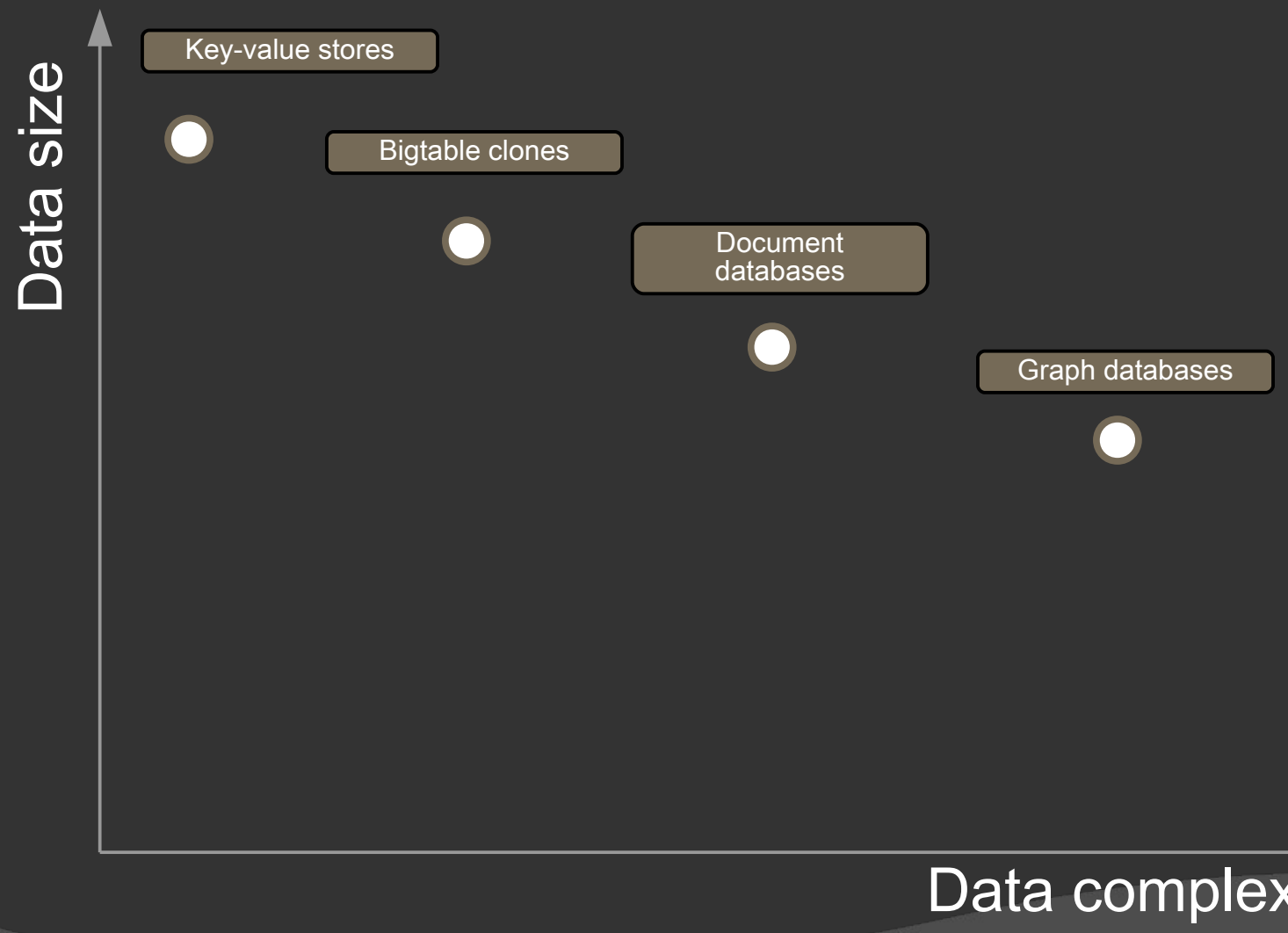
Big Data, Complex Data

Managing Data and Complexity in Graph Databases

Peter Neubauer
Neo Technology

#neo4j
@peterneubauer
peter@neotechnology.com

NOSQL data models



Neo4j
the graph database

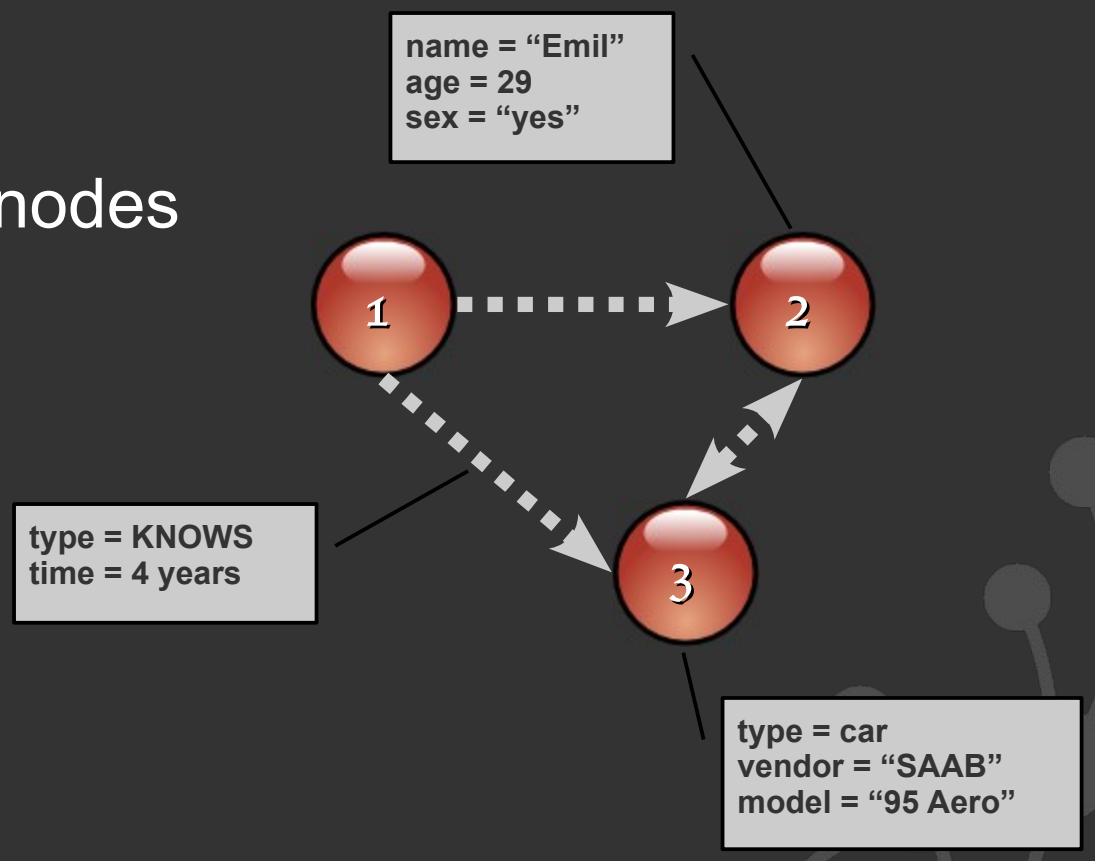
neotechnology
value in relationships

- Full Open Source, Java
  - Embeddable and server based
  - Index-free traversal of data on disk level
  - ACID, JTA compliant
  - Transactional Indexing framework
- 24/7 since 2003
- High Availability clustering support
- Great community
- Great bindings (Jruby, Python, REST, C#, PHP, Gremlin)
- Spring Data for access from Spring framework
- Cloud deployments: AWS, Azure, Cloudfoundry(VMWare), Heroku

The Neo4j model: Property Graph

Core abstractions:

- Nodes
- Relationships between nodes
- Properties on both



Building a node space (core API)

```
GraphDatabaseService graphDb = ... // Get factory


// Create Thomas 'Neo' Anderson
Node mrAnderson = graphDb.createNode();
mrAnderson.setProperty( "name", "Thomas Anderson" );
mrAnderson.setProperty( "age", 29 );

// Create Morpheus
Node morpheus = graphDb.createNode();
morpheus.setProperty( "name", "Morpheus" );
morpheus.setProperty( "rank", "Captain" );
morpheus.setProperty( "occupation", "Total bad ass" );

// Create a relationship representing that they know each other
mrAnderson.createRelationshipTo( morpheus, RelTypes.KNOWS );
// ...create Trinity, Cypher, Agent Smith, Architect similarly
```

