



Graph Storage and Query: An Industrial Perspective

Dr. Jim Webber

Chief Scientist, Neo Technology

@jimwebber

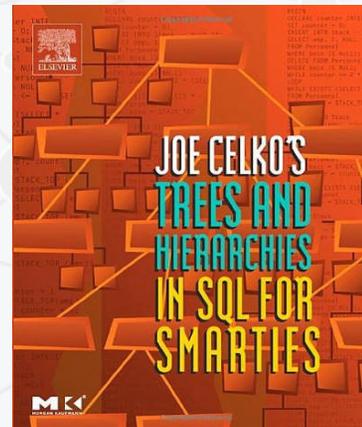
Common Perception:

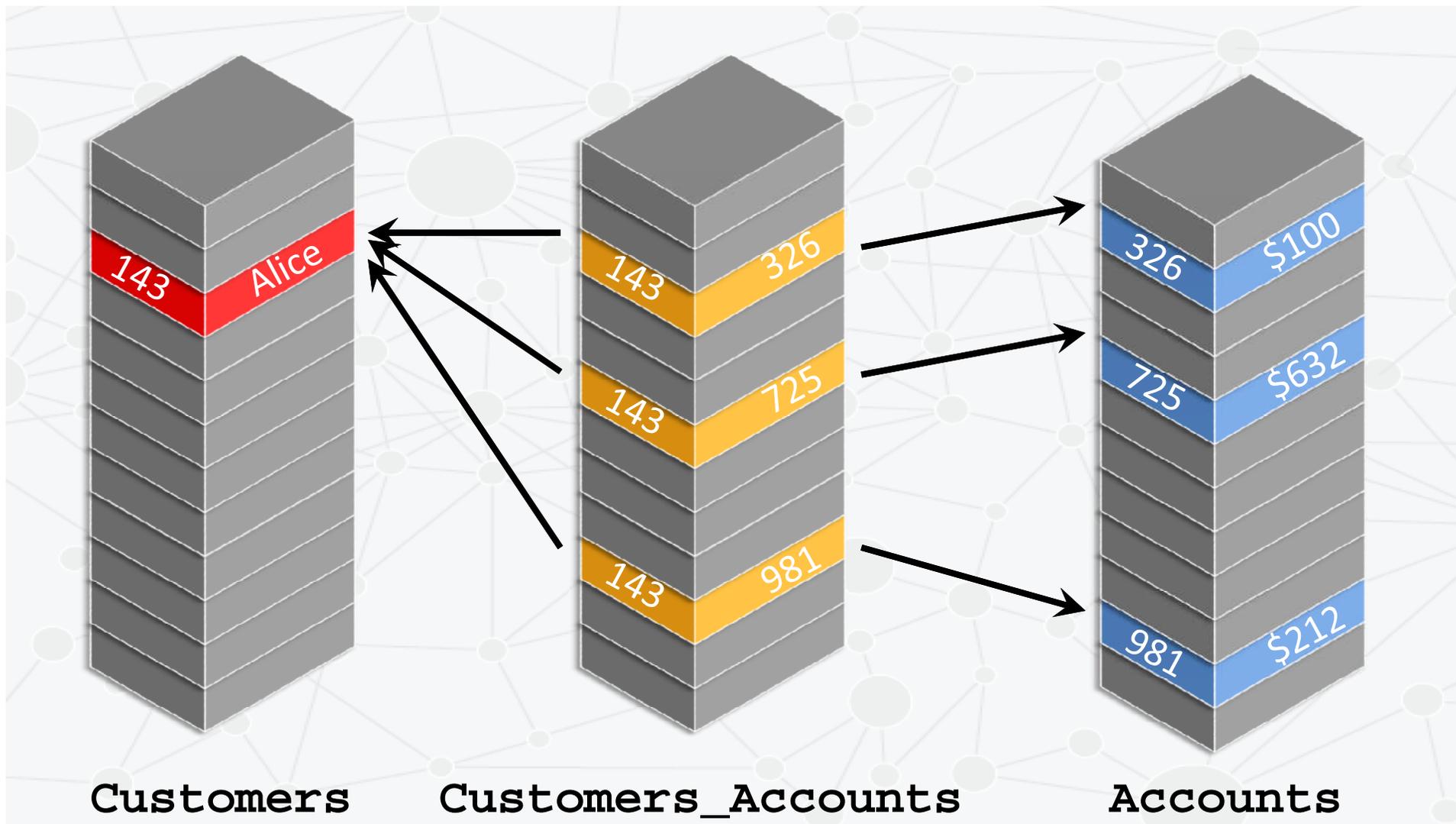
Data storage/query is a done deal

- Commercial developers already have the golden hammer of RDBMS
- Lots of academic research still in RDBMS systems
 - Query optimisation, “semantic” stores, etc
- Graph is just a subset of relational, so...

