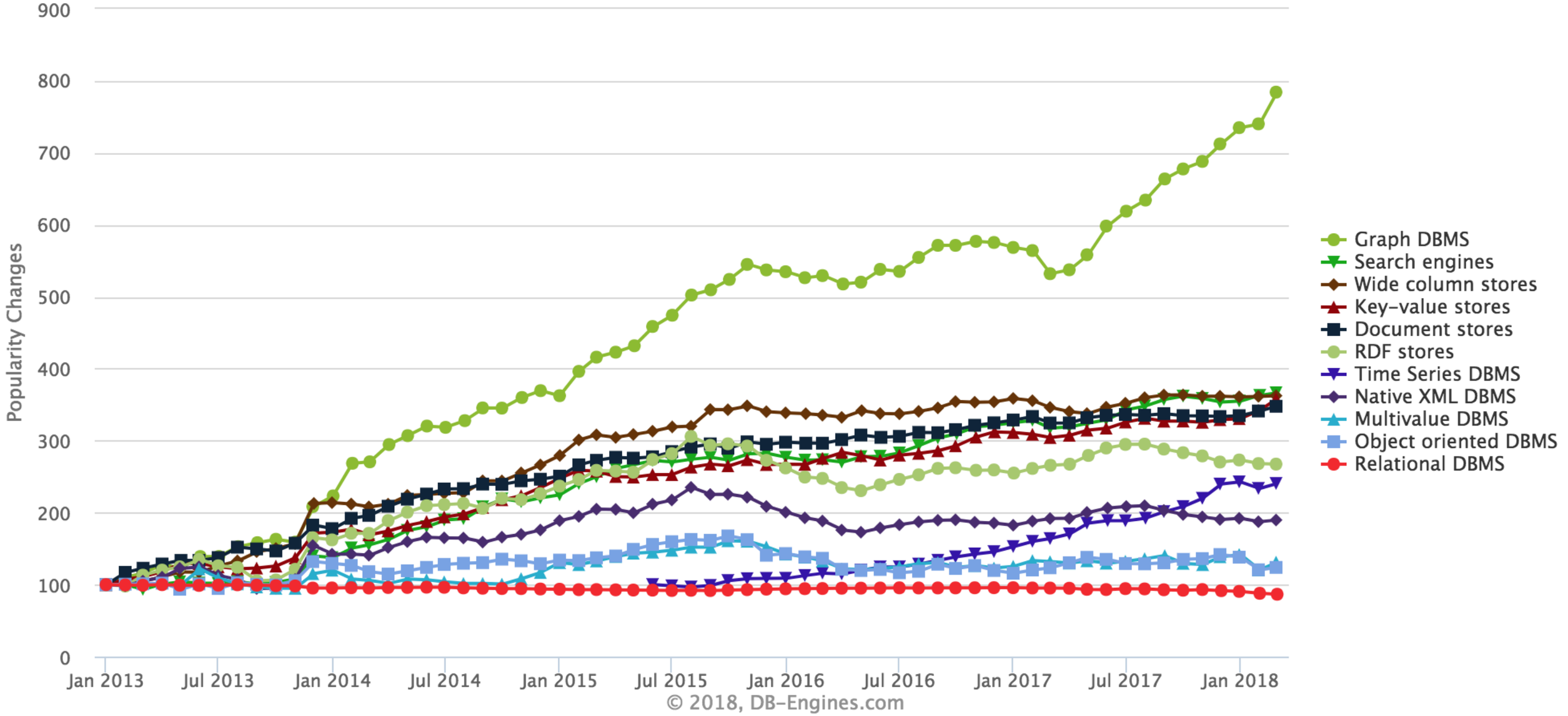
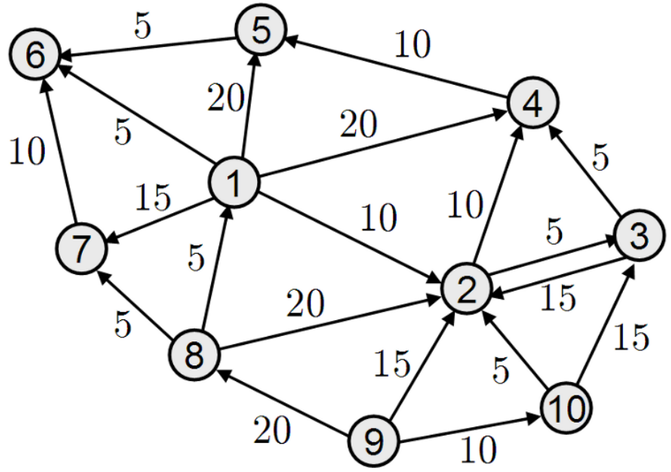
Introduction to Graphs & their Use Cases

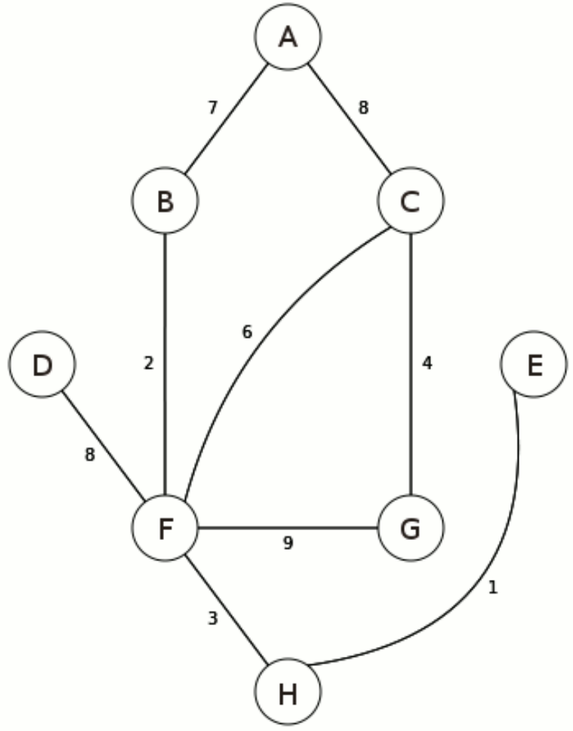
Brandon Richards
Sales Lead - Asia Pacific
Neo4j