Building a node space

```
GraphDatabaseService graphDb = ... // Get factory
Transaction tx = graphdb.beginTx();

// Create Thomas 'Neo' Anderson
Node mrAnderson = graphDb.createNode();
mrAnderson.setProperty( "name", "Thomas Anderson" );
mrAnderson.setProperty( "age", 29 );

// Create Morpheus
Node morpheus = graphDb.createNode();
morpheus.setProperty( "name", "Morpheus" );
morpheus.setProperty( "rank", "Captain" );
morpheus.setProperty( "occupation", "Total bad ass" );

// Create a relationship representing that they know each other
mrAnderson.createRelationshipTo( morpheus, RelTypes.KNOWS );
// ...create Trinity, Cypher, Agent Smith, Architect similarly
tx.commit();
```

Code (2): Traversing a node space

```
// Instantiate a traverser that returns Mr Anderson's friends
Traverser friendsTraverser = mrAnderson.traverse(
    Traverser.Order.BREADTH_FIRST,
    StopEvaluator.END_OF_GRAPH,
    ReturnableEvaluator.ALL_BUT_START_NODE,
    RelTypes.KNOWS,
    Direction.OUTGOING );

// Traverse the node space and print out the result
System.out.println( "Mr Anderson's friends:" );
for ( Node friend : friendsTraverser )
{
    System.out.printf( "At depth %d => %s%n",
        friendsTraverser.currentPosition().getDepth(),
        friend.getProperty( "name" ) );
}
```

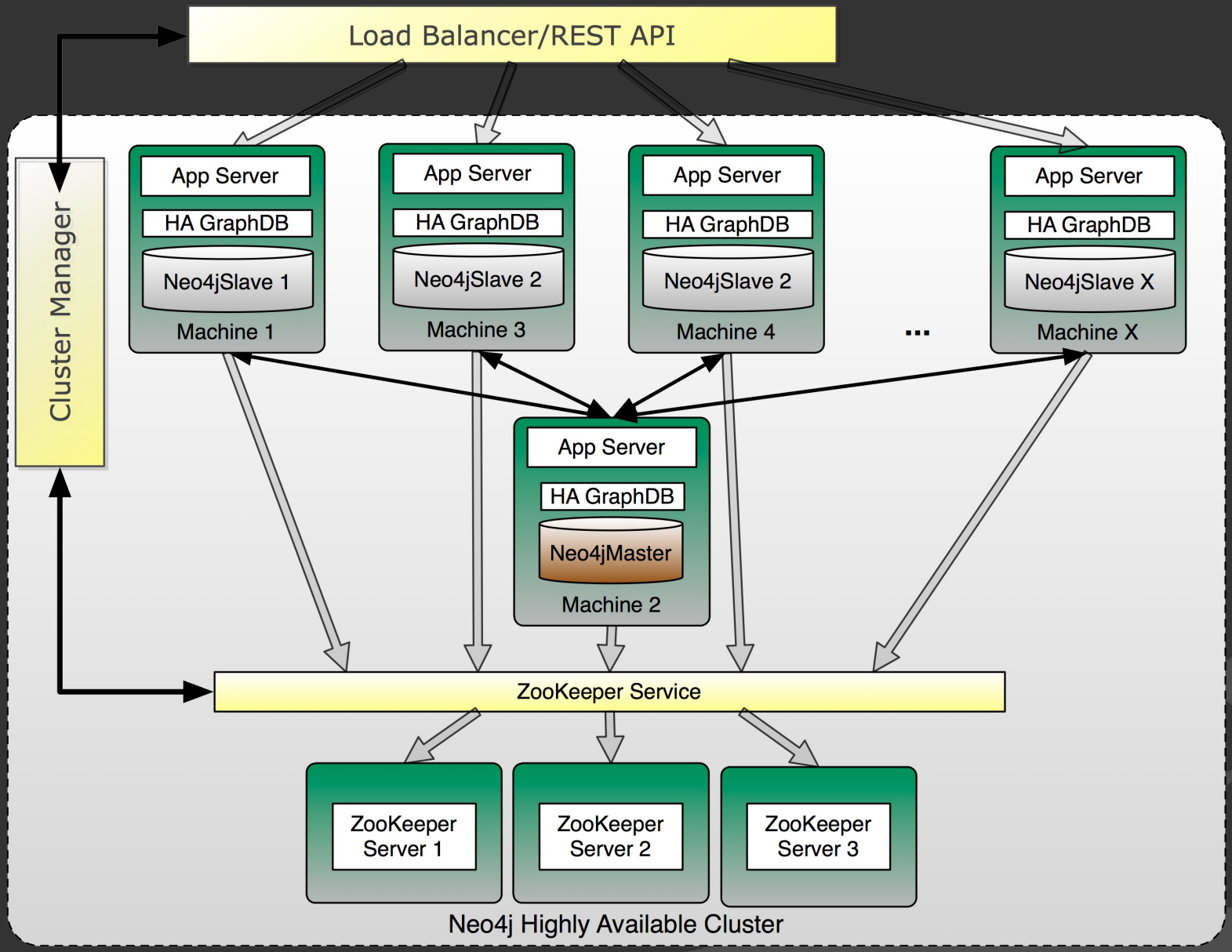
Ruby

```
gem install neo4j

require "rubygems"
require 'neo4j'

class Person
  include Neo4j::NodeMixin
  property :name, :age, :occupation
  index :name
  has_n :friends
end

Neo4j::Transactoin.run do
  neo = Person.new :name=>'Neo', :age=>29
  morpheus = Person.new :name=>'Morpheus', :occupation=>'badass'
  neo.friends << morpheus
end

neo.friends.each {|p|...}
```

Neo4j 1.3 news

- GPL Community Edition
- 128 Billion primitives address space
- Spring Data integration
- Short string → long
- Web visualization
- Server plugins for moving code to data
- HA improvements