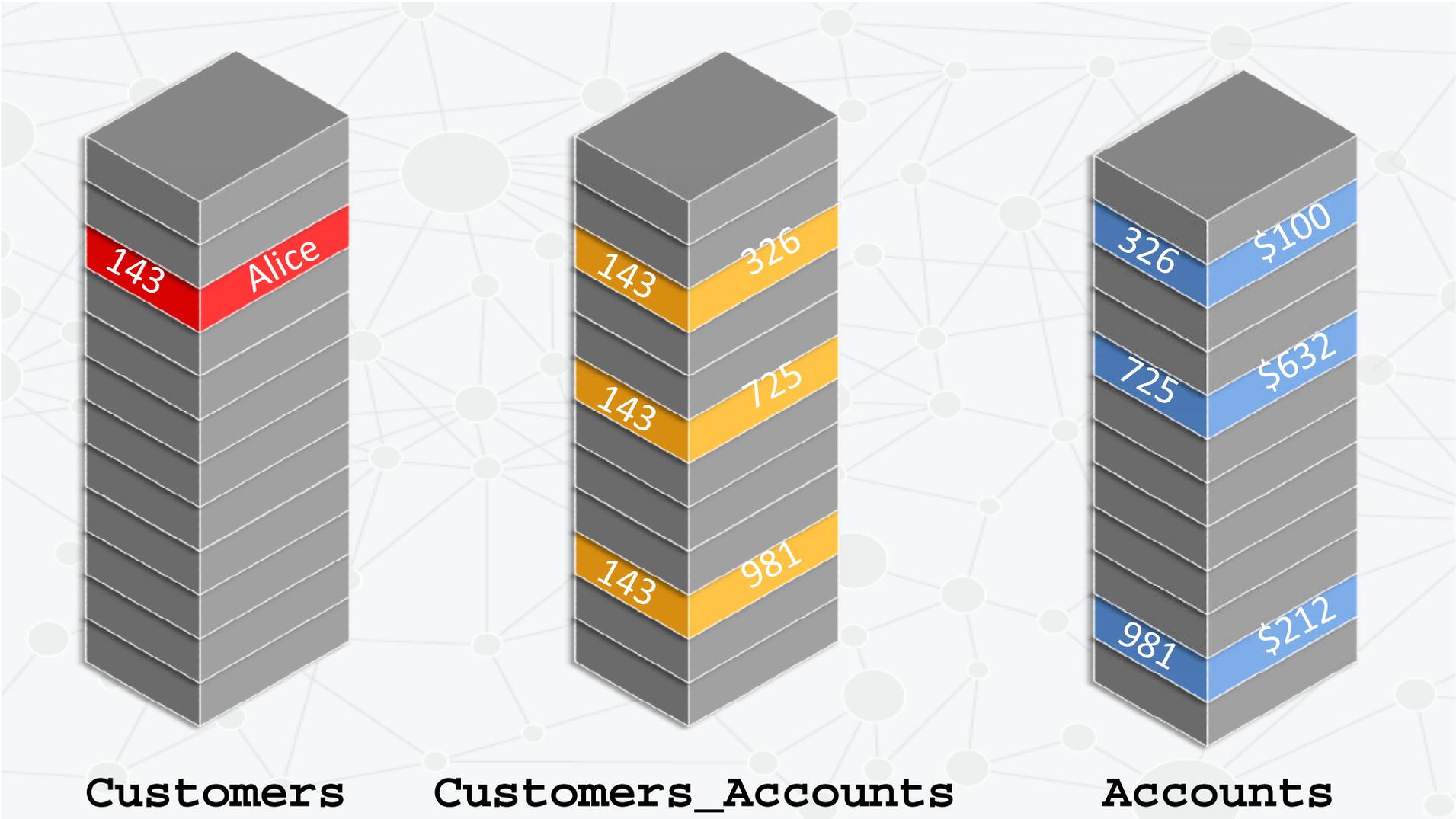
My RDBMS/triple store/etc already
does this

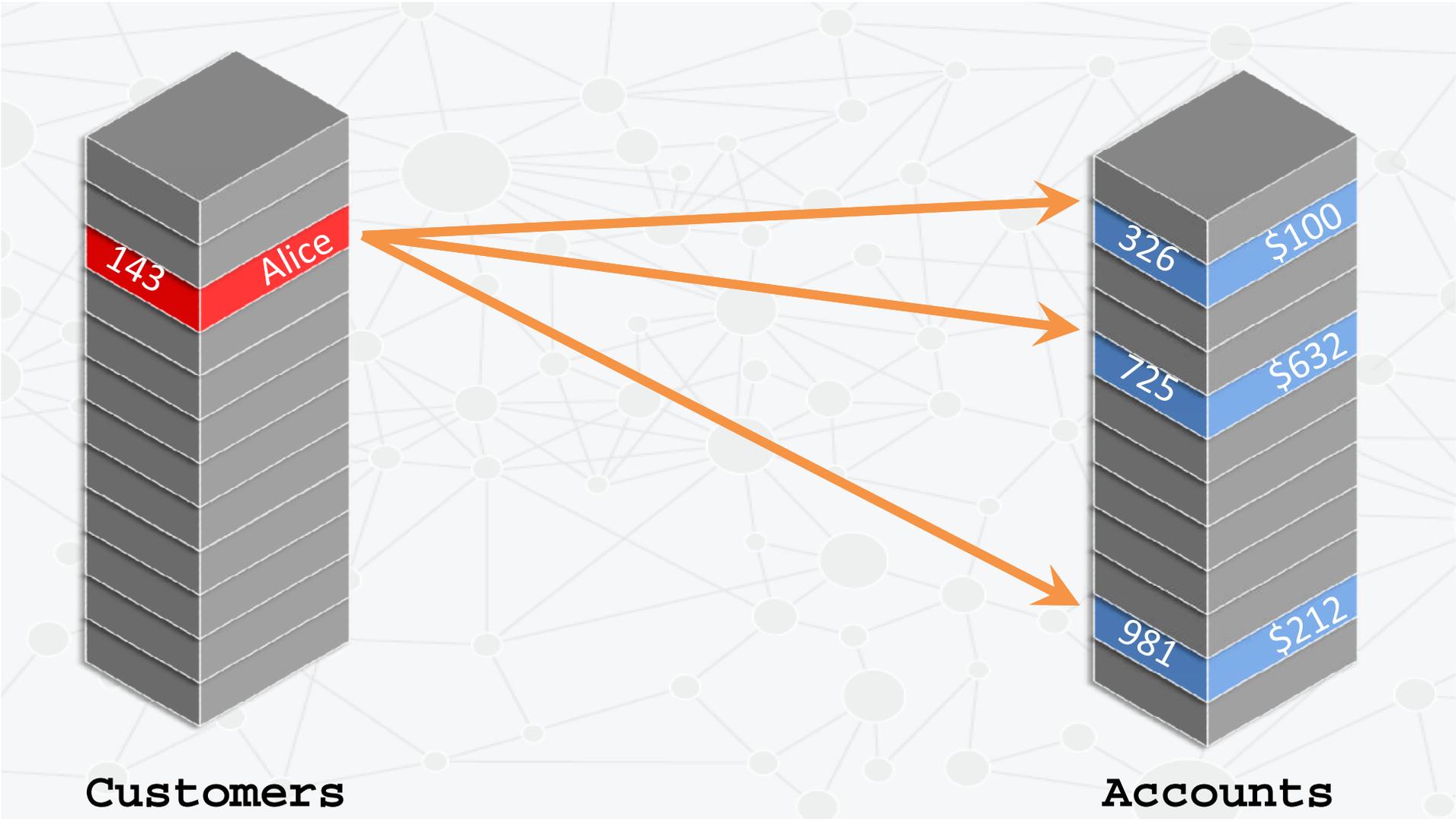
There's even a book
about it (sort of)