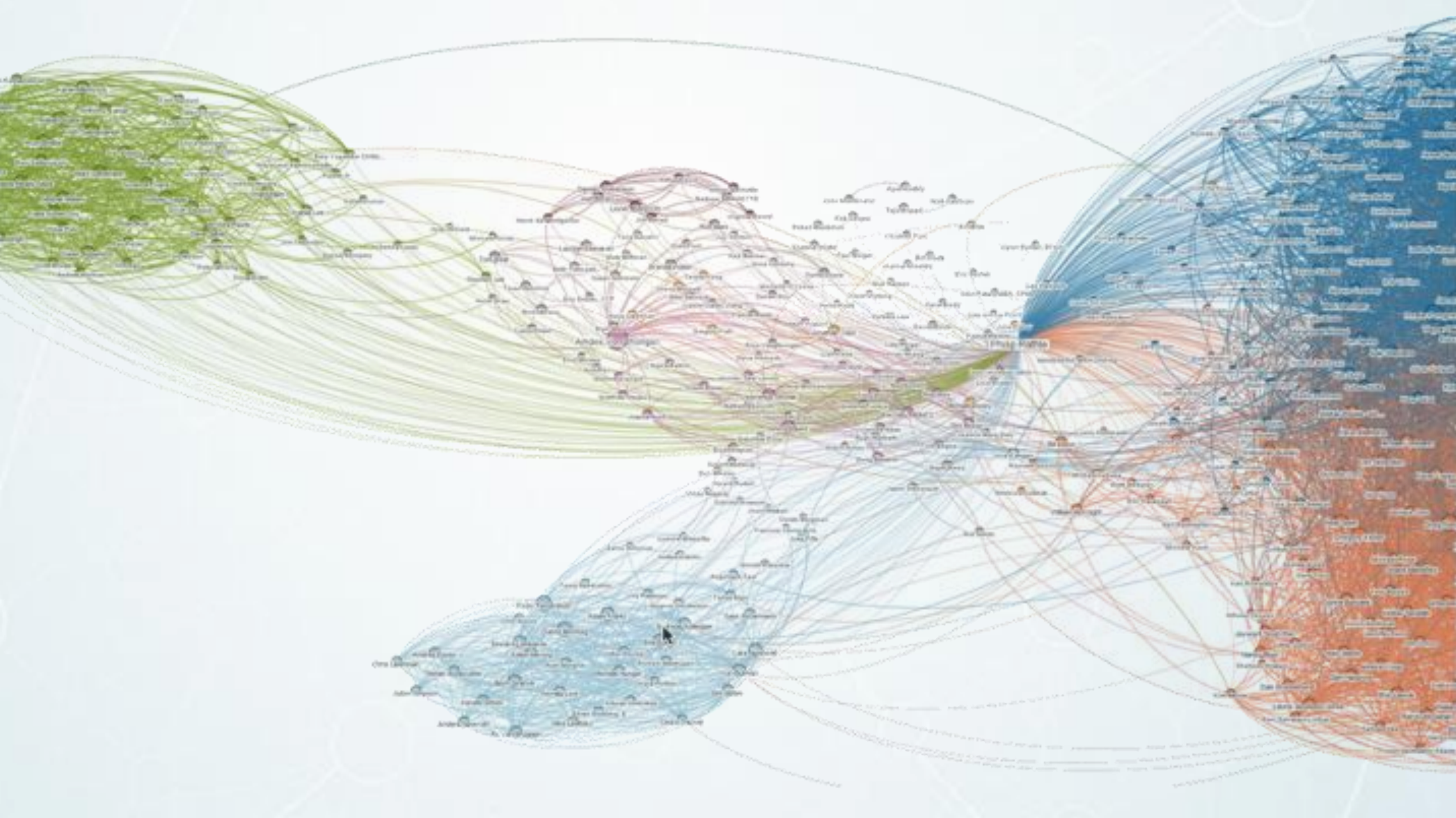
Fastest Growing Database Category

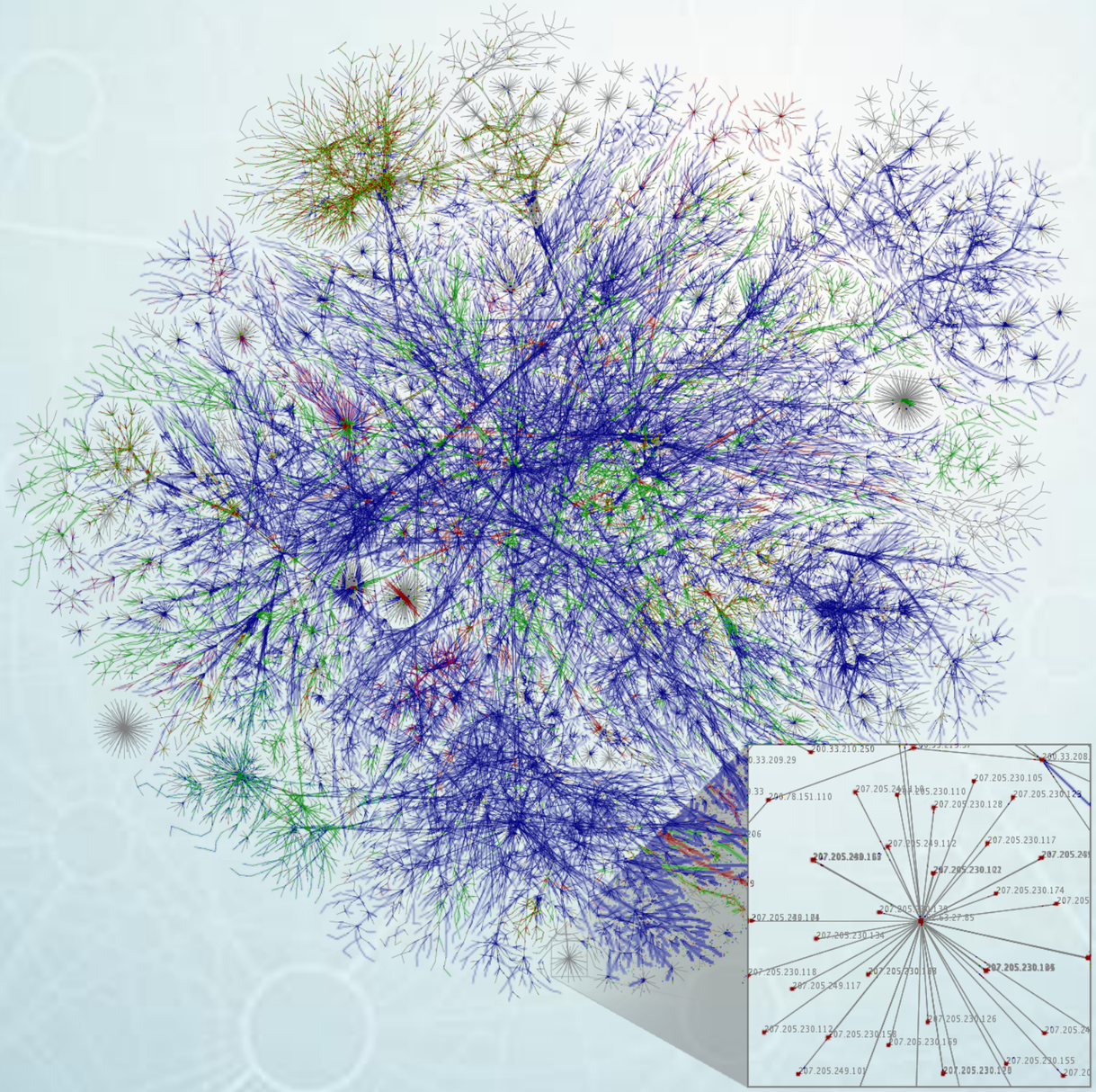


Number of Orders and Amount of Sales (last year)
Number of Orders
Amount of Sales in €
OrderID
Amount
mbination chart
Total fruit consumption
Apples
Oranges
Pears
Plums
Jane
John
Joe
Average
Highch
Sine Wave
Cosine Wave
Voltage (V)
Time (ms)