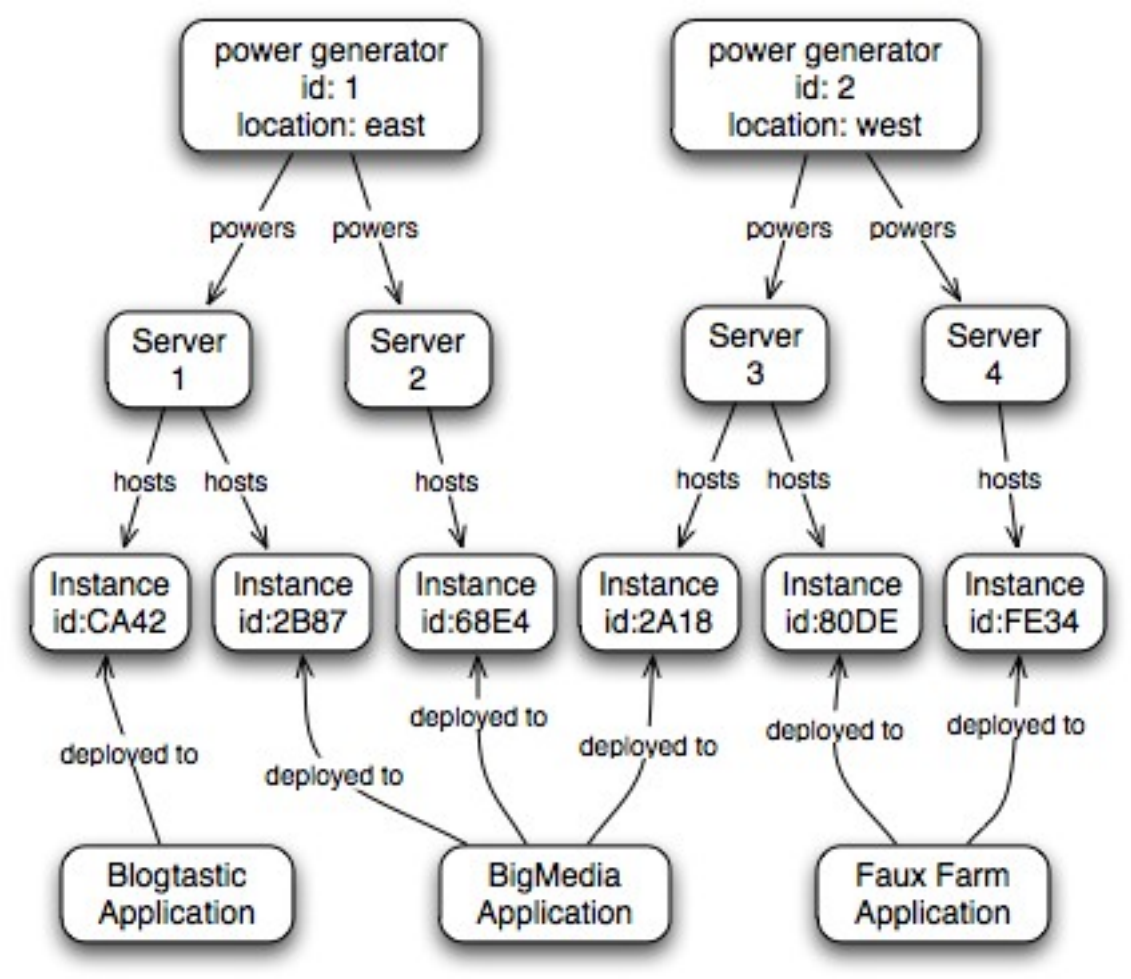
Neo4j High Availability



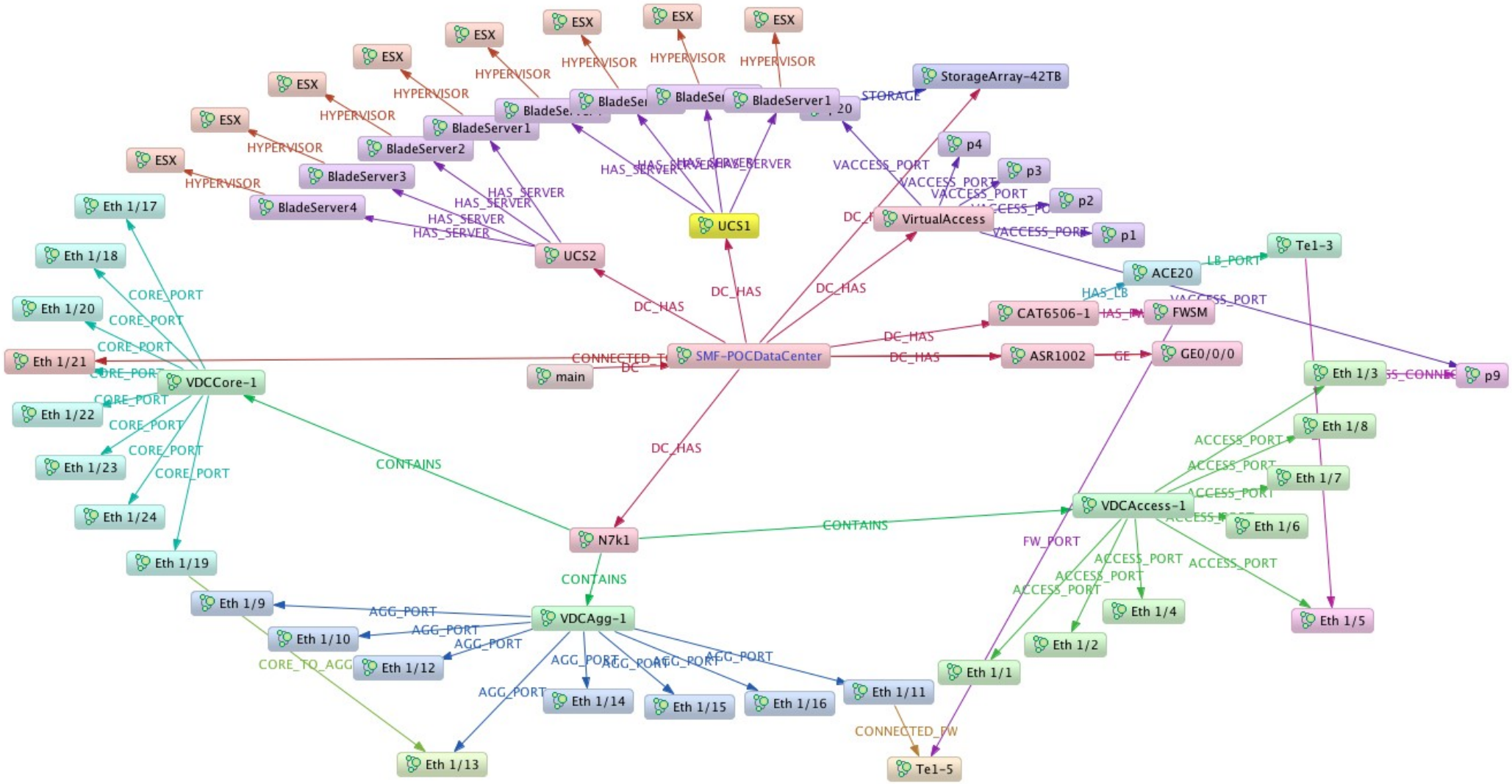
Neo4j – what do do people with it?

- Complex and connected data
  - >> 1M nodes
  - Realtime analytics on transactional data (< 1s)
- Network and Cloud Management
- Master Data Management
- Social applications (recommendations)
- Finance, Insurance (Fraud detection, financial BI)
- Spatial, Geolocation services
- Other
  - Bioinformatics, genomics
  - Semantic applications (RDF)
  - Intelligence

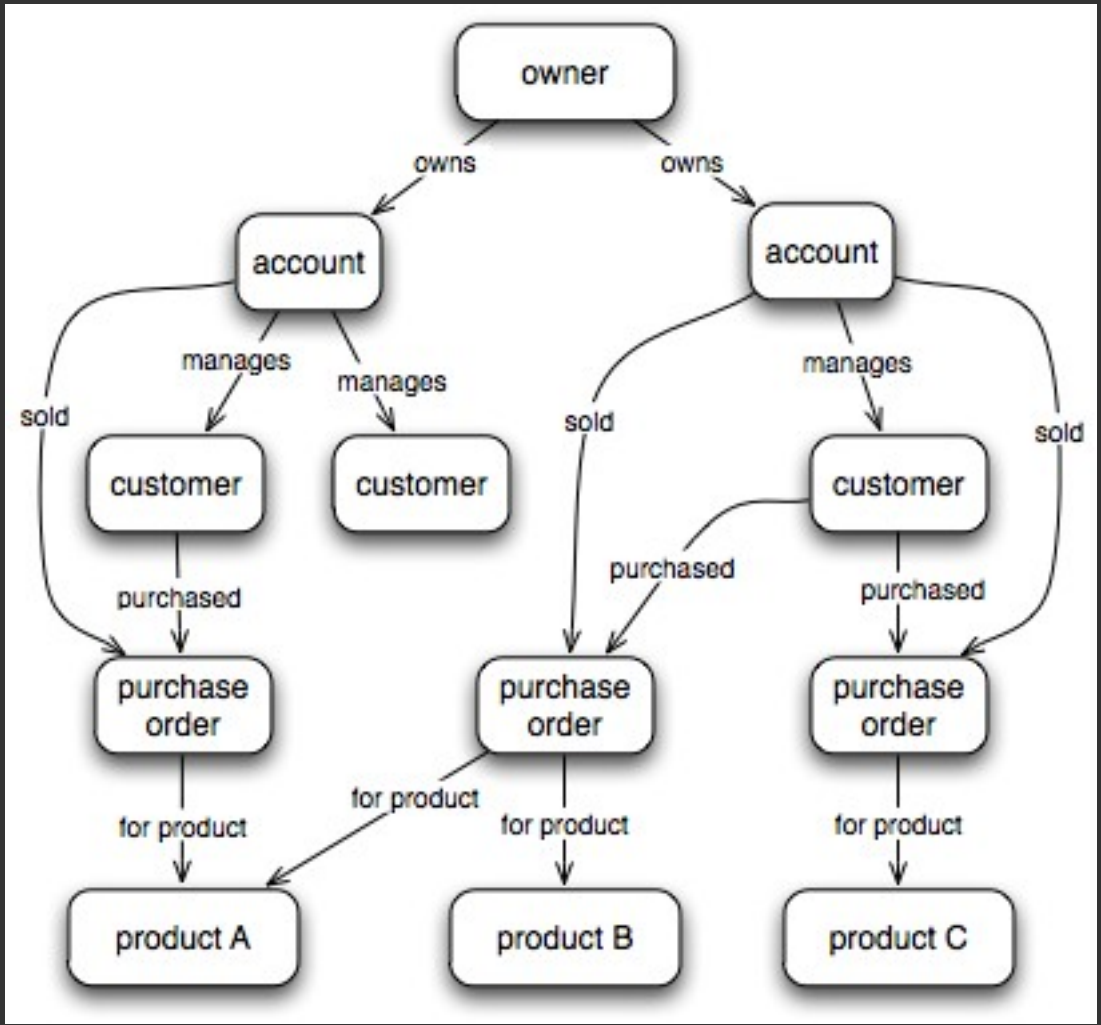
Impact Analytics, CMDB, Network Management, Provisioning



