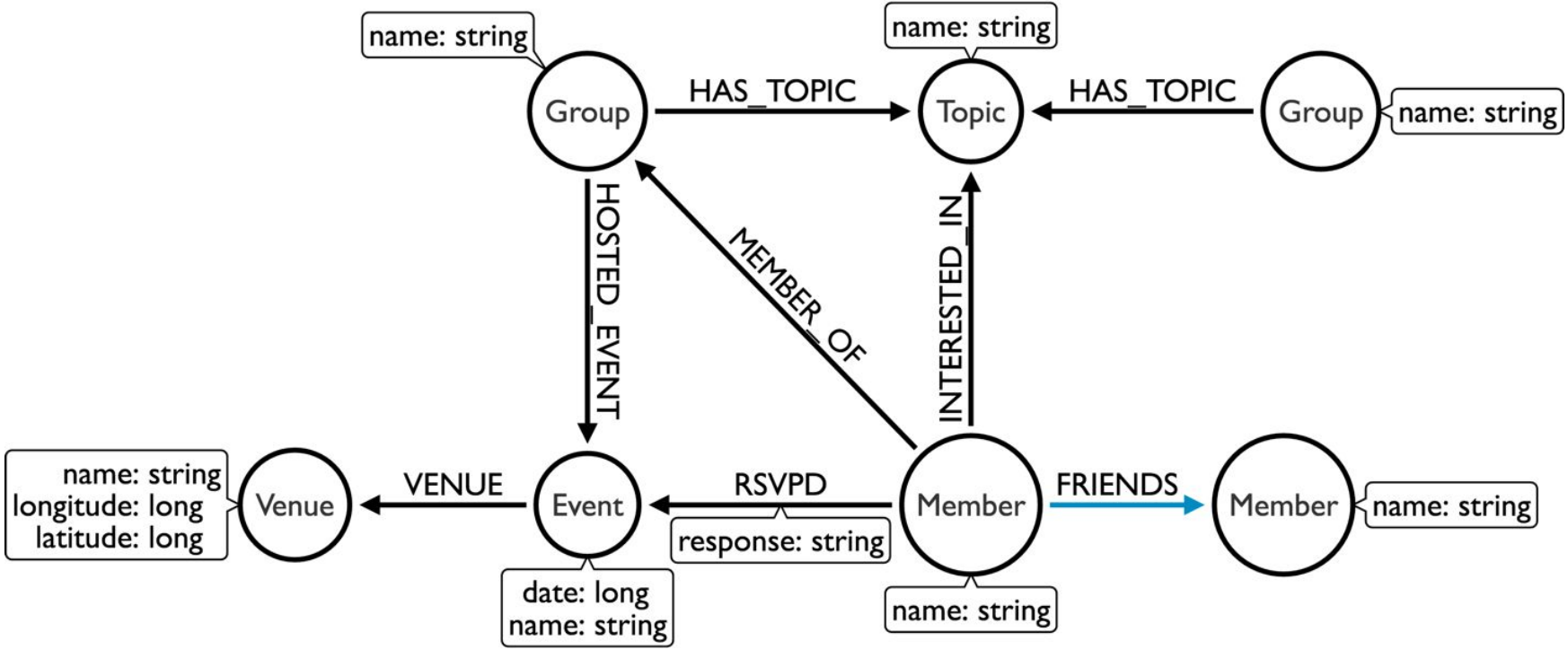
The meetup.com API allows you to find people's friends/followers on the Facebook and twitter social networks, but we actually have a potential friendship graph lurking within our graph.

People who've attended multiple same events probably know each other!

What are we going to do?

- Find the social graph hiding inside the data
- Materialise the Meetup social graph
- Recommend events that our friends are attending

Adding friends to the model



Start playing the next guide….

...if you aren’t playing it already

Latent Social Graph

:play http://guides.neo4j.com/reco/file/08_latent_social_graph.html

Transaction State

Periodic Commit

Cypher keeps **all transaction state in memory** while running a query which is fine most of the time.

But when refactoring the graph, this state can get **very large** and may result in an **OutOfMemory** exception.

Batching updates

We therefore need to take a batched approach to large scale refactorings.

```
MATCH (m1:Process)
WITH m1 LIMIT 1000
REMOVE m1:Process
WITH m1
// do the refactoring
```

Continue with the guide

It’s your turn to import the data

Answers

Exercise 8 Answers

Type the following command into the Neo4j browser query pane to see the answers:

```
:play http://guides.neo4j.com/reco/answers/8.html
```

Scoring Recommendations

Scoring recommendations

In the previous few sections we've been adding features to our recommendation but we haven't stopped to consider whether each of them should add equal weight to the recommendation.

In this section we'll look at how we can score our recommendations so that we show the user the best ones first.

What are we going to do?

- The pareto scoring function
- Score components of a recommendation
- Score our friendships

Start playing the next guide….

neo4j

...if you aren’t playing it already

Scoring

:play http://guides.neo4j.com/reco/file/09_scoring.html

Final Thoughts

How do we know if the recommendations are any good?

- If we're using recommendations in an application we'd need a process for checking the quality of the recommendations.
- Once we've collected that data we could then have a machine learning algorithm tell us what weights we should apply to each feature rather than guessing ourselves.

Thank you for coming

We hope you had a good day and picked up some ideas that you can use in your own application.

- Please fill out the feedback form: bit.ly/neo-survey
- Join the #training-recommendations Slack channel