1
2
3
4
5
6
7
8
9
10


A
B
C
D
E
F
G
H

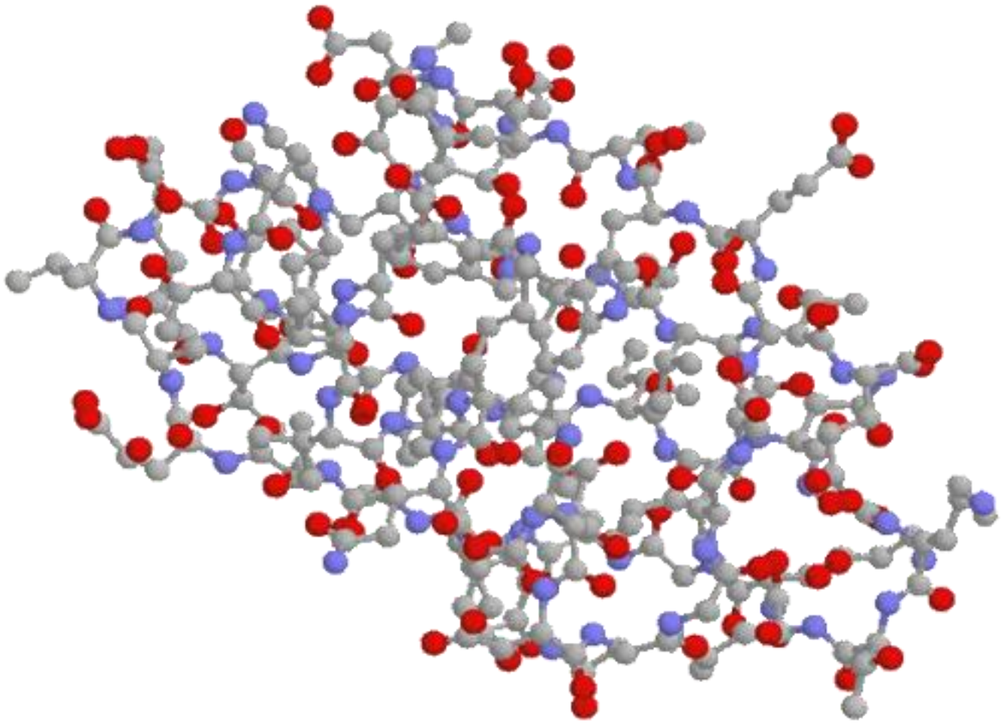


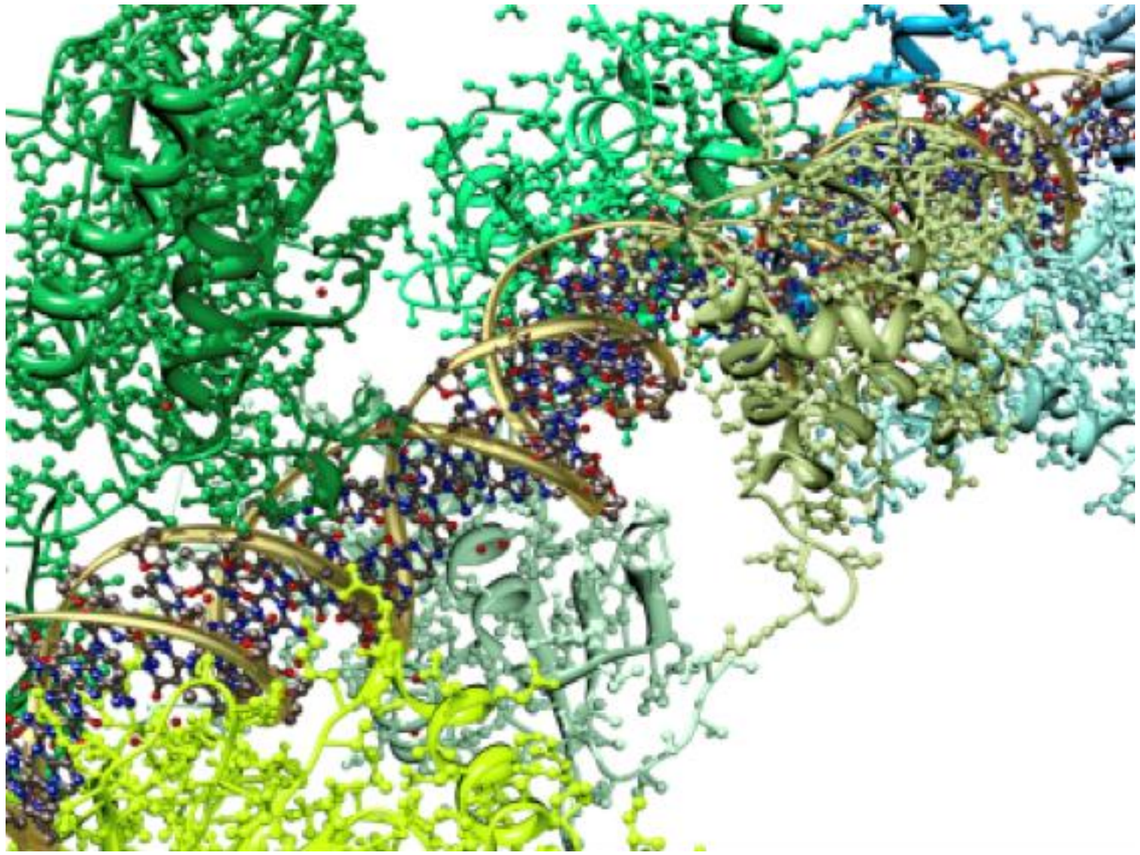


The internet

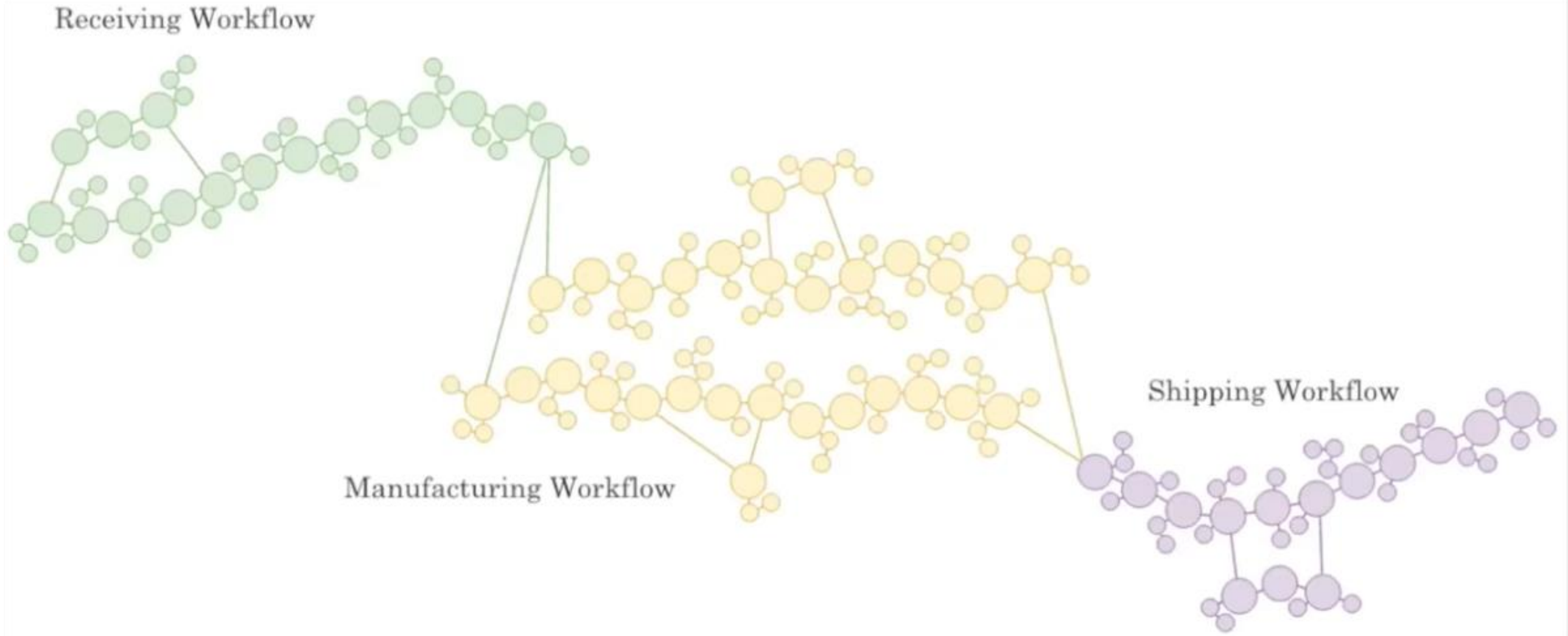
Genetic Ancestry of One Single Corn Variety