Impact Analytics, CMDB, Network Management, Provisioning

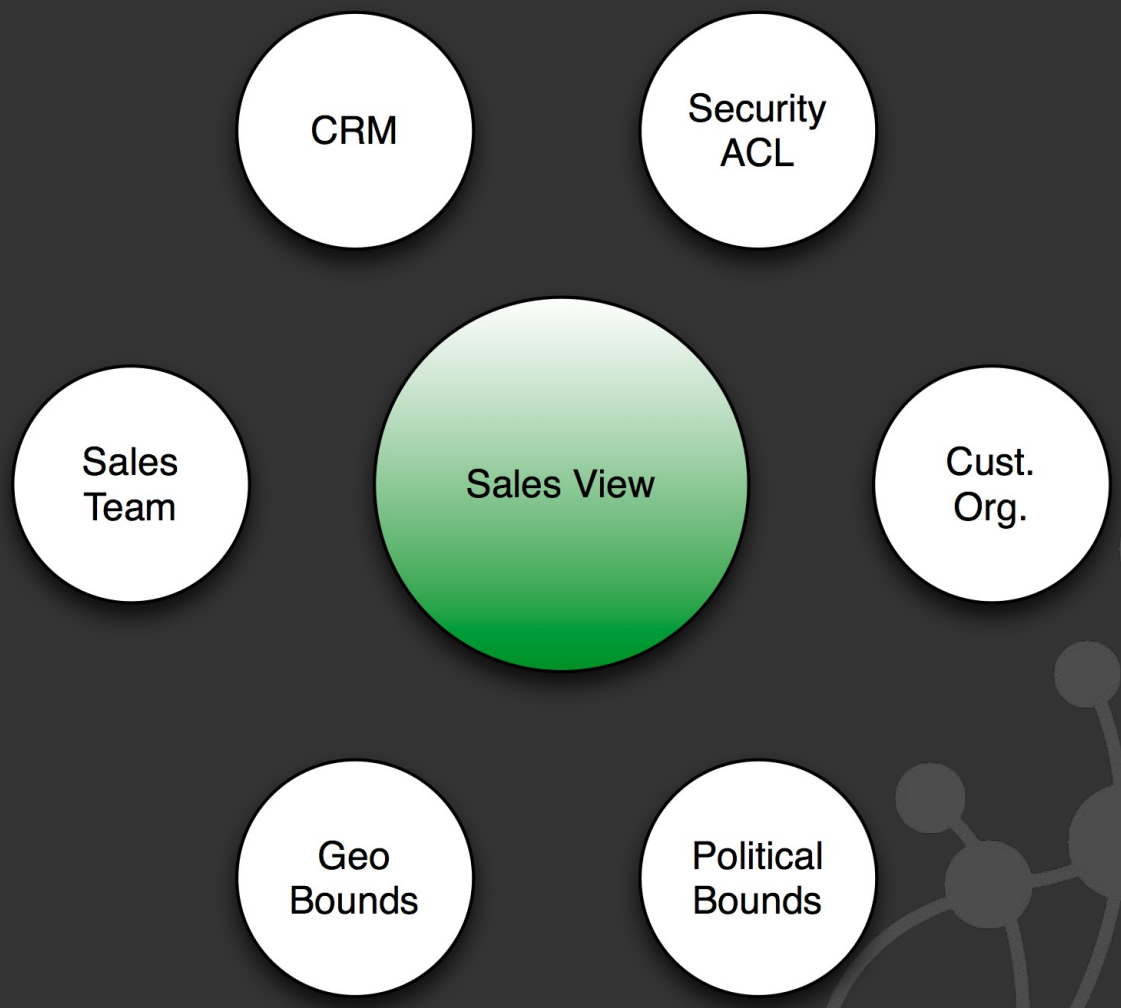


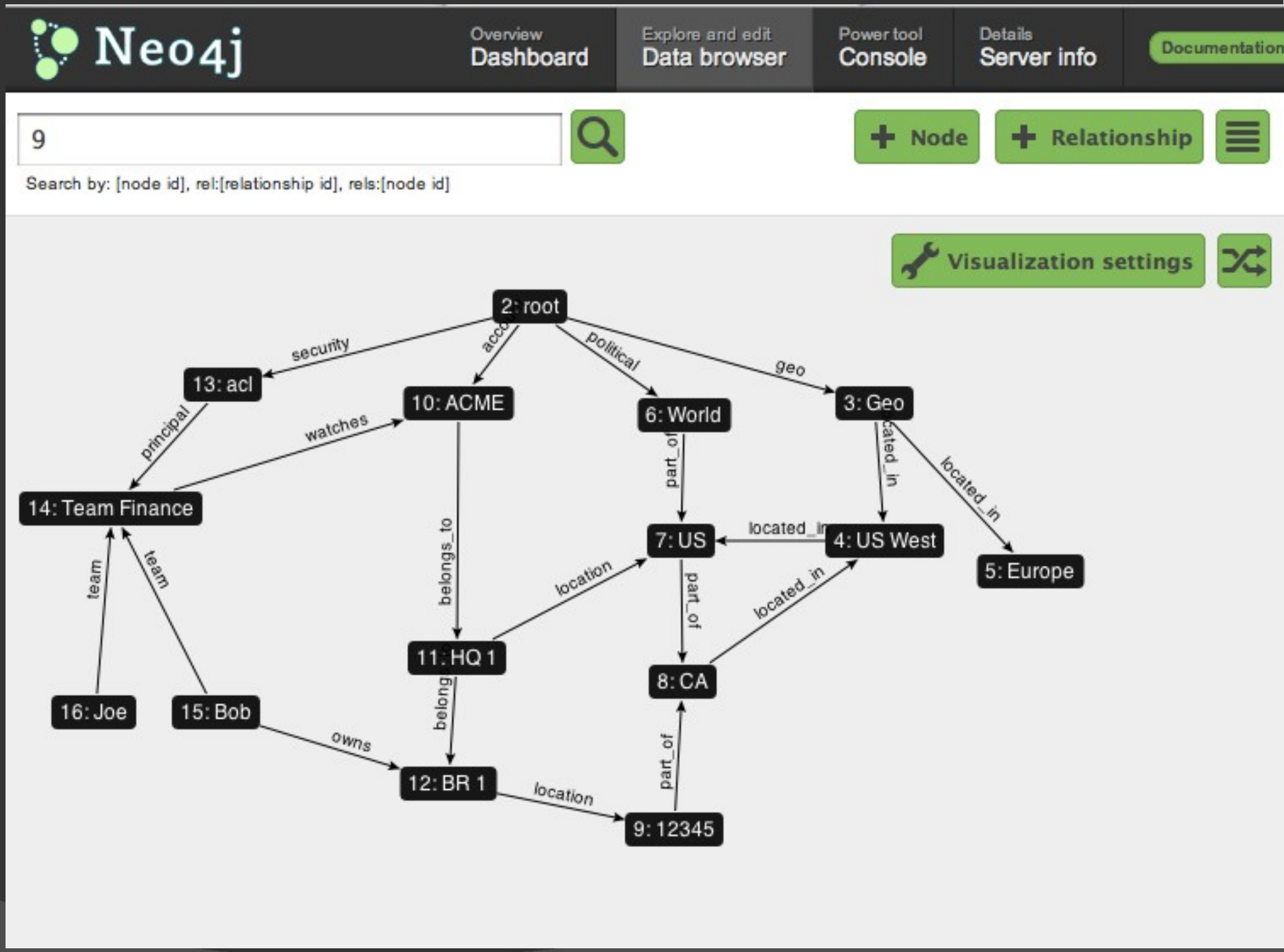
Master Data Management



Customer MDM

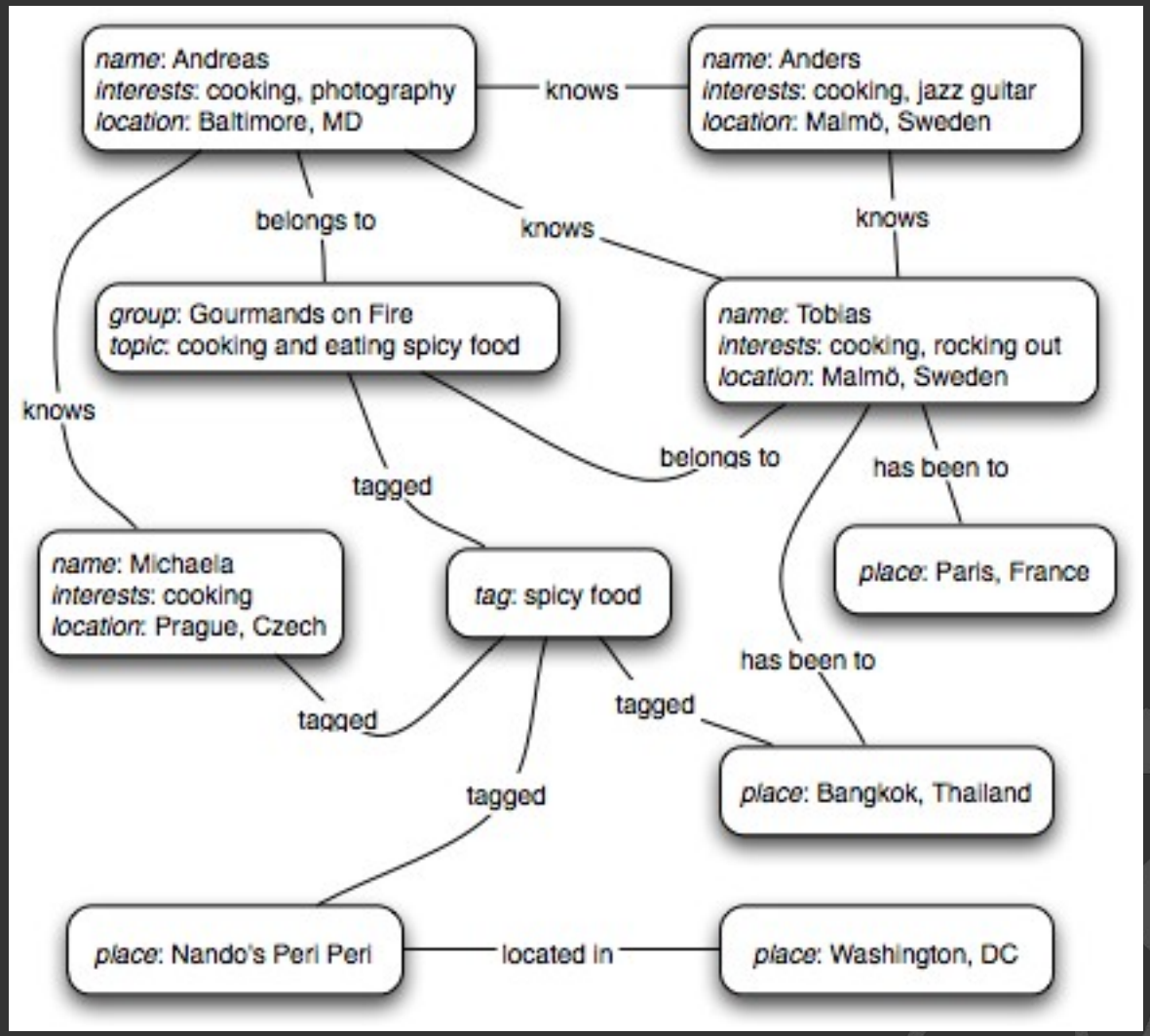
- All US HQ susidiaries with geo boundary validation (only SFO, LA)




Neo4j
Overview
Dashboard
Explore and edit
Data browser
Power tool
Console
Details
Server info
Documentation
9
Search by: [node id], rel:[relationship id], rels:[node id]
Node
Relationship
Visualization settings
2: root
security
political
geo
13: acl
10: ACME
6: World
3: Geo
principal
watches
14: Team Finance
belongs_to
part_of
located_in