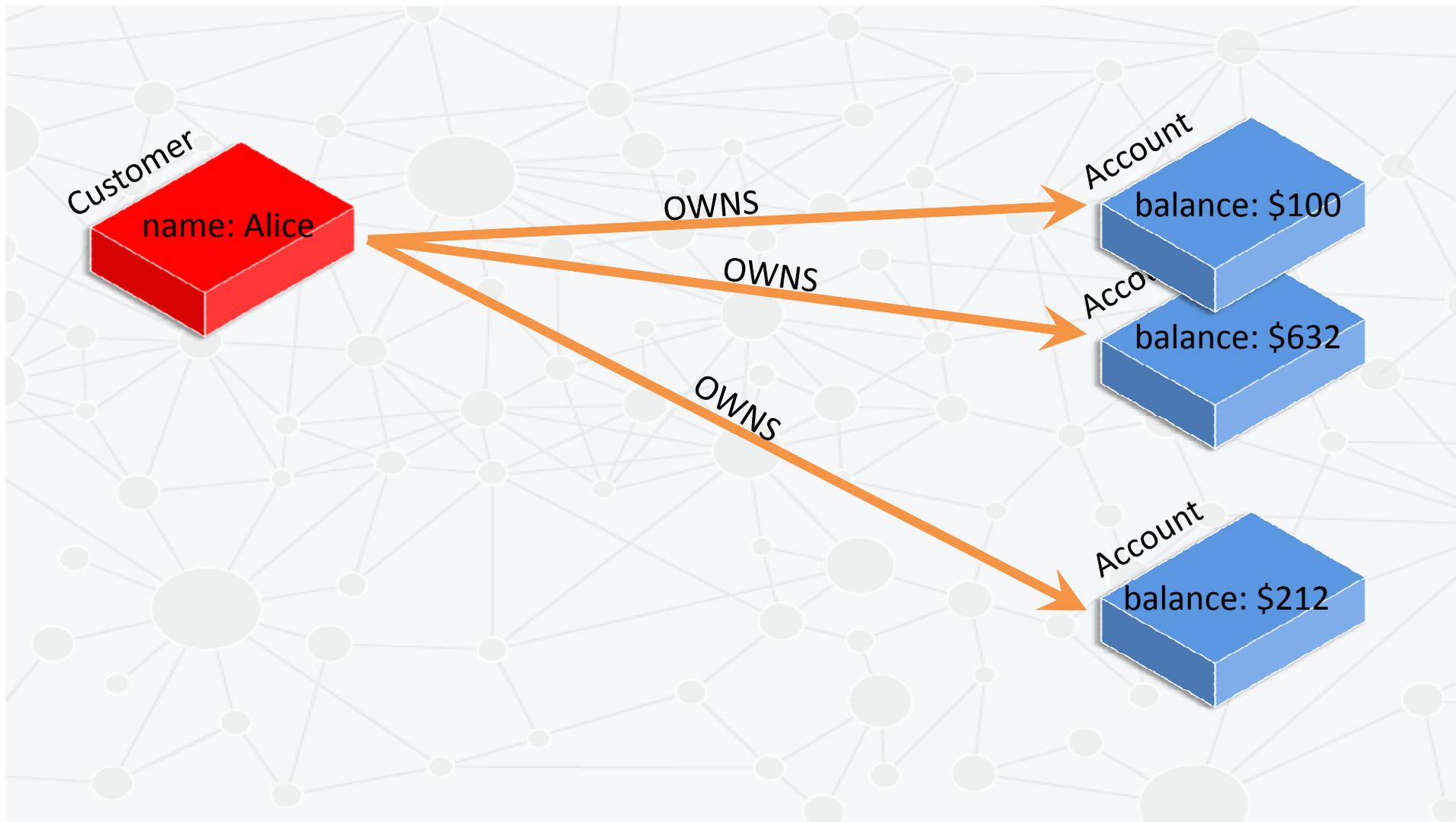


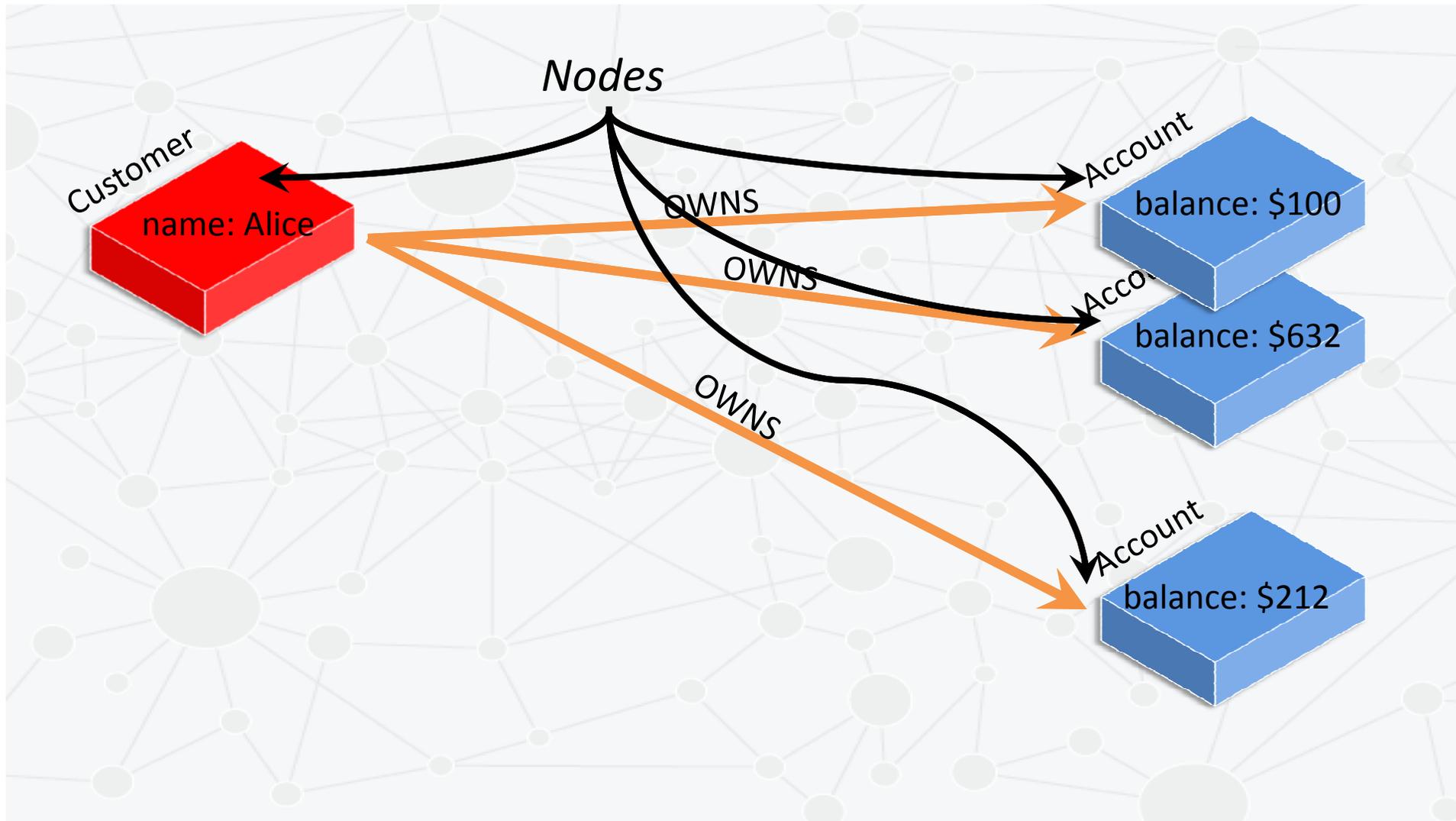


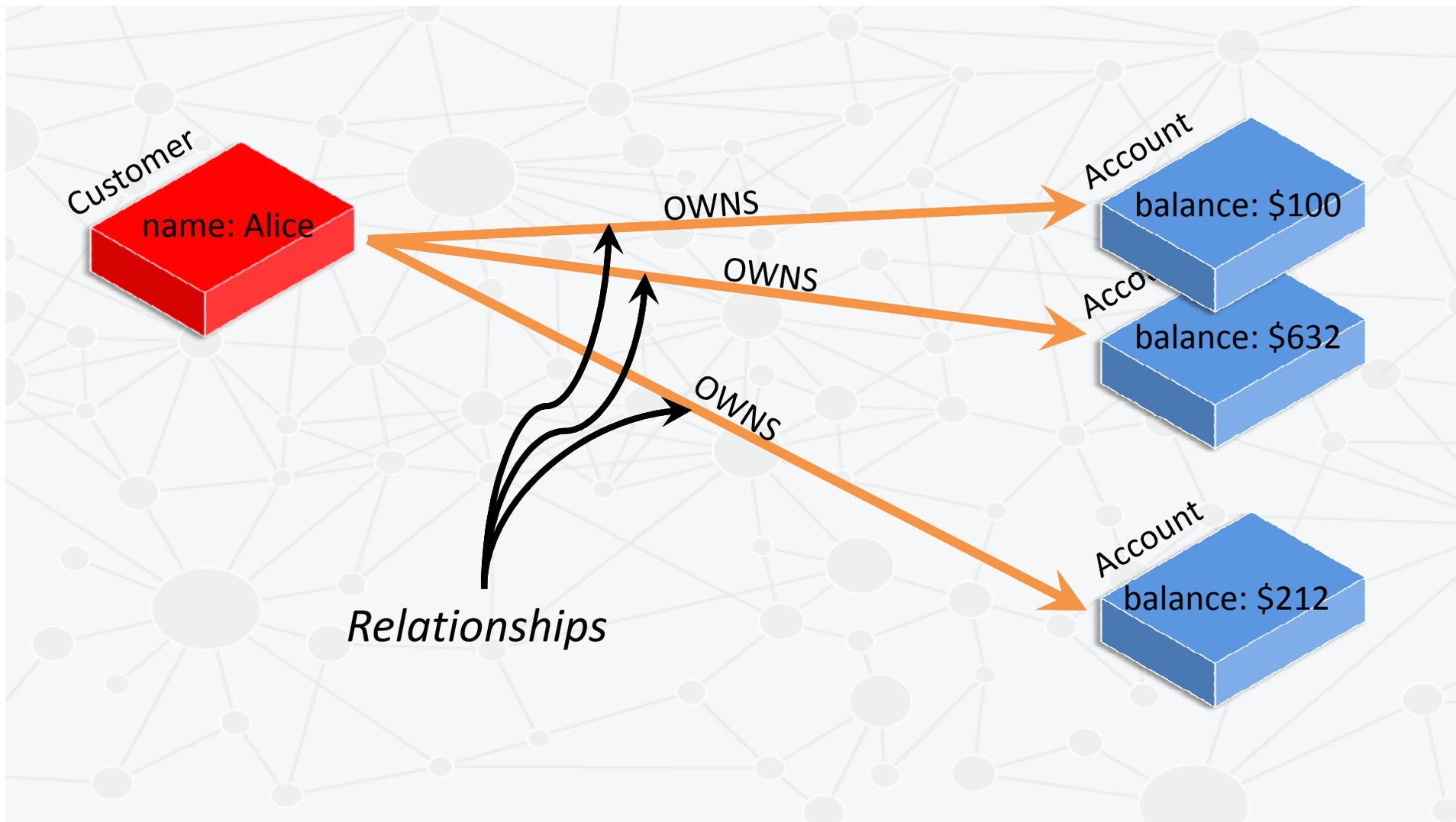