Protein Molecule

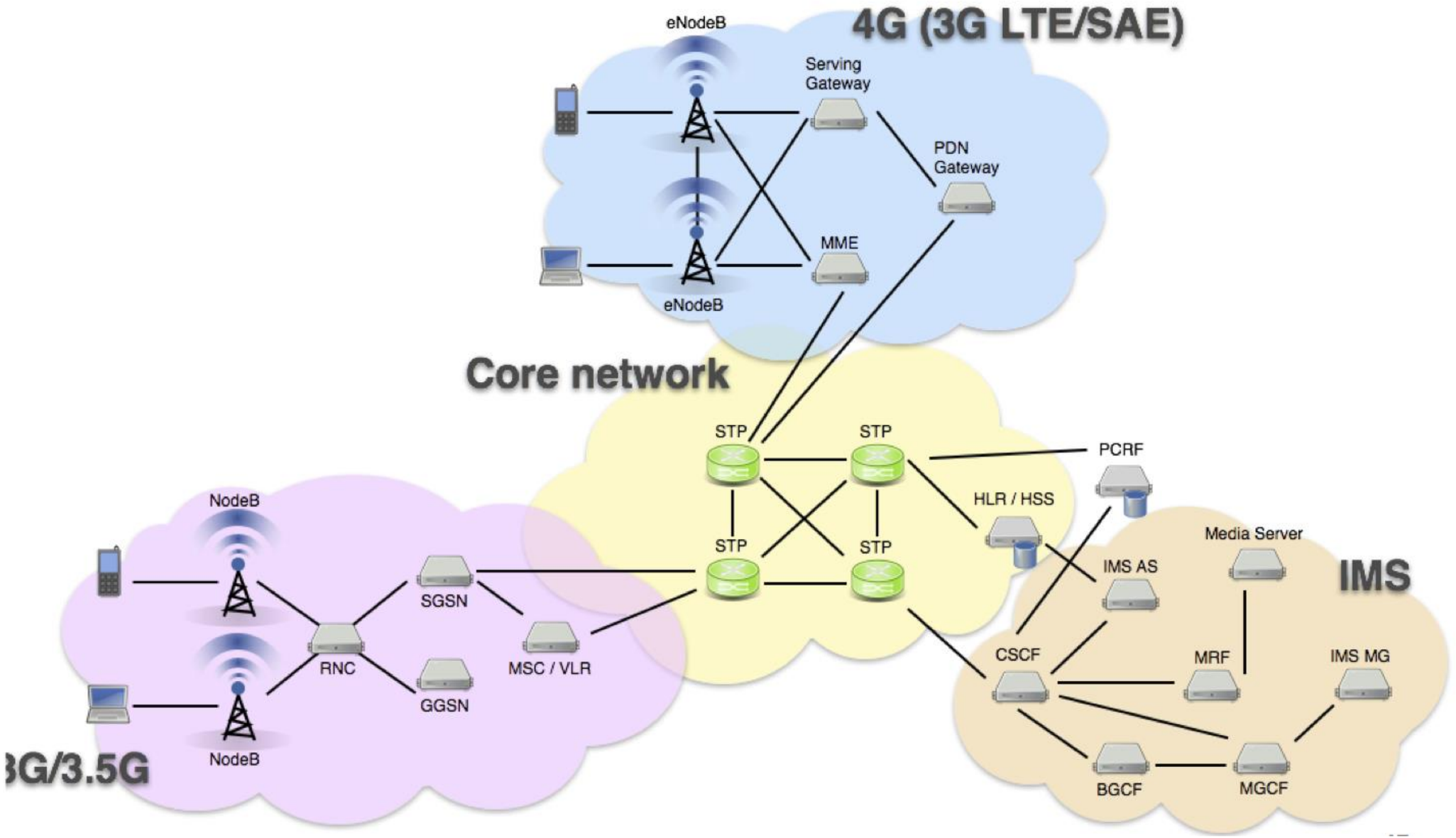


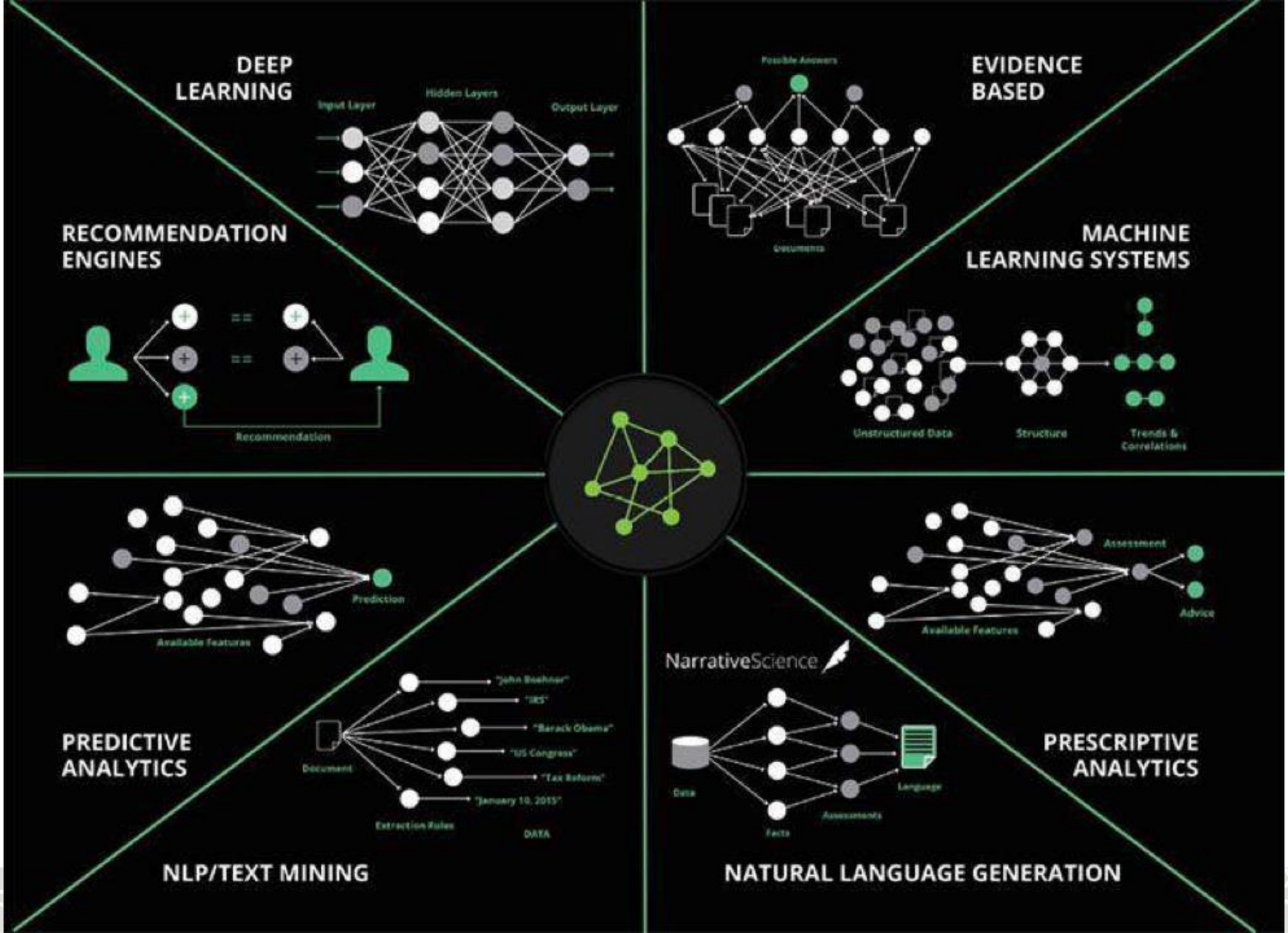


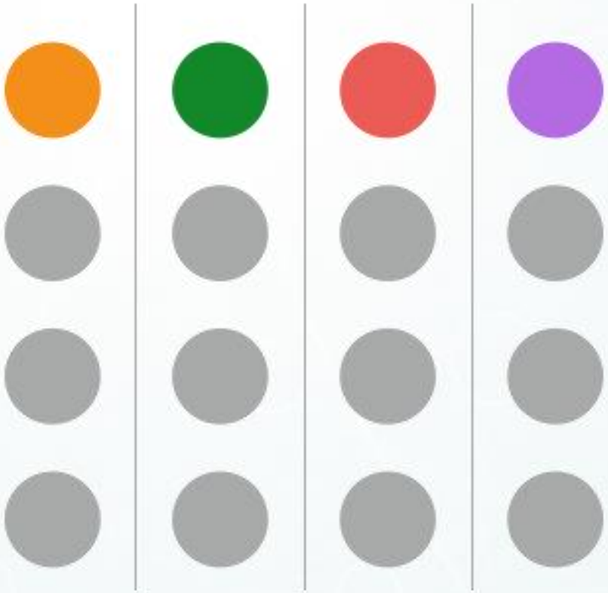
Process Workflow



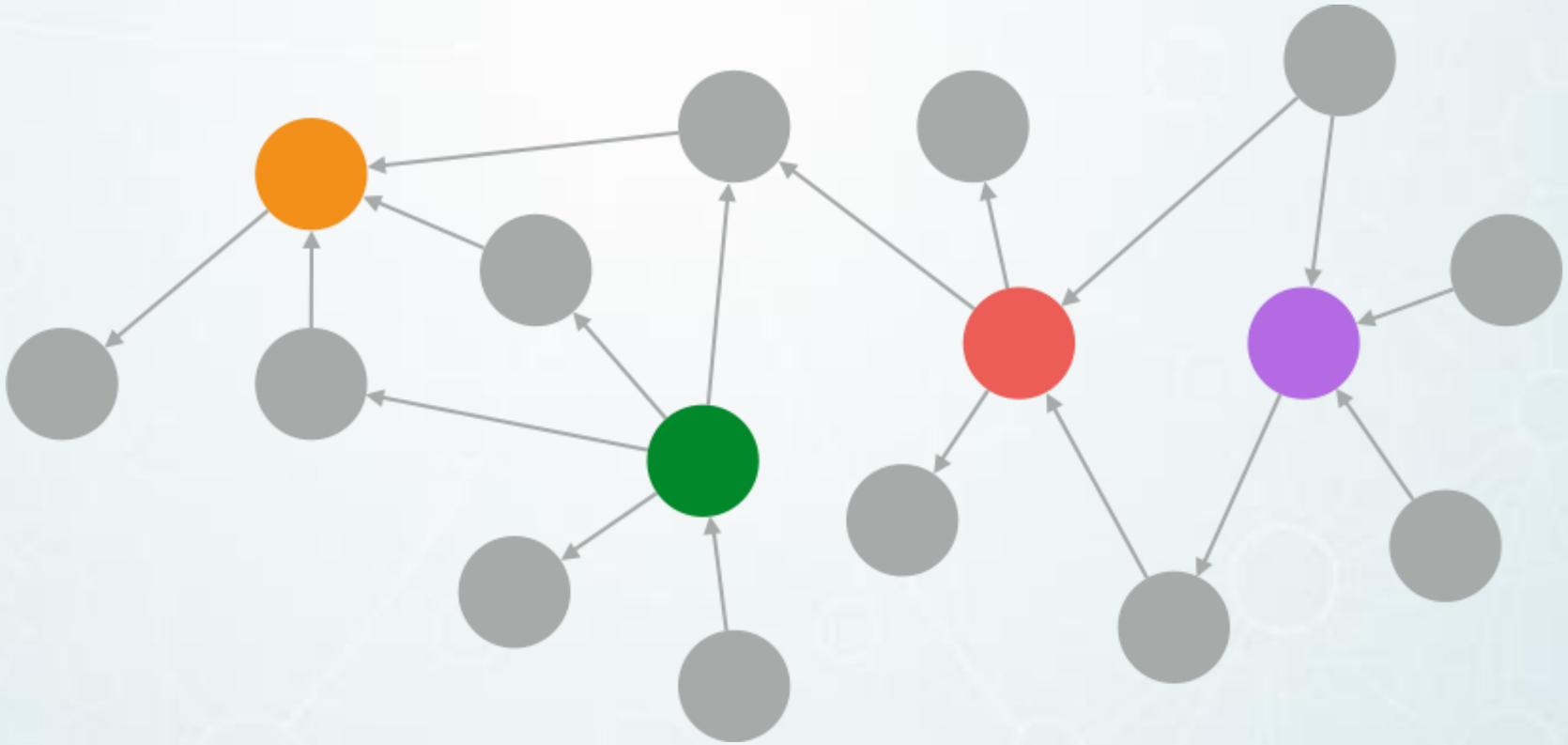
Telecom network architecture

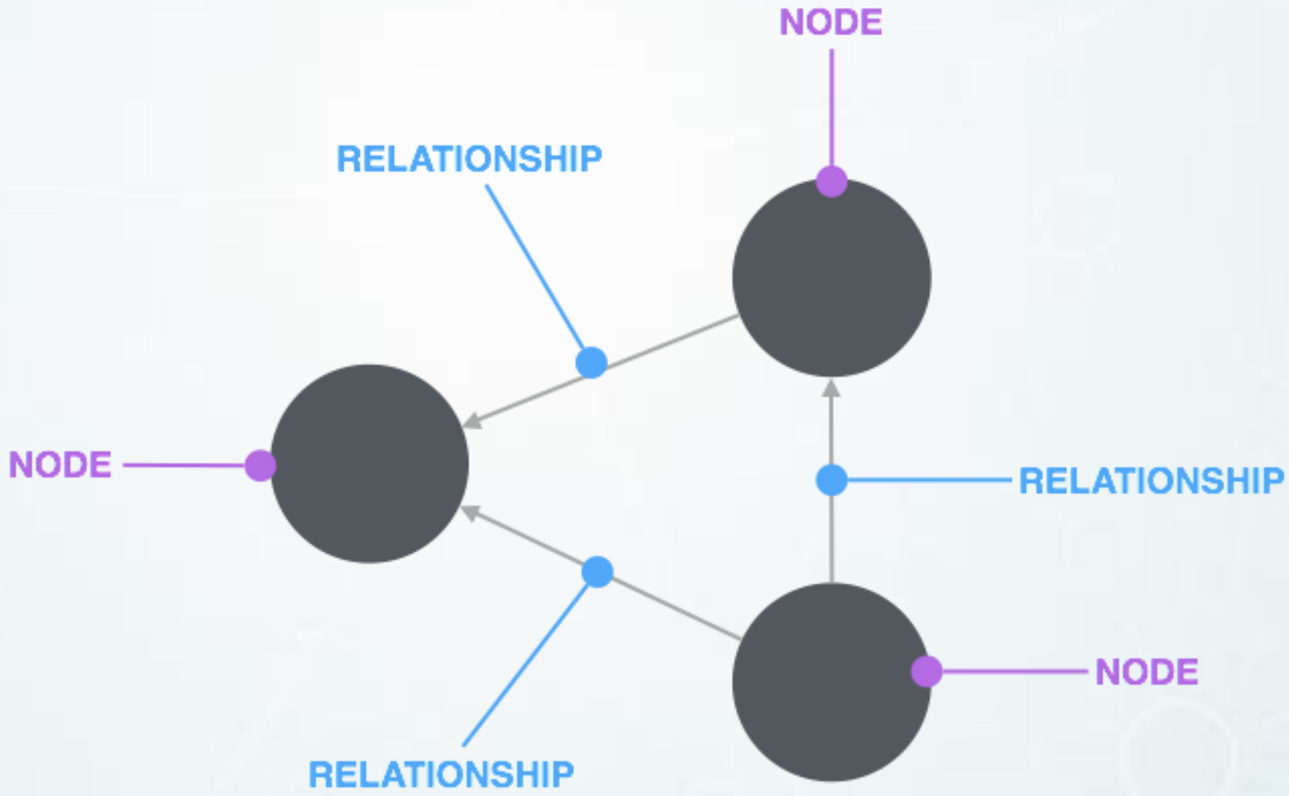



DEEP LEARNING
Input Layer
Hidden Layers
Output Layer
EVIDENCE BASED
Possible Answers
RECOMMENDATION ENGINES
Recommendation
MACHINE LEARNING SYSTEMS
Unstructured Data
Structure
Trends & Correlations
Prediction
Available Features
Assessment
Advice
Available Features
NarrativeScience
PREDICTIVE ANALYTICS
"John Boehner"
"IRS"
"Barack Obama"
"US Congress"
"Tax Reform"
"January 10, 2015"
Document
Extraction Rules
DATA
Data
Language
Assessments
Facts
PRESCRIPTIVE ANALYTICS
NLP/TEXT MINING
NATURAL LANGUAGE GENERATION