7: US
4: US West
5: Europe
team
team
location
part_of
located_in
11: HQ 1
8: CA
16: Joe
15: Bob
belong
owns
part_of
12: BR 1
location
9: 12345

Social graphs

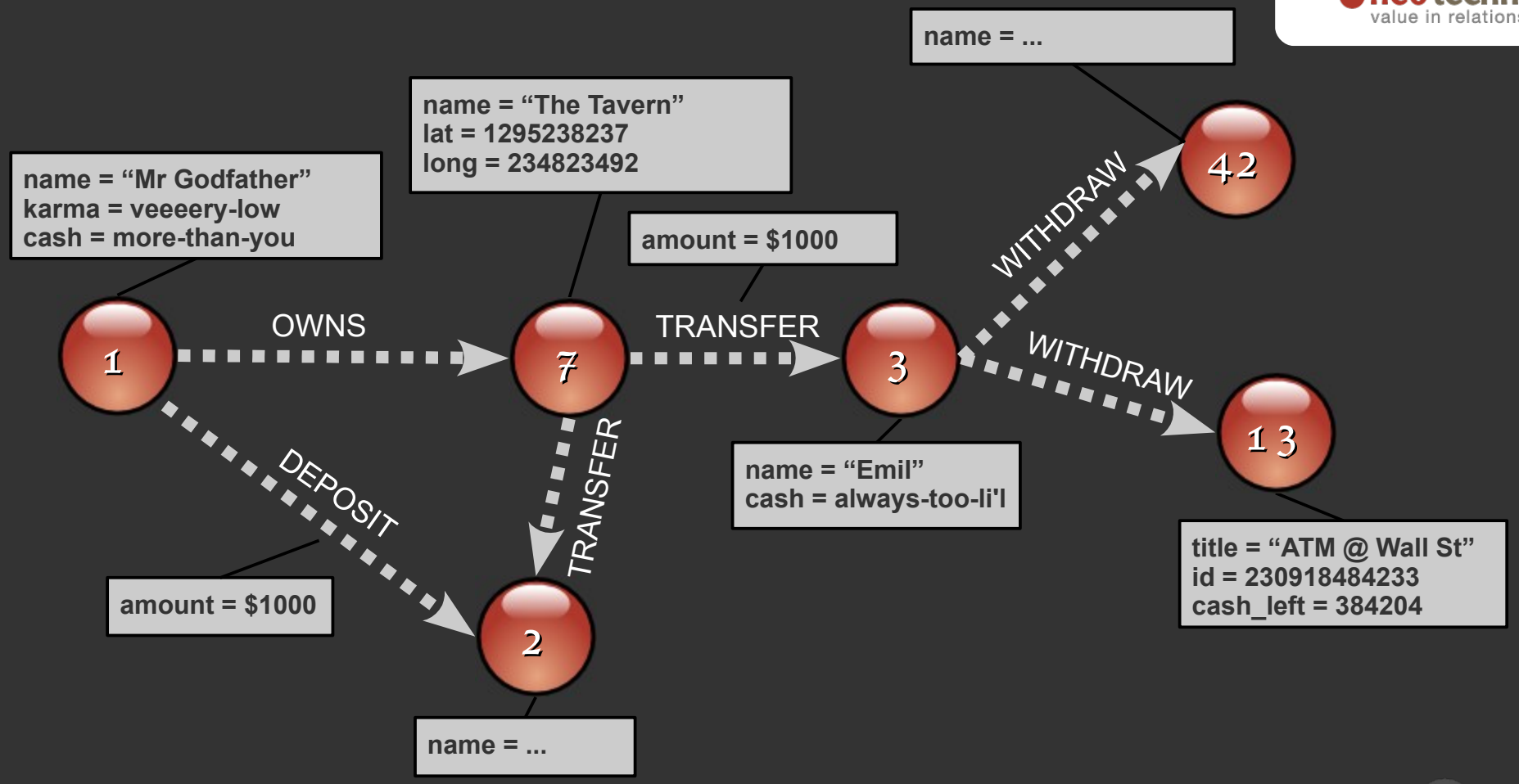
- Recommendations
- Location based services
- Influencers
- Shortest path



Recommendation engines

- Usage – the Web of Recommendation (Nova Spivak)
  - E-commerce
  - Ads
  - Friends, interests, places, partners etc.
- Techniques – data locality
  - Global heuristics
    - Page rank
  - Local recommendations
    - Shortest paths
    - Hammock functions
    - Random walks, energy diffusion
    - Dijkstra, A*, Shooting star etc