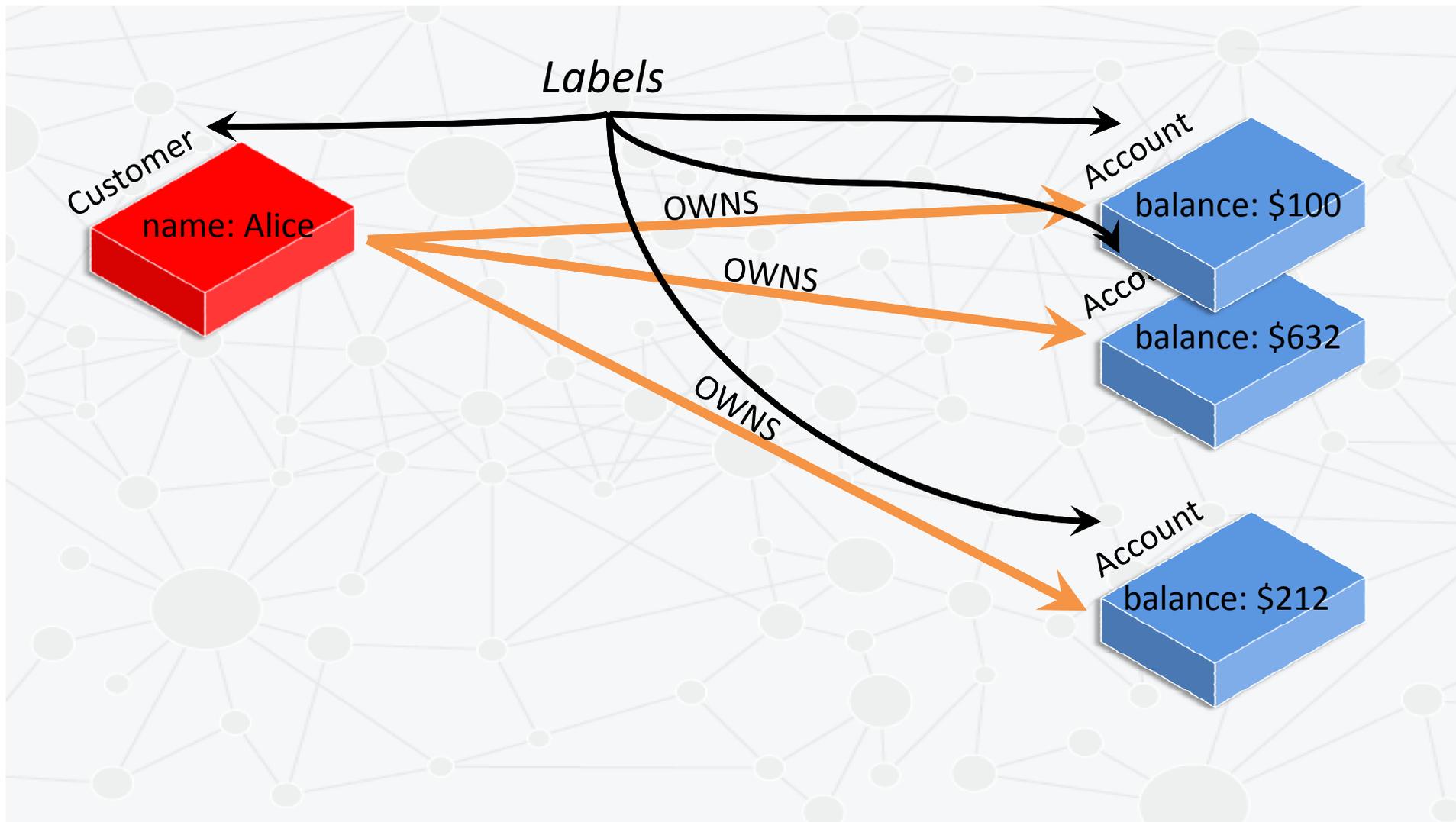






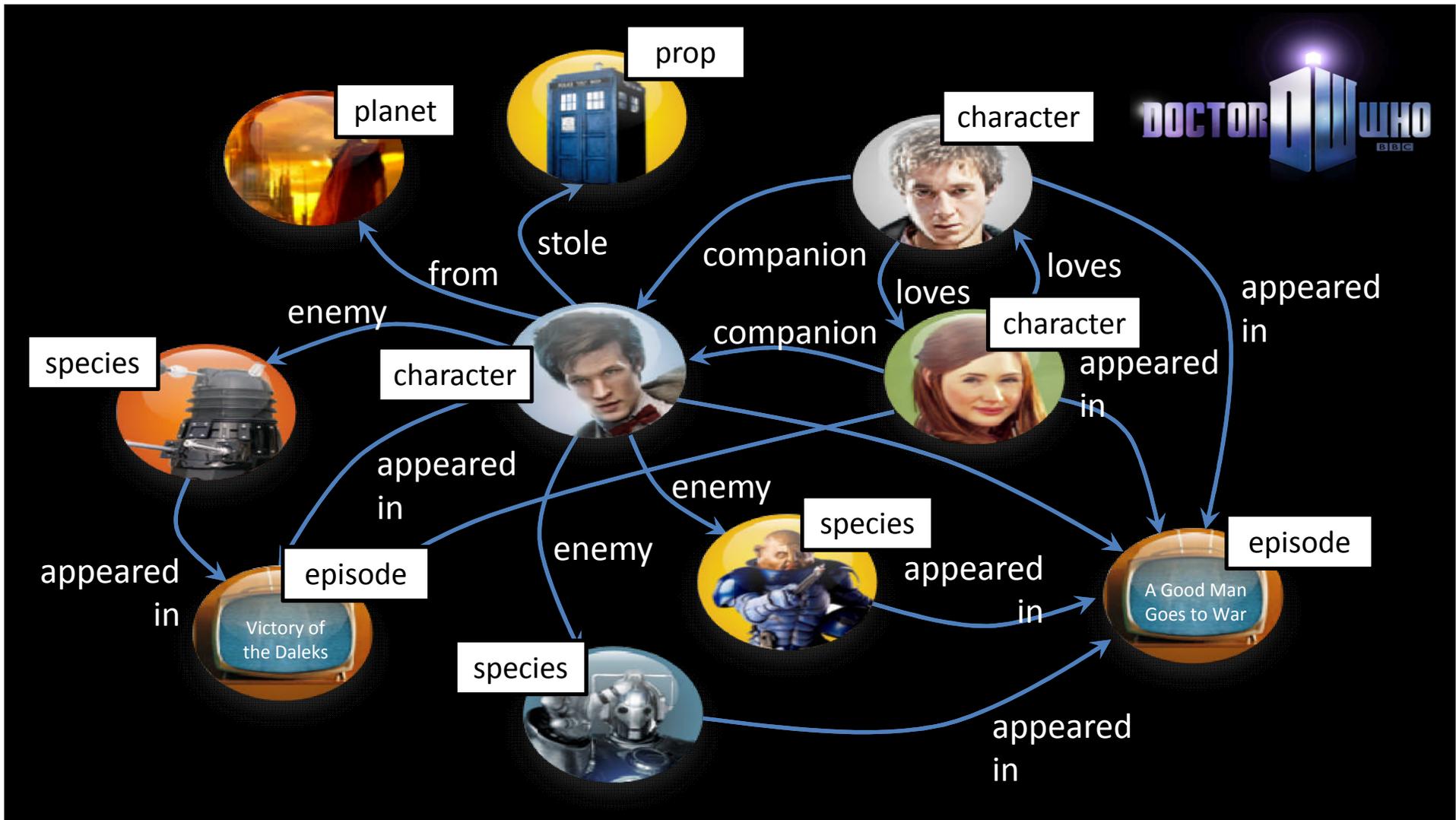