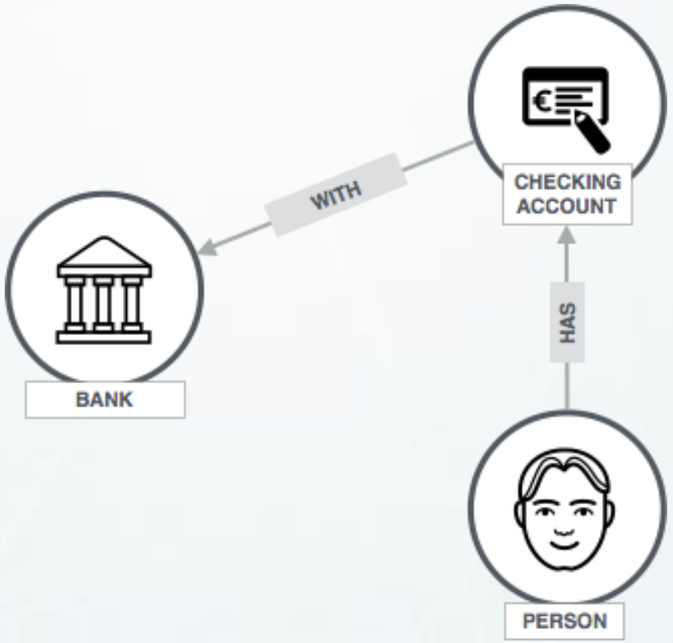


Data stored as a graph

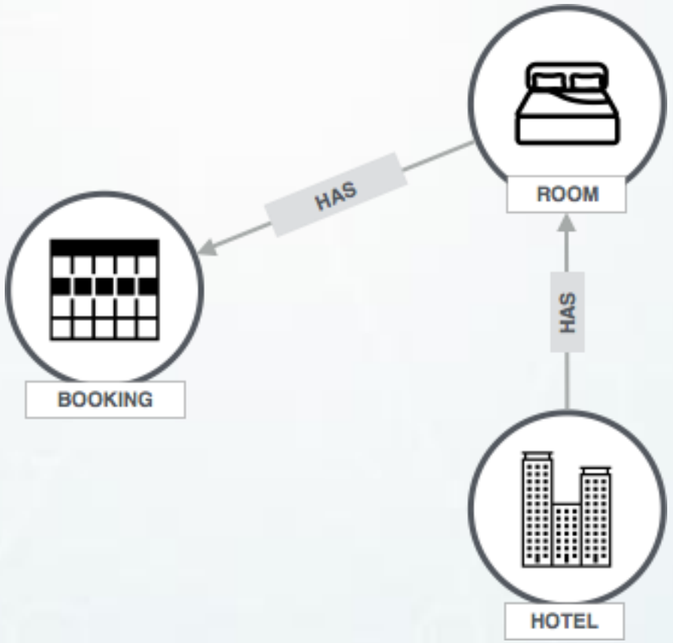




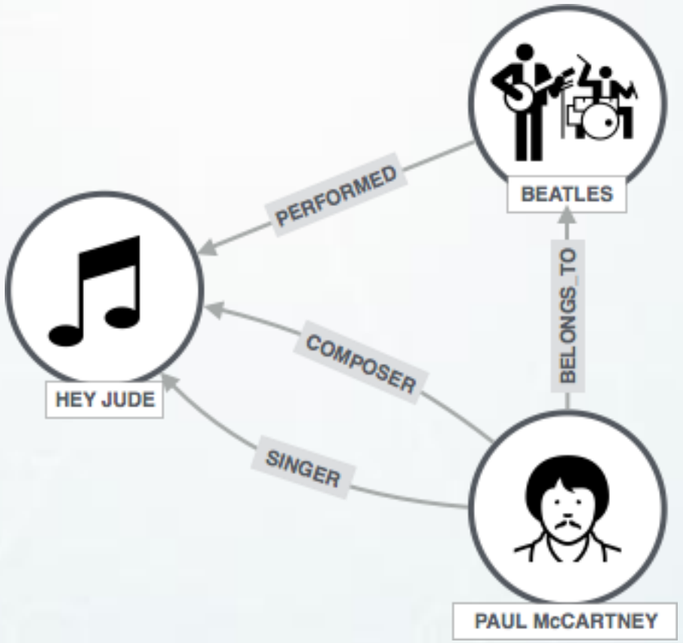
A Graph Is



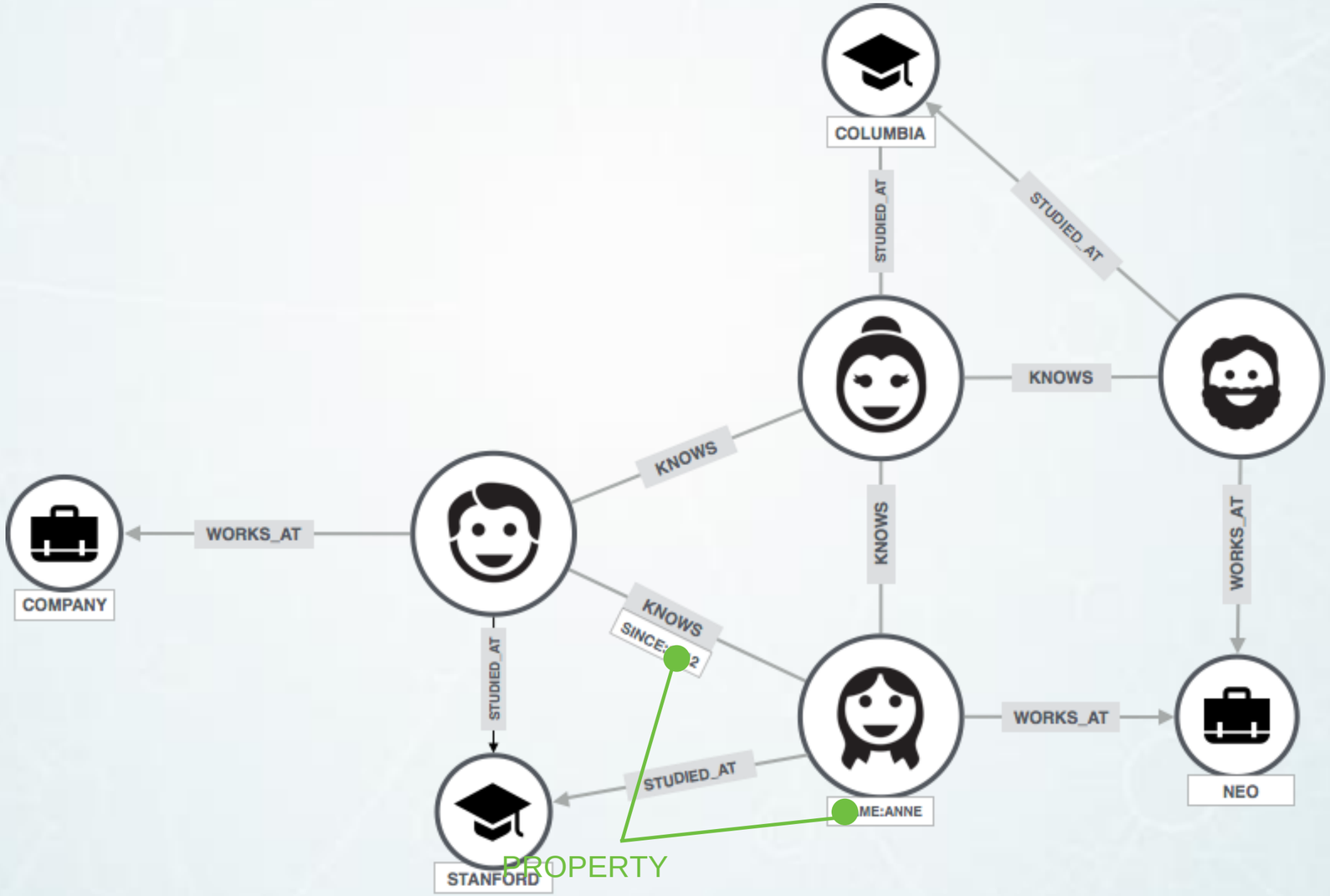
A Graph Is



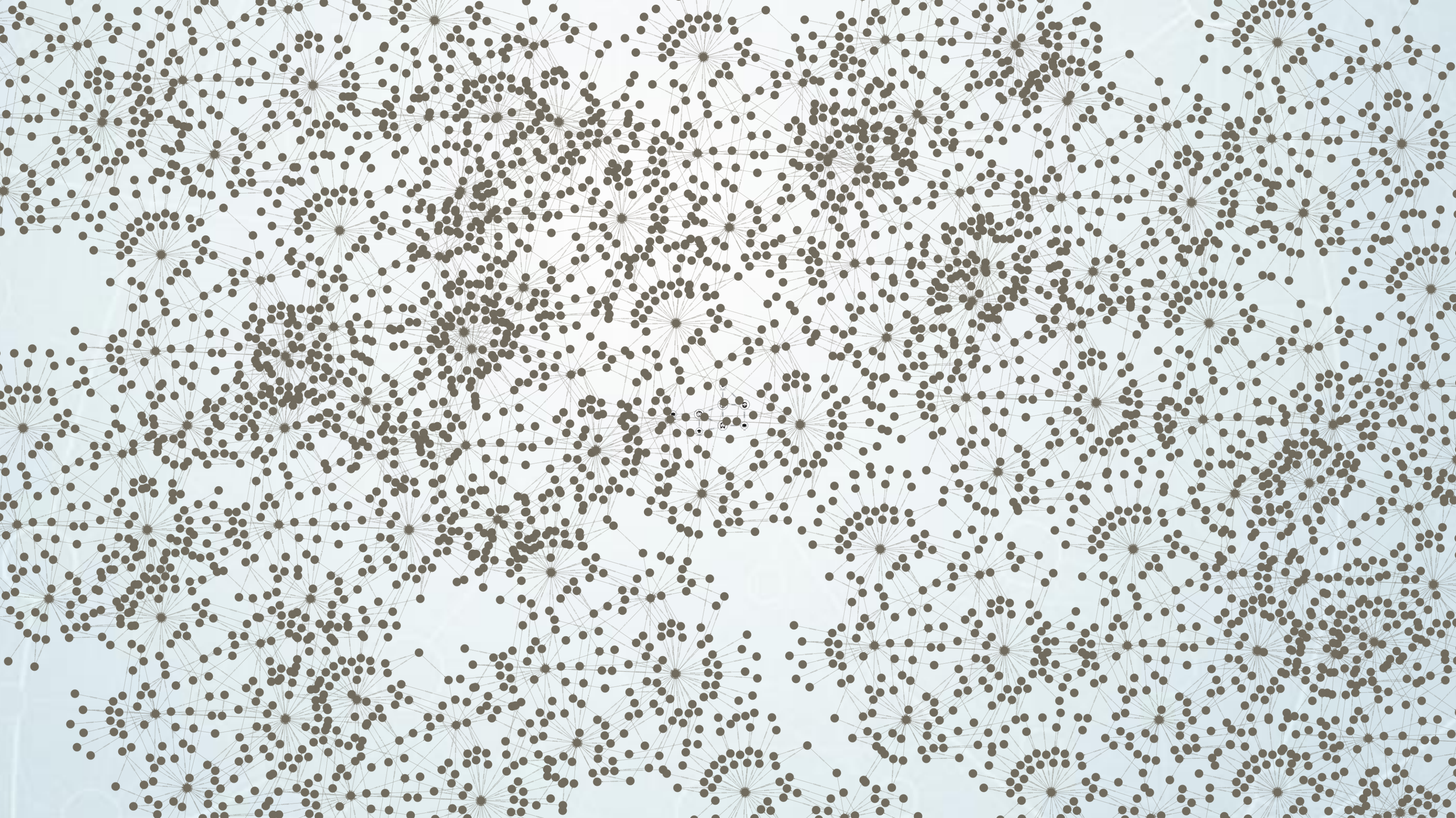
A Graph Is


BEATLES
PERFORMED
HEY JUDE
COMPOSER
BELONGS_TO
SINGER
PAUL McCARTNEY

A Graph Is


COLUMBIA
STUDIED_AT
STUDIED_AT
KNOWS
KNOWS
KNOWS
KNOWS
WORKS_AT
WORKS_AT
WORKS_AT
COMPANY
STUDIED_AT
STUDIED_AT
NEO
STANFORD
PROPERTY

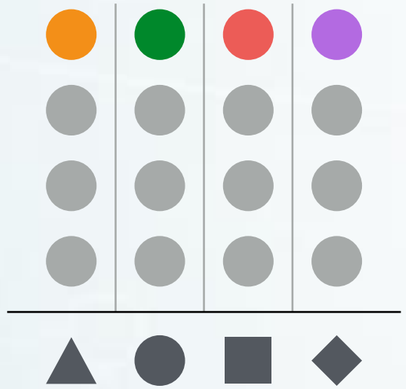
A Graph Is