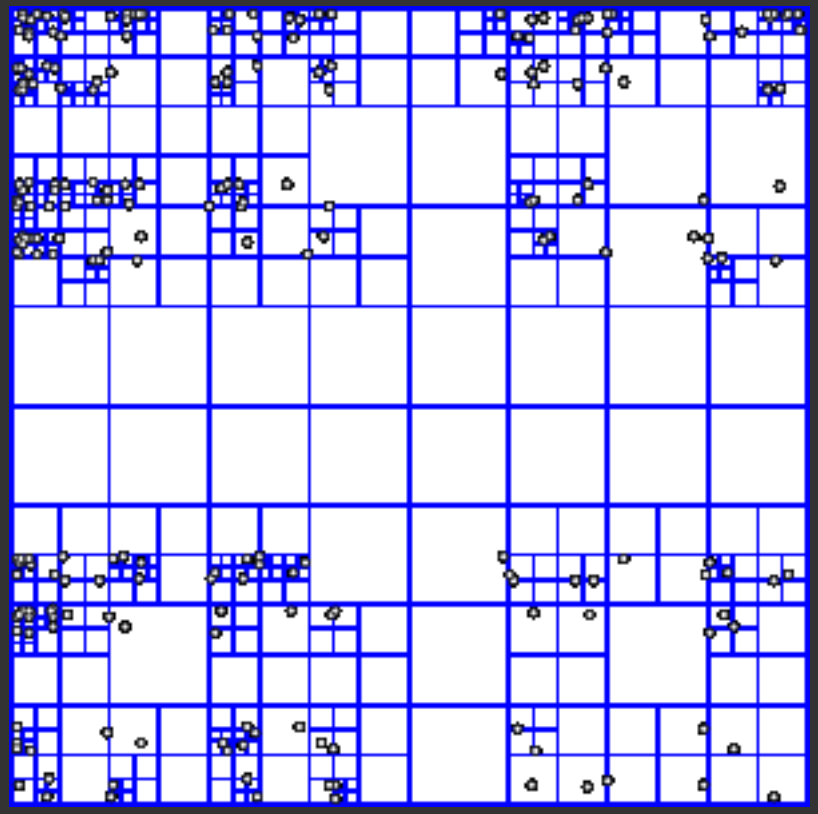
Financial data – fraud detection

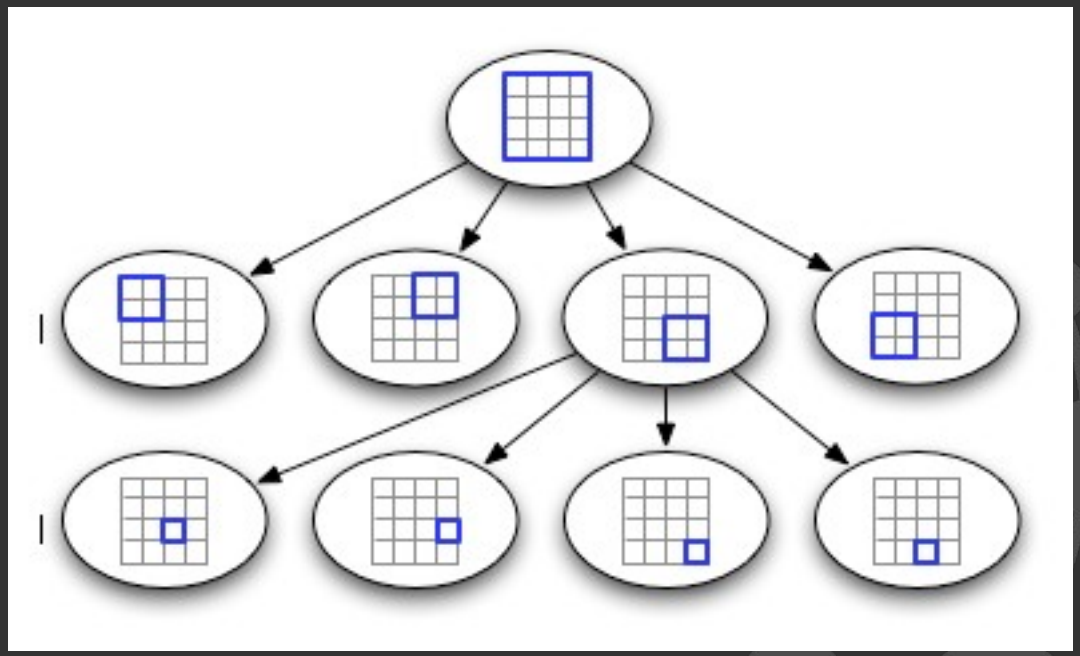


Call Data Records (CDR)

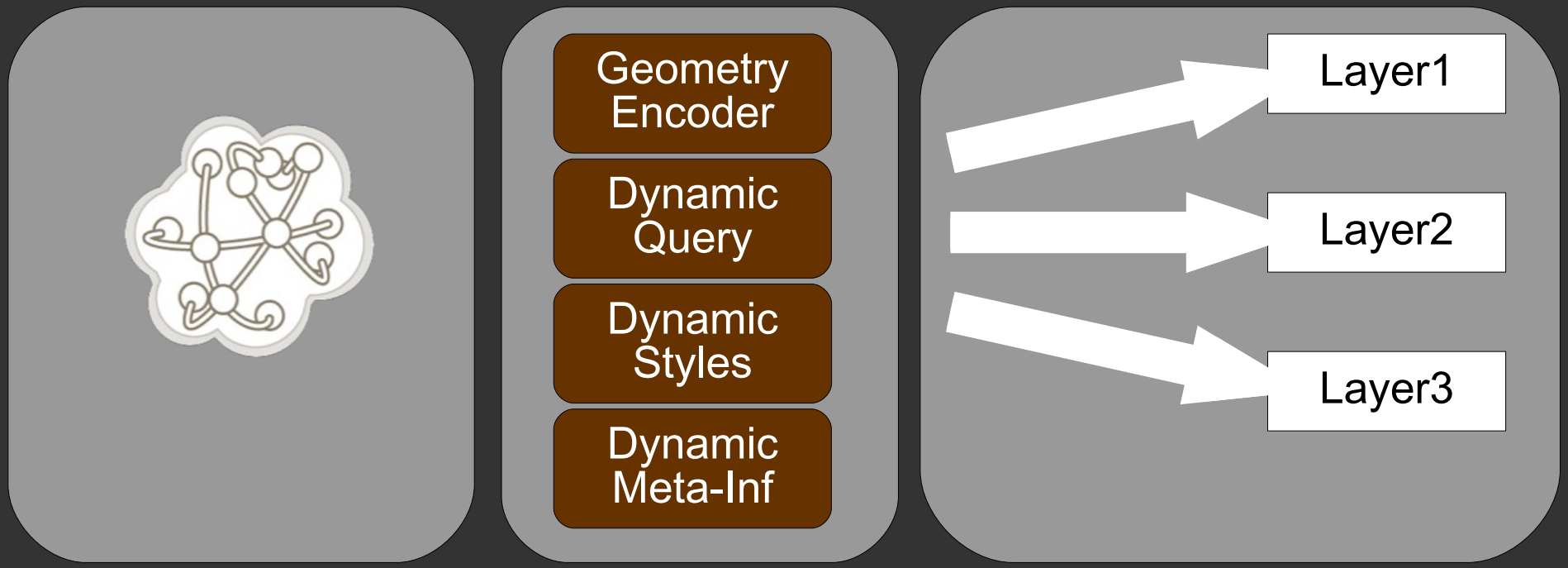
- Forming a (social) graph
- Location based
- Possible uses:
  - Find clusters (better plans)
  - Build social connections
  - Find influencers