What's Neo4j?

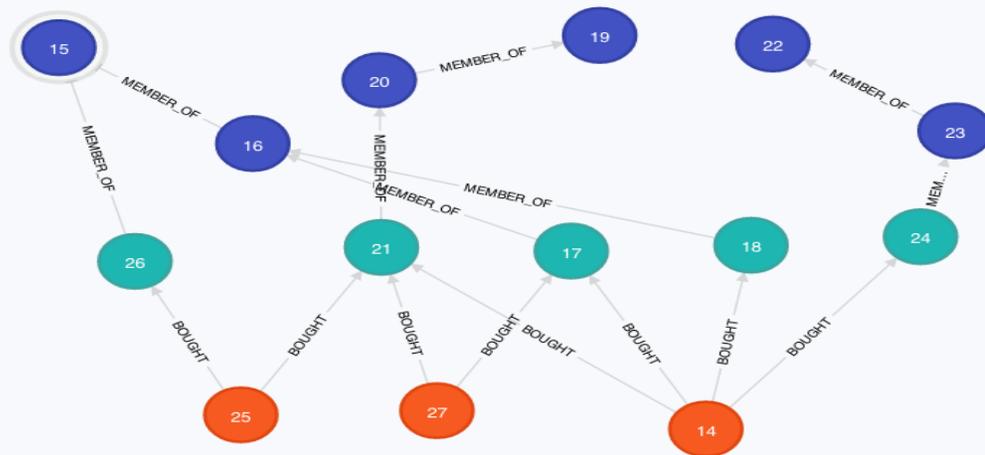
- It's a graph database
- OLTP-friendly (low latency)
- ACID transactional
- Clusterable
- Friendly query language - Cypher
- Programmatic APIs
 - For arbitrary graph algorithms