"We found Neo4j to be literally **thousands of times faster** than our prior MySQL solution, with queries that require **10-100 times less code**. Today, Neo4j provides eBay with **functionality that was previously impossible**."

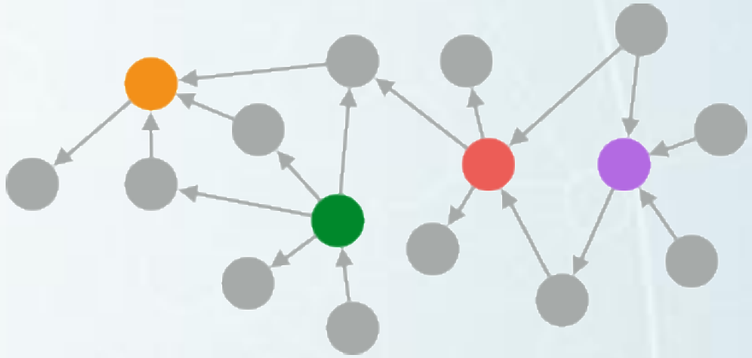
- Volker Pacher, Senior Developer

A way of representing data



Relational Database

Graph Database



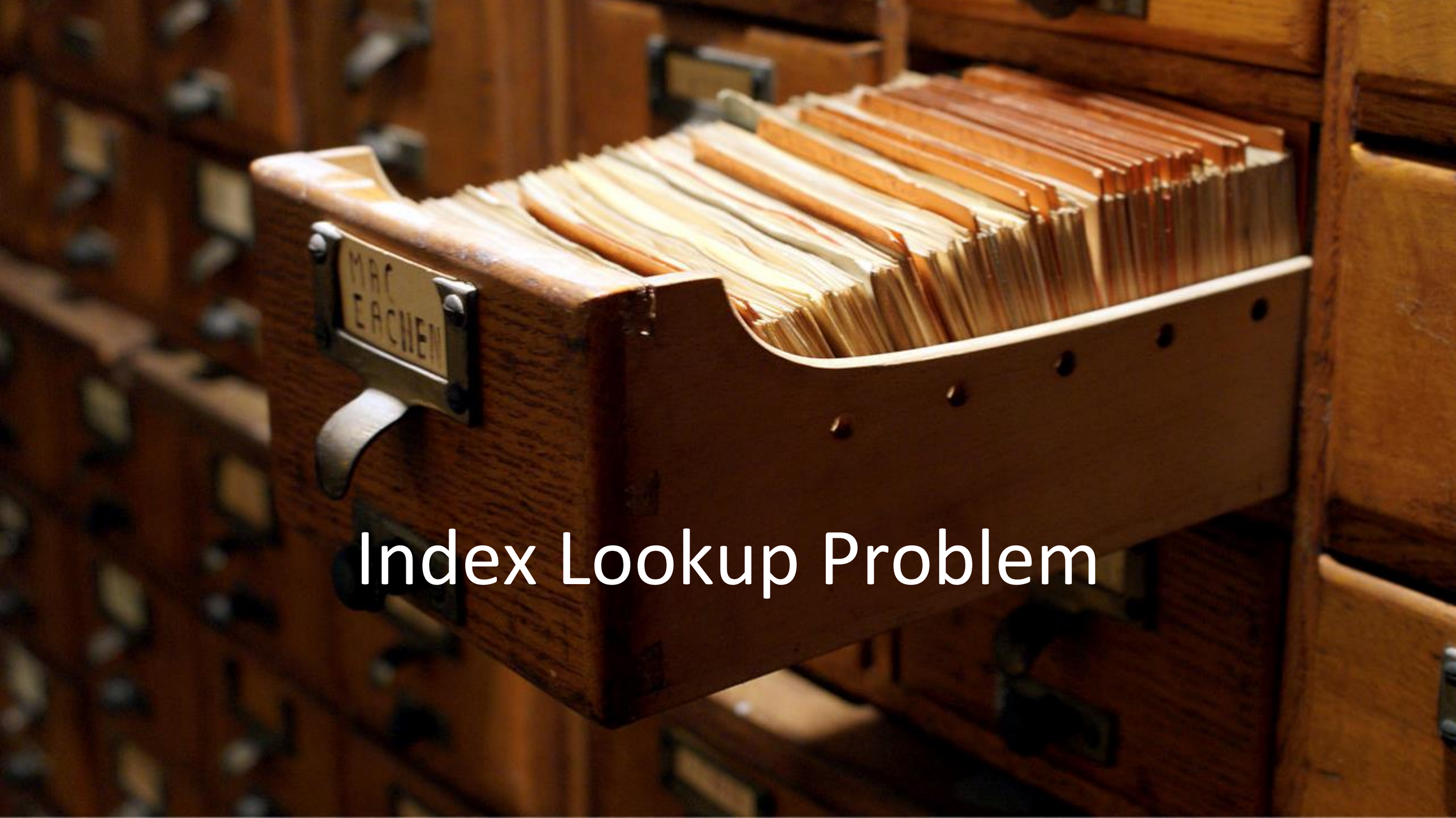
Good for:

- Well-understood data structures that don't change too frequently
- Known problems involving discrete parts of the data, or minimal connectivity

Good for:

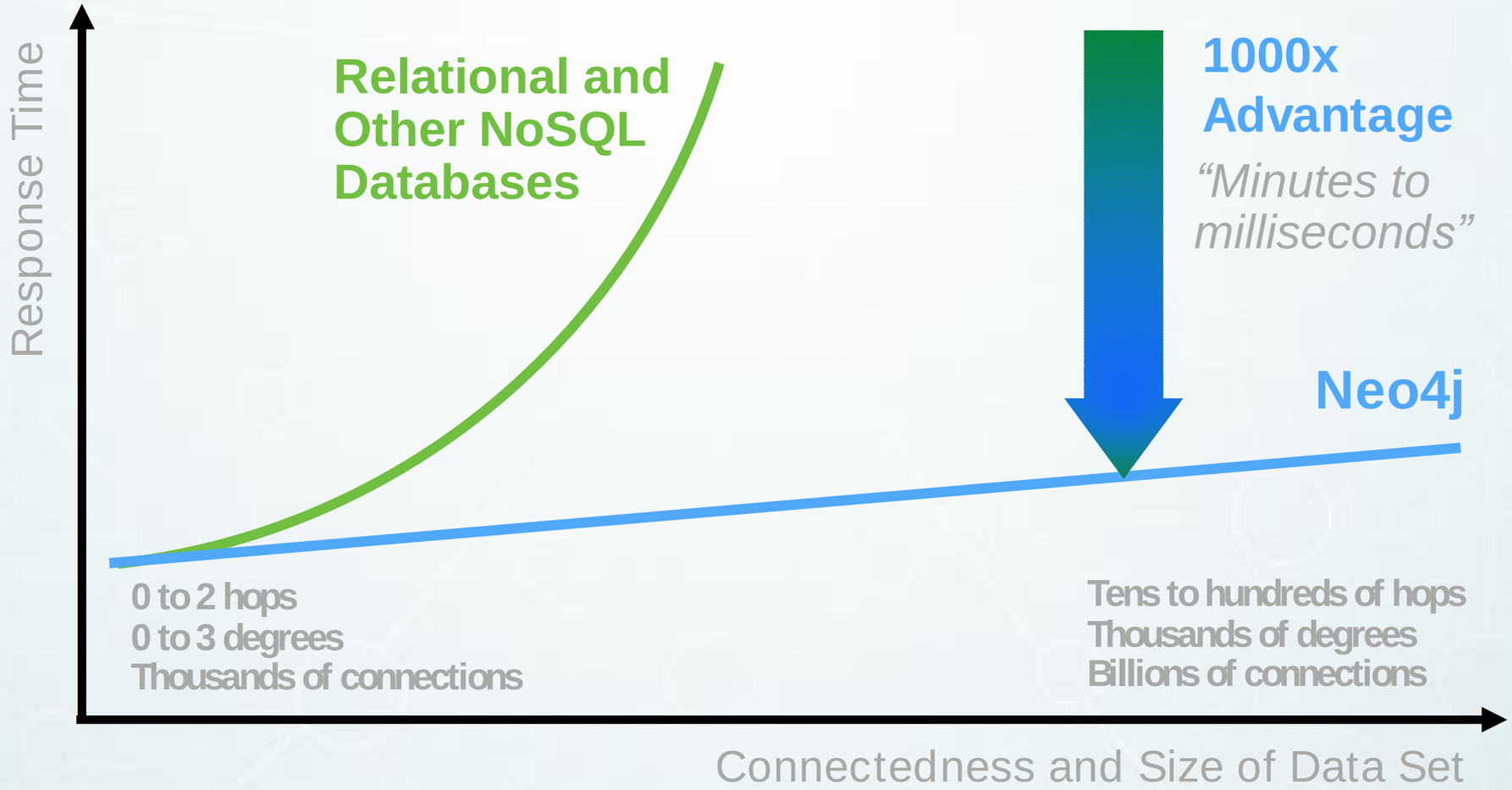
- Dynamic systems: where the data topology is difficult to predict
- Dynamic requirements: the evolve with the business
- Problems where the relationships in data contribute meaning & value

IT IS NOT THE STRONGEST OF THE SPECIES THAT SURVIVE NOR THE MOST INTELLIGENT BUT THE ONE MOST RESPONSIVE TO CHANGE
Charles Darwin (1809 - 1882)



Index Lookup Problem

Real-Time Query Performance

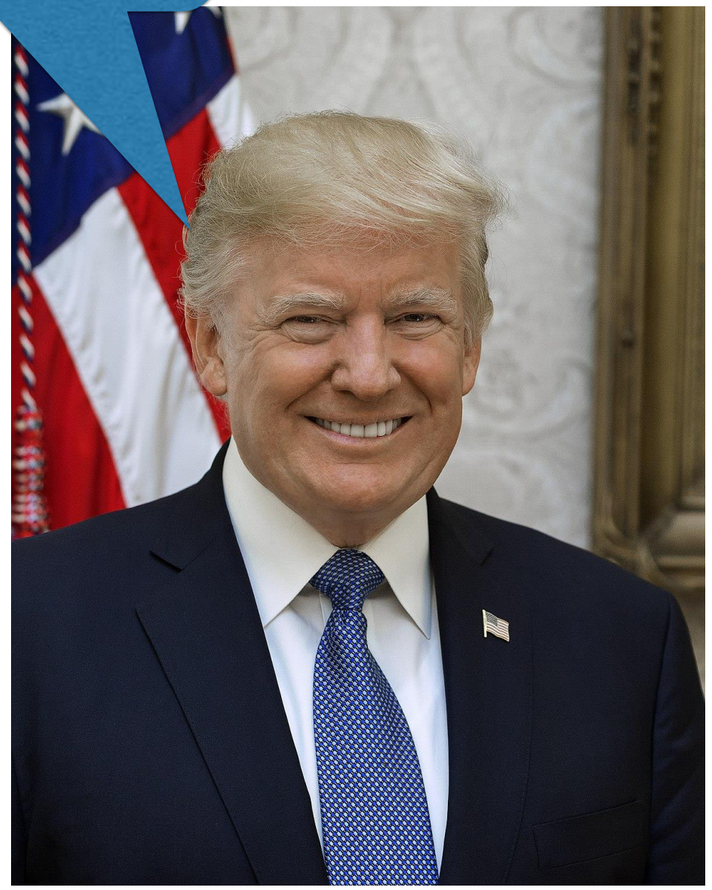


Native Graph

Non-Native Graph

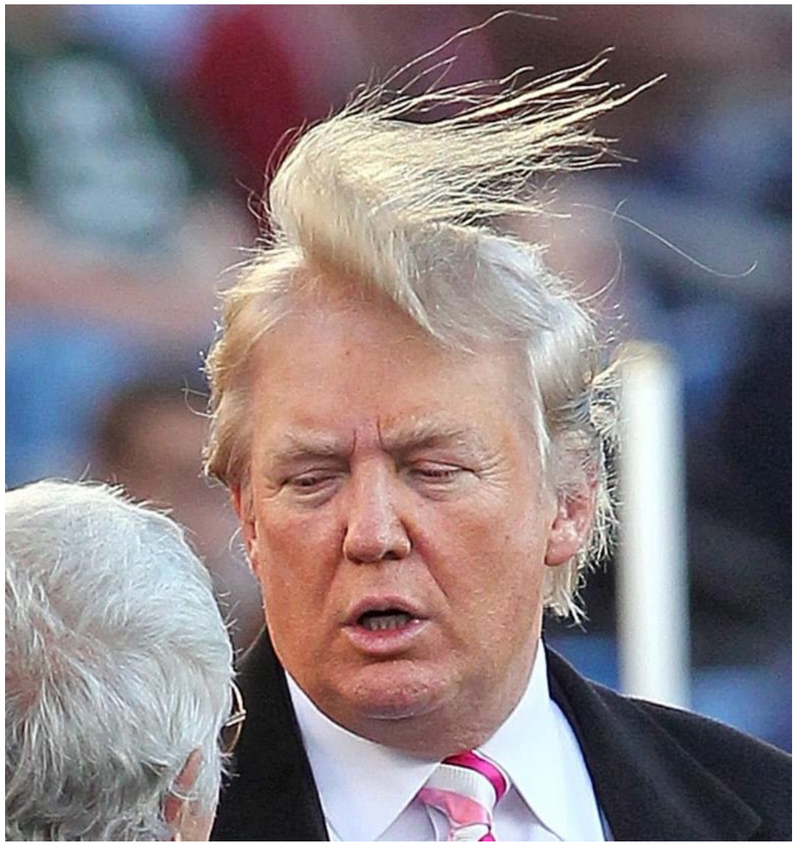
I’m a
Native Graph

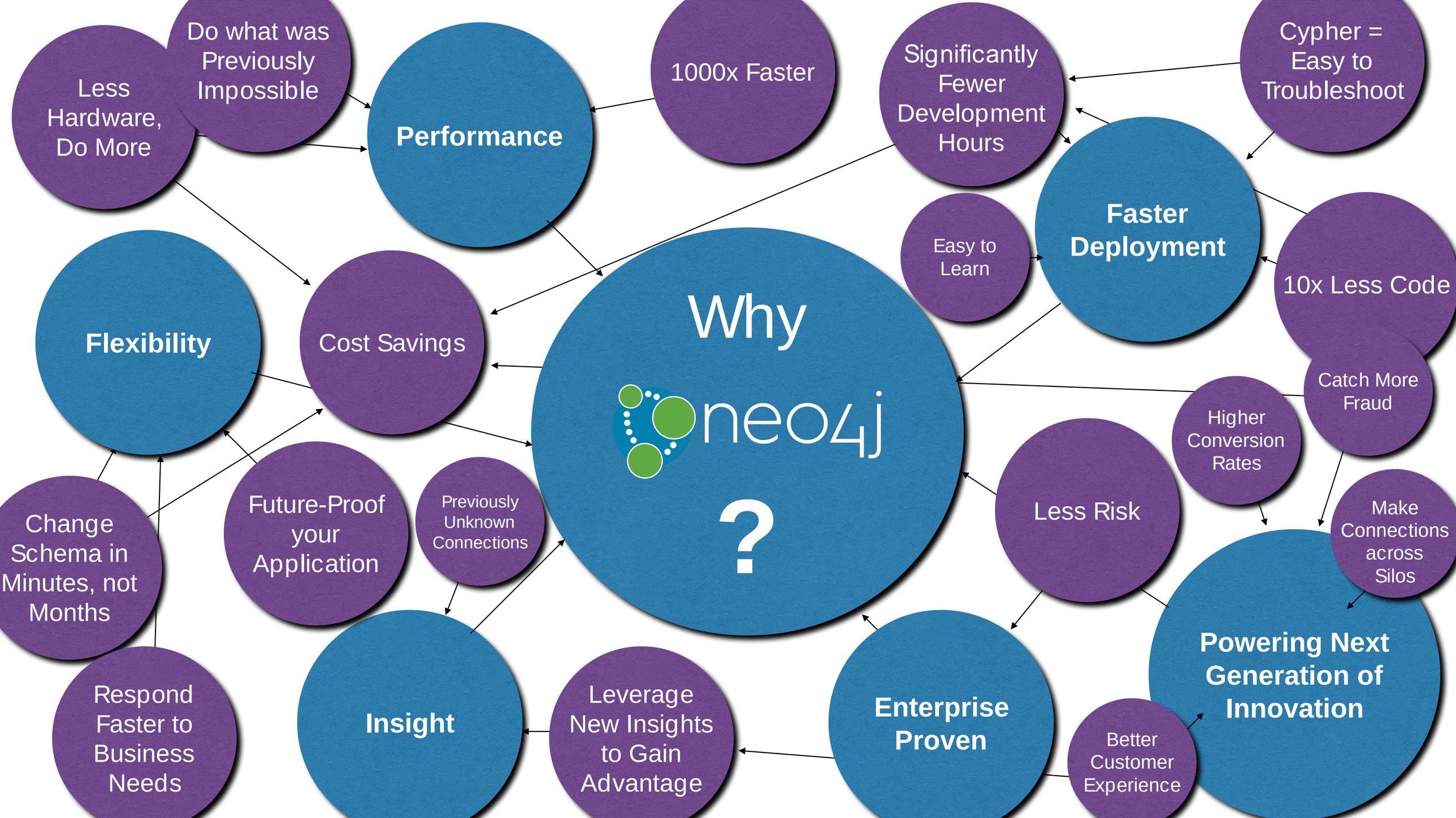
I’m a
Native Graph



I’m still a Native Graph

Non-Native Graph!