Multiple indexes - GIS





Neo4j dynamic layers

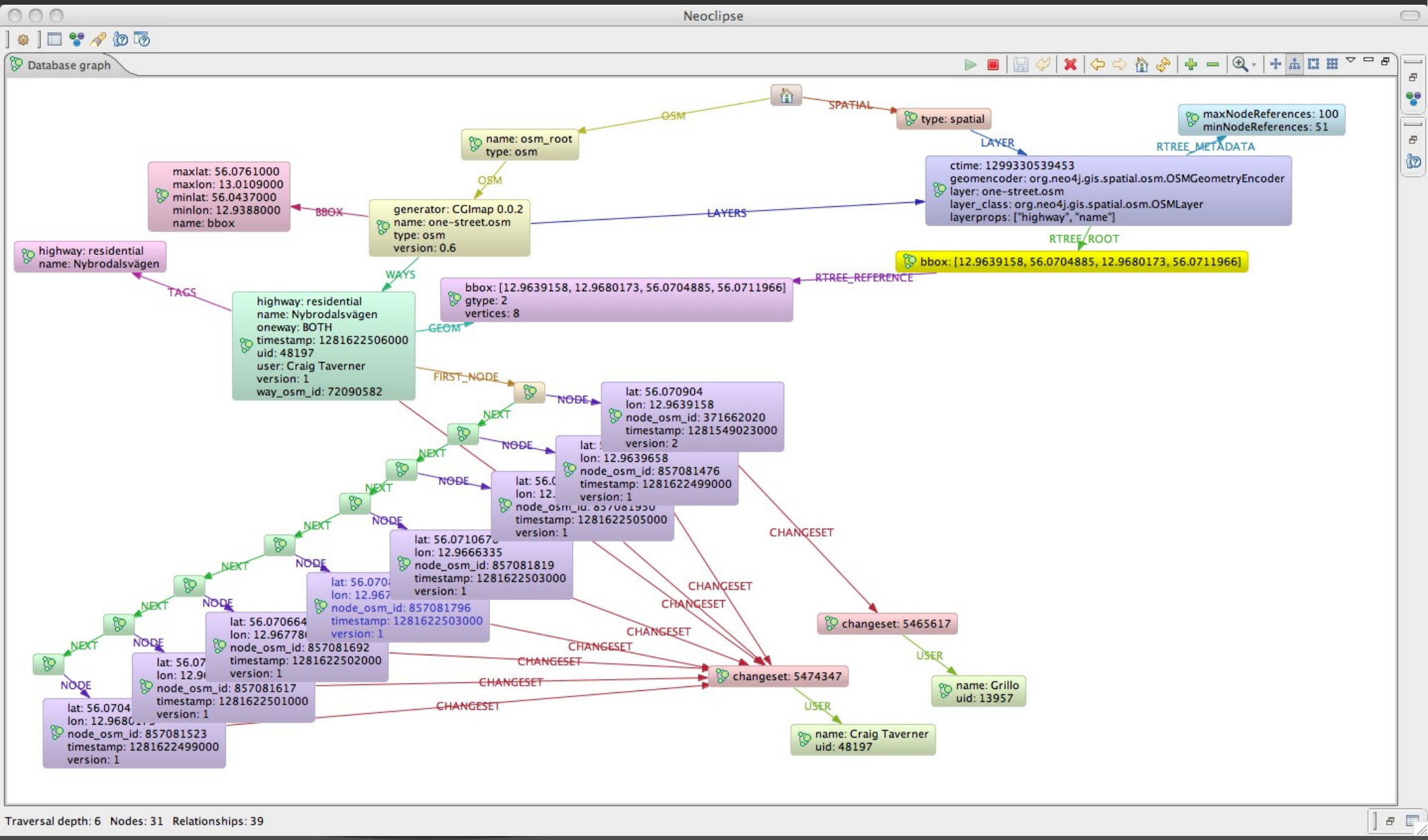


Connected domain data

Neo4j Spatial

GIS and Spatial stacks

OpenStreetMap



Network Topology analysis

- Analytics of network coverage and frequencies
  - Cell towers
  - Drive data
  - Infrastructure
- Analytics
  - Spatial signal strength
  - Antenna placement and azimuth
  - Frequency planning
  - Network differences over time
  - Reporting and charting

Cell network analysis

neotechnology
value in relationships

Amanzi Wireless Explorer

GeOptimA Report

Projects | Ruby Expl...

project
- Net work 2 Cell Database.c
  - Net work 2 Cell Databa
  - ROMES Drive Data 1.as

La... | Net... | Dat...

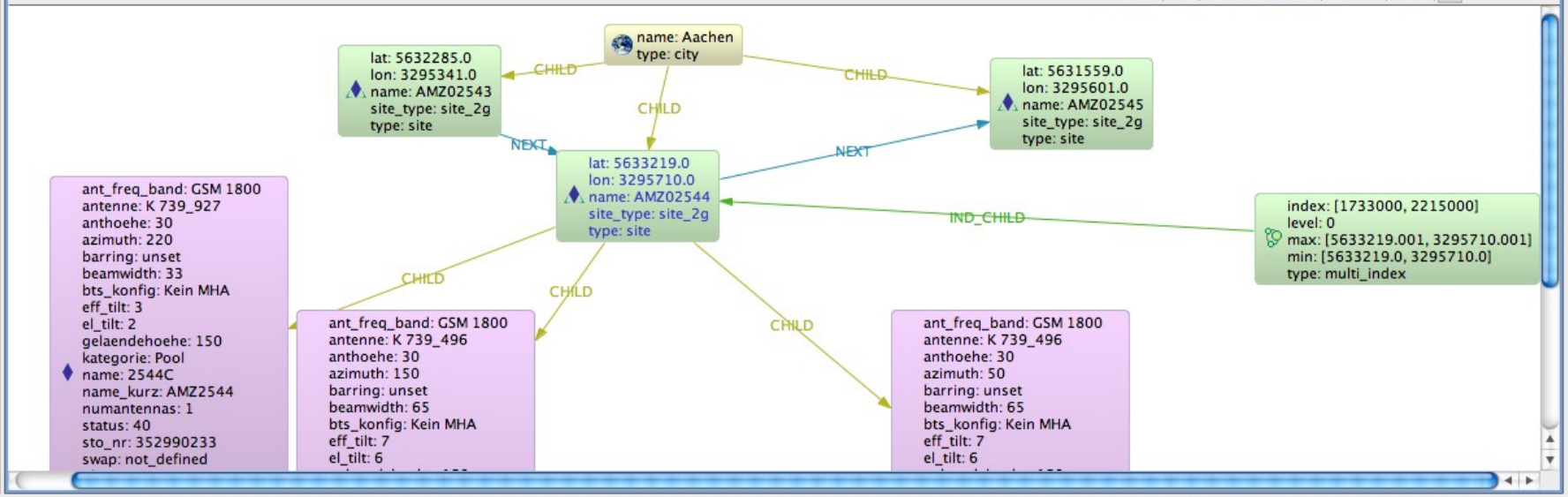
AMZ02533

/Users/peterneubauer/.aman
- Net work 2 Cell Database.
  - Aachen
    - AMZ02533
    - AMZ02536
    - AMZ02537
    - AMZ02538
    - AMZ02539
    - AMZ02540
    - AMZ02541
    - AMZ02542
    - AMZ02543
    - AMZ02544
      - 2544A
      - 2544B
      - 2544C
    - AMZ02545
    - AMZ02546
    - AMZ02547
    - AMZ02548
    - AMZ02549
    - AMZ02550
    - AMZ02551
    - AMZ02552
    - AMZ02553
    - AMZ02556

Net work 2 Cell Database.csv

1:20,837 | DHDN /...zone 3 | 3297424, 5634101

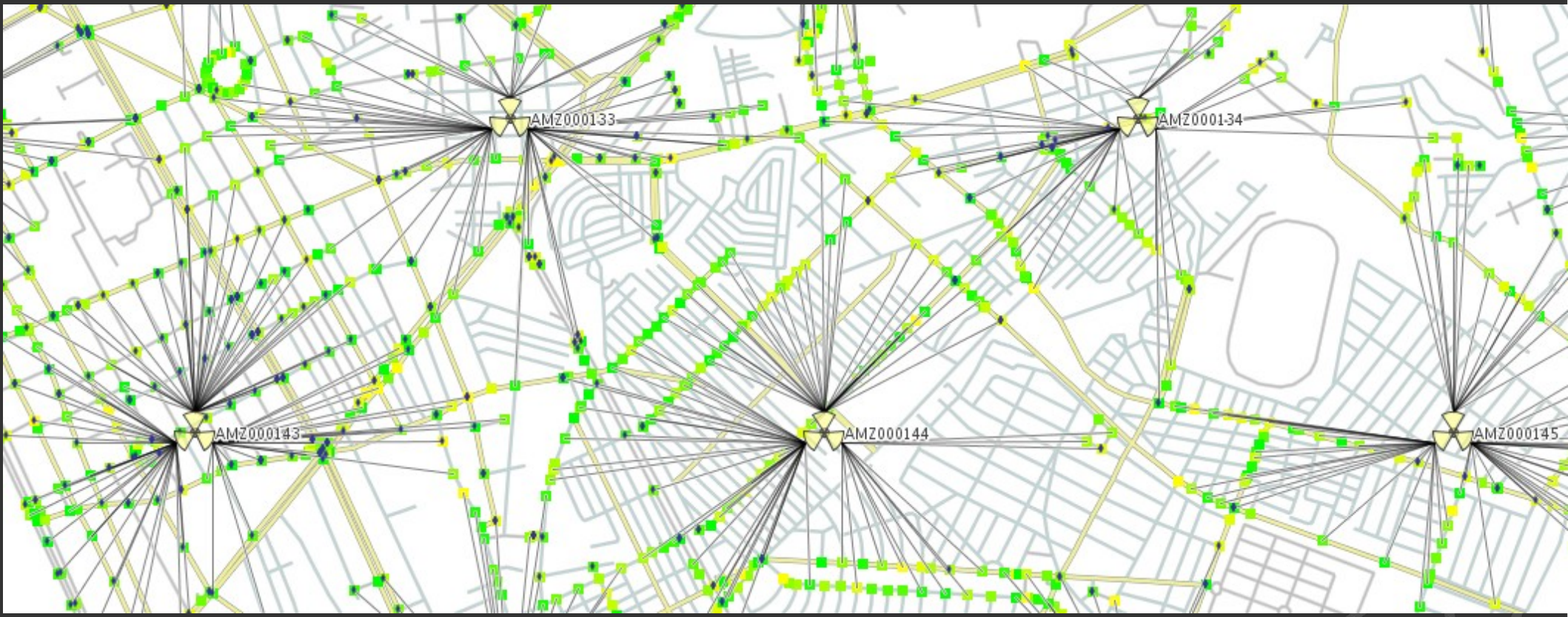
Catal... | Data... | Table | Search | Web | Book... | Cons... | Prop... | Filters | AWES... | Drive | Corr... | KPI B... | Neig... | Tran... | Distri... | Prop... | Info...