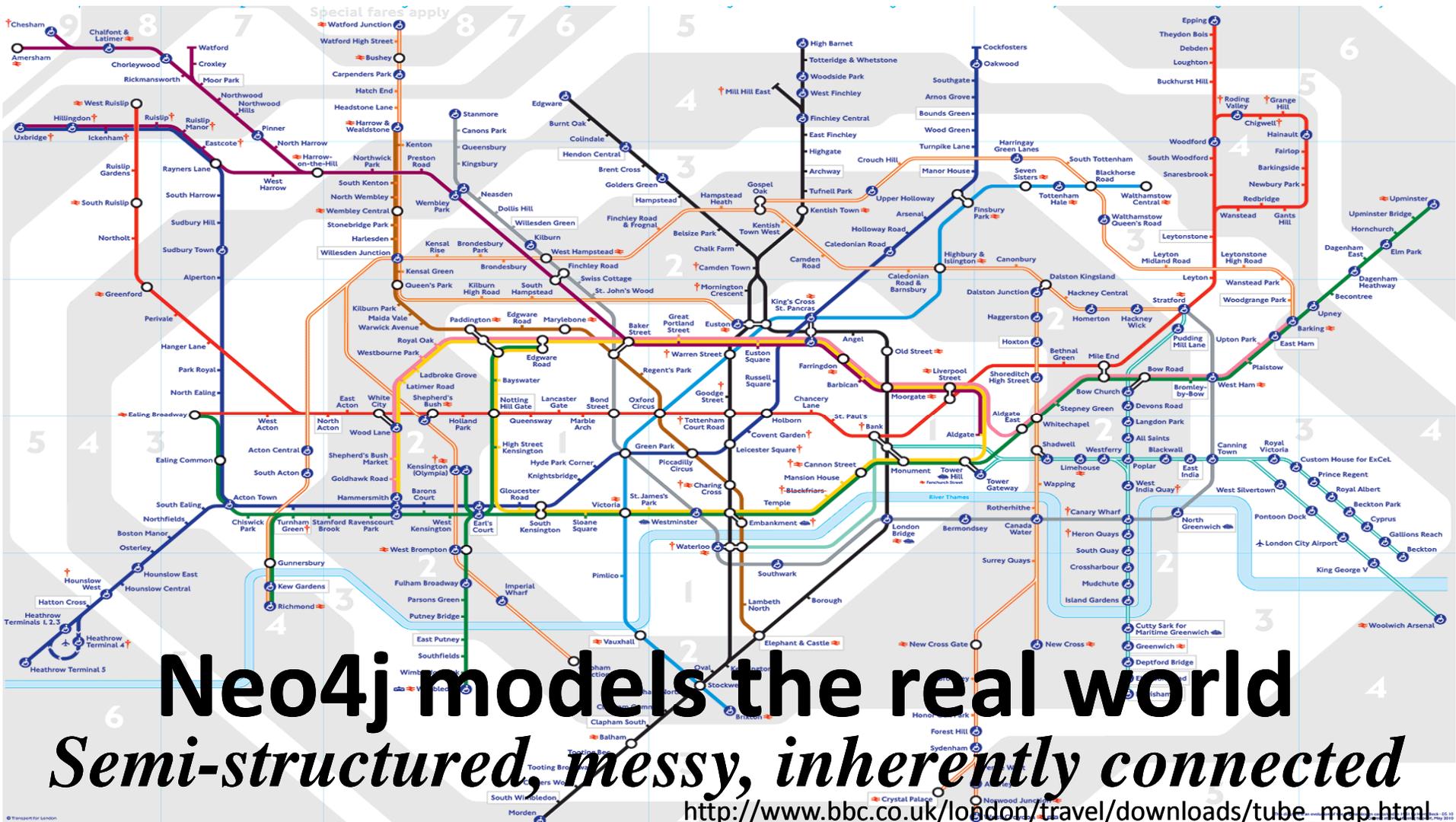
```
1 CREATE (daddy1:Person { name: 'Mickey Smith', dob: 19781006 })
2
3 CREATE (alcohol:Category { category : 'alcoholic drinks'})
4 CREATE (beer:Category { category : 'beer'})
5 CREATE beer-[:MEMBER_OF]->alcohol
6
```

CYPHER MATCH n RETURN n

- Default
- Customer
- Basket
- Product
- VirtualMachine
- OS
- Patch
- Owner
- Person
- Species
- Character
- Planet



Displaying 14 nodes, 16 relationships



Neo4j models the real world

Semi-structured, messy, inherently connected

http://www.bbc.co.uk/london/travel/downloads/tube_map.html

© Transport for London. Last updated: 15.11.2015



Neo4j supports *humane* queries with Cypher
(*Andres Taylor not always humane*)



Mechanical sympathy makes

Neo4j fast for graph traversals

Millions rels/sec/core on commodity hardware

<http://www.eachan.com/2010/09/mechanical-sympathy.html>



Neo4j is operationally boring
Which is a good thing!

<http://garethcase.com/casefiles/2013/04/B2B-Boring.jpg>



**Neo4j is an ACID
transactional database**
It values and protects your data

<http://foxandfoxfw.files.wordpress.com/2011/12/airbag1.jpg>

NICKYWIRE



Neo4j is optimised for graph structures

- Good graph databases exhibit *Index-free Adjacency*
 - Neubauer and Rodriguez, 2013
- We explicitly store:
 - Nodes
 - **Relationships**
 - Labels
 - Properties
- **Not** tables, **not** keys and values, **not** triples
 - Triples suck.
 - RDF is for the Web, not the database

But Research Challenges Remain

- Query planning and optimisation for Cypher
 - Cypher Algebra exists, pattern matcher constantly improving
- Very large scale mutable graphs
 - Transaction processing (OLTP)
 - Partitioned storage (domain-specific)
 - Performant distributed querying (caches! c.f. Beehive)
- I/O
 - Low-latency, safe storage to durable media
 - Low-contention, low resource use access to graph data
- Runtime Adaptation
 - Ensure fast traversals (gather heuristics from cache misses)
 - Graph-affined caching (shape, volatility, connectivity)