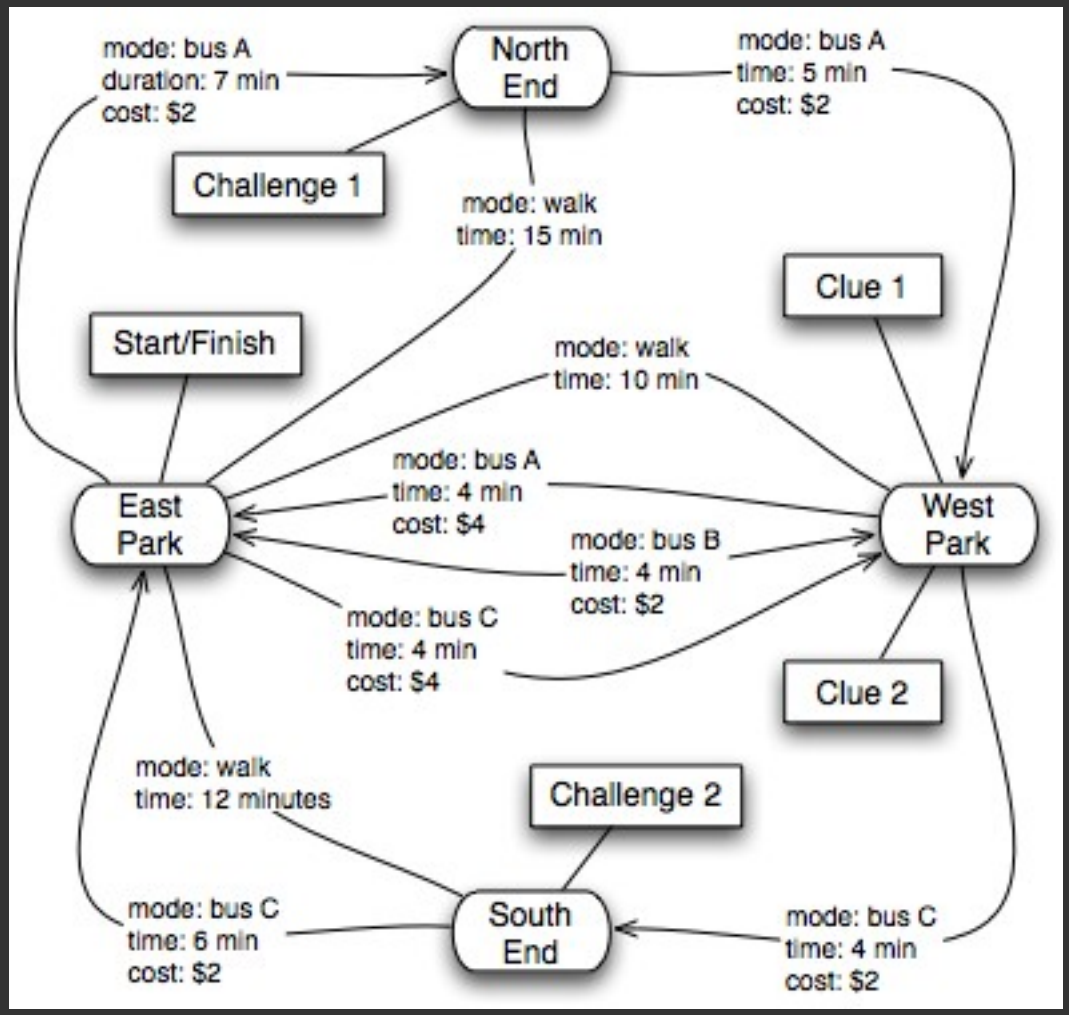
Traversal depth: 1 Nodes: 8 Relationships: 9

...: (100%)

Cell network analysis

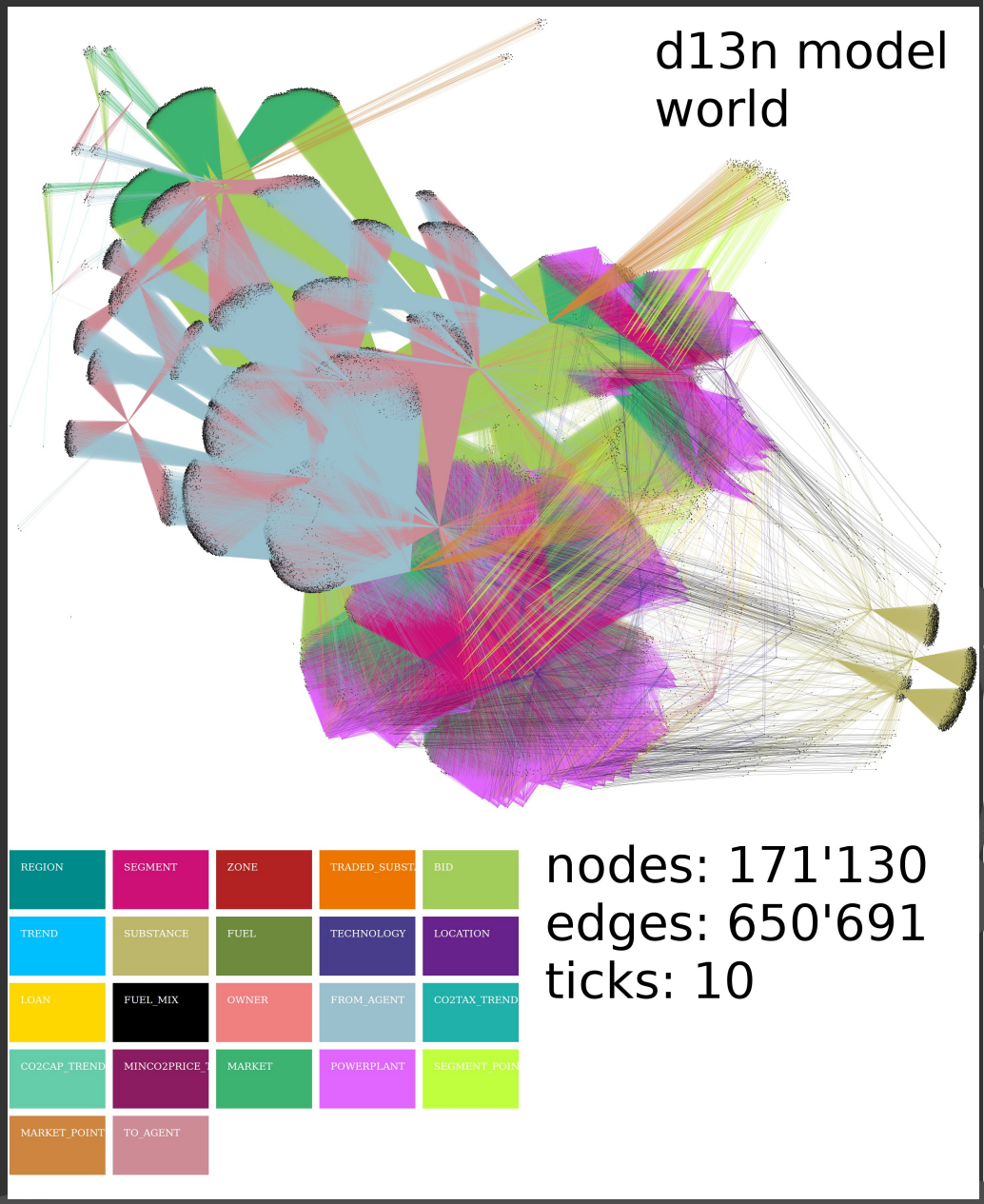


Routing, Logistics



Simulations (Energy market dynamics)

- agents, markets, power plants, bids, substances, technologies
- Analysis of volatilities for market changes
- Applicable for f nancial forecasting and regulations



Summary

- Value is in complexity of data
- Traditional models not good at handling the complexity of data
  - Index lookups and joins don't scale
- Graph databases are optimized for connections between data elements
- Transactional data and analytical data is the same
  - Real-time analytics on all data
  - Deeper and faster insight from complexity of data

Questions?

Image credit: lost again! Sorry :(

neo technology
value in relationships