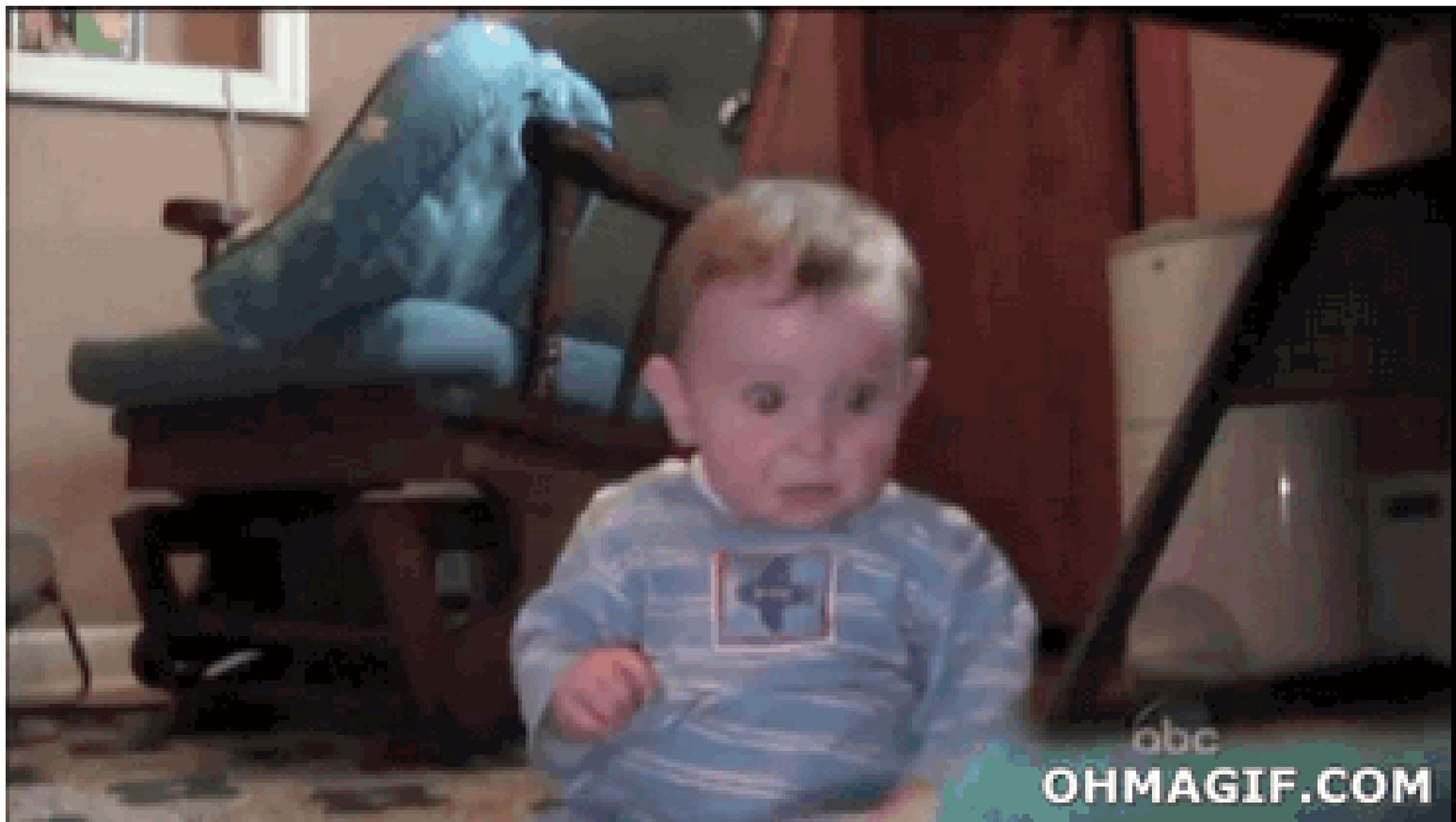


What do we need?

- We're (mostly) consumers of research right now
- We need research on:
 - Query language design/optimisation
 - Efficient data structures for graph storage/query/localised indexing
 - Highly available (ACID) transaction/coordination protocols
 - Very large scale graph storage/query
 - Data structures
 - Efficient I/O
 - Etc.

Better tools/perspectives

- Relational algebra not very useful
 - We treat graph storage/query more like data structure/algorithm analysis
 - Except in our query language team
- RDF is **not** gaining much traction commercially
 - Triple model is far less expressive than labelled property graph model
 - http://googlefight.com/index.php?lang=en_GB&word1=%22Resource+Description+Framework%22&word2=neo4j



DE MUNDI CREATIONE, RERUM

& ornatu; deque hominis formatione, etc.



In principio creavit Deus
caelum, & terram.

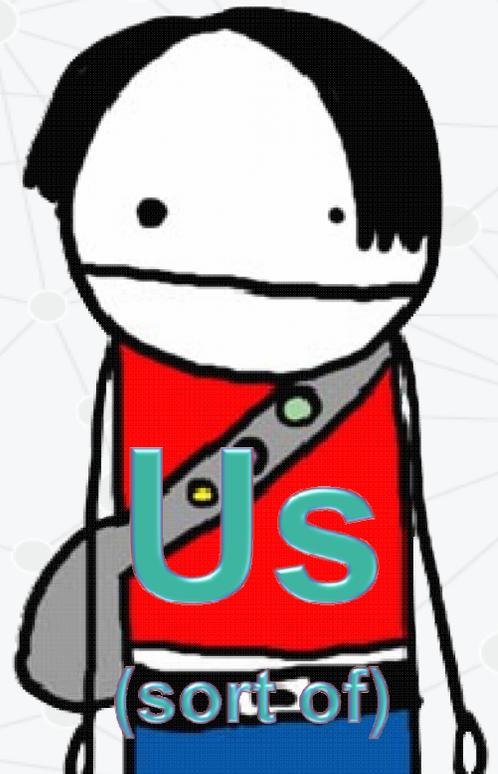
2. Terra autem erat inanis
& vacua, & tenebrae
erant super faciem abyssi:
& spiritus Dei ferebatur

Academia

4. Et vidit Deus lucem quod esset bona: &
divisit lucem a tenebris.

6. Appellavitque lucem Diem, & tenebras

HIPSTER



Us
(sort of)

Neo4j Tutorial

Wednesday @ 15:45

- Neo4j is the leading commercial graph database
 - It's open source, so you can see how it works
- Hands-on, practical tutorial
 - Bring your laptop
 - Focus on Cypher query language
- **Free Teaching materials** and **free text book** you can use



Thanks for listening
@jimwebber

Free Full O'Reilly eBook!
<http://graphdatabases.com>