Why
neo4j
?
Performance
Do what was Previously Impossible
Less Hardware, Do More
1000x Faster
Significantly Fewer Development Hours
Cypher = Easy to Troubleshoot
Faster Deployment
Easy to Learn
10x Less Code
Flexibility
Cost Savings
Catch More Fraud
Higher Conversion Rates
Less Risk
Make Connections across Silos
Powering Next Generation of Innovation
Better Customer Experience
Enterprise Proven
Leverage New Insights to Gain Advantage
Insight
Respond Faster to Business Needs
Change Schema in Minutes, not Months
Future-Proof your Application
Previously Unknown Connections

neo4j

Common Graph Technology Use Cases

Real-Time Recommendations

Fraud Detection

Network & IT Operations

Master Data Management

Knowledge Graph

Identity & Access Management

Walmart

Itaú

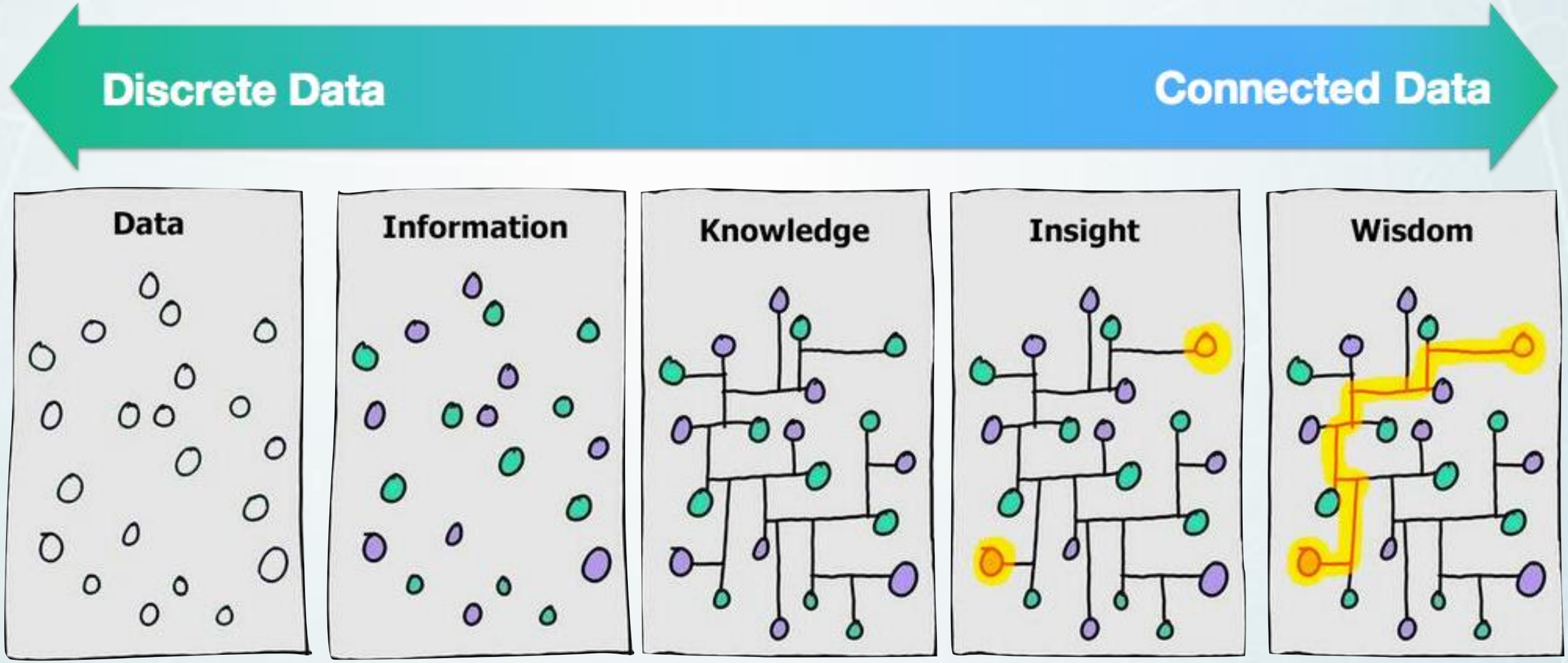
hp

cisco

AirBnb

UBS

A View of the Data Management Portfolio



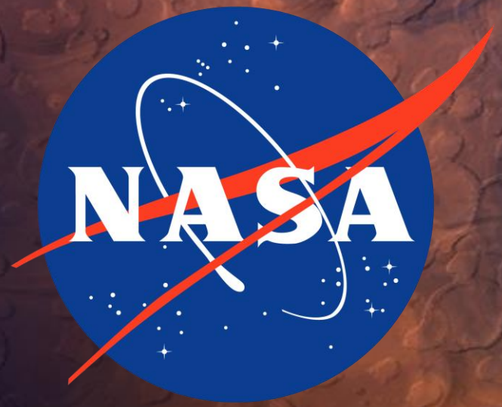
RDBMS & Aggregate-Oriented NoSQL

Hadoop / MapReduce

|<————————- Graph Database & Graph Compute Engine ————————>|

Illustration by David Somerville based on the original by Hugh McLeod (@gapingvoid)

Neo4j Impact at NASA



"Neo4j saved well over two years of work and one million dollars of taxpayer funds."

— David Meza, Chief Knowledge Architect at NASA

TOPIC MODELING

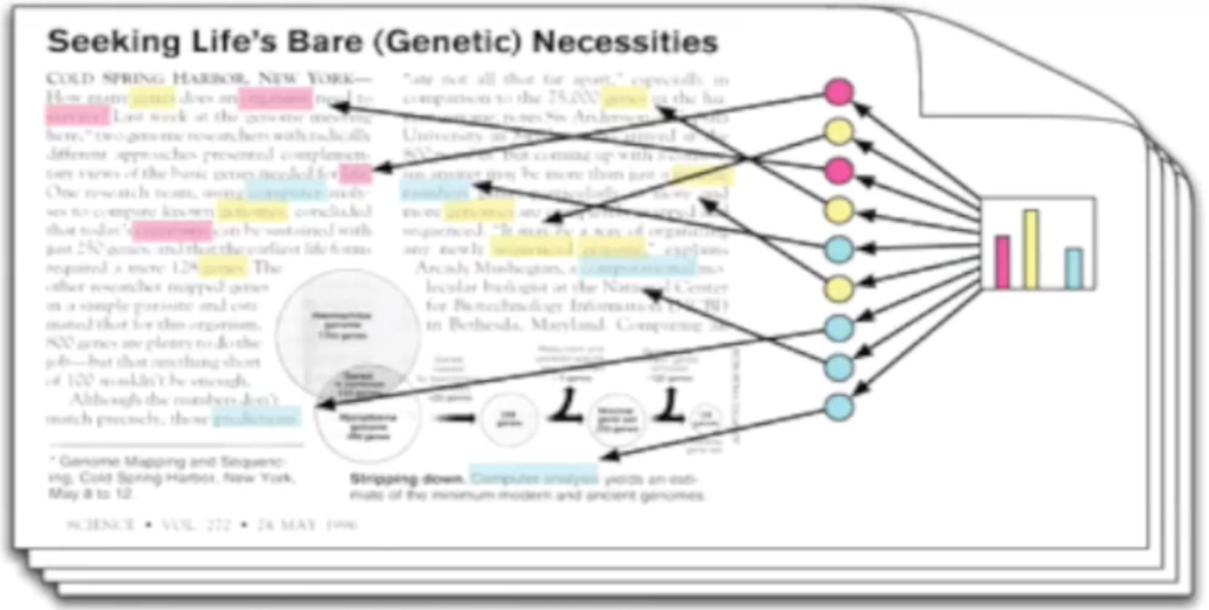
Topics | *Documents* | *Topic proportions and assignments*

gene	0.04
dna	0.02
genetic	0.01
...	

life	0.02
evolve	0.01
organism	0.01
...	

brain	0.04
neuron	0.02
nerve	0.01
...	

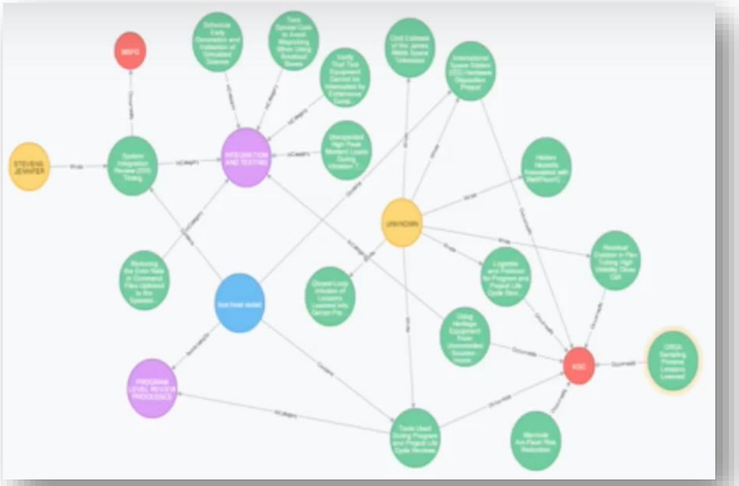
data	0.02
number	0.02
computer	0.01
...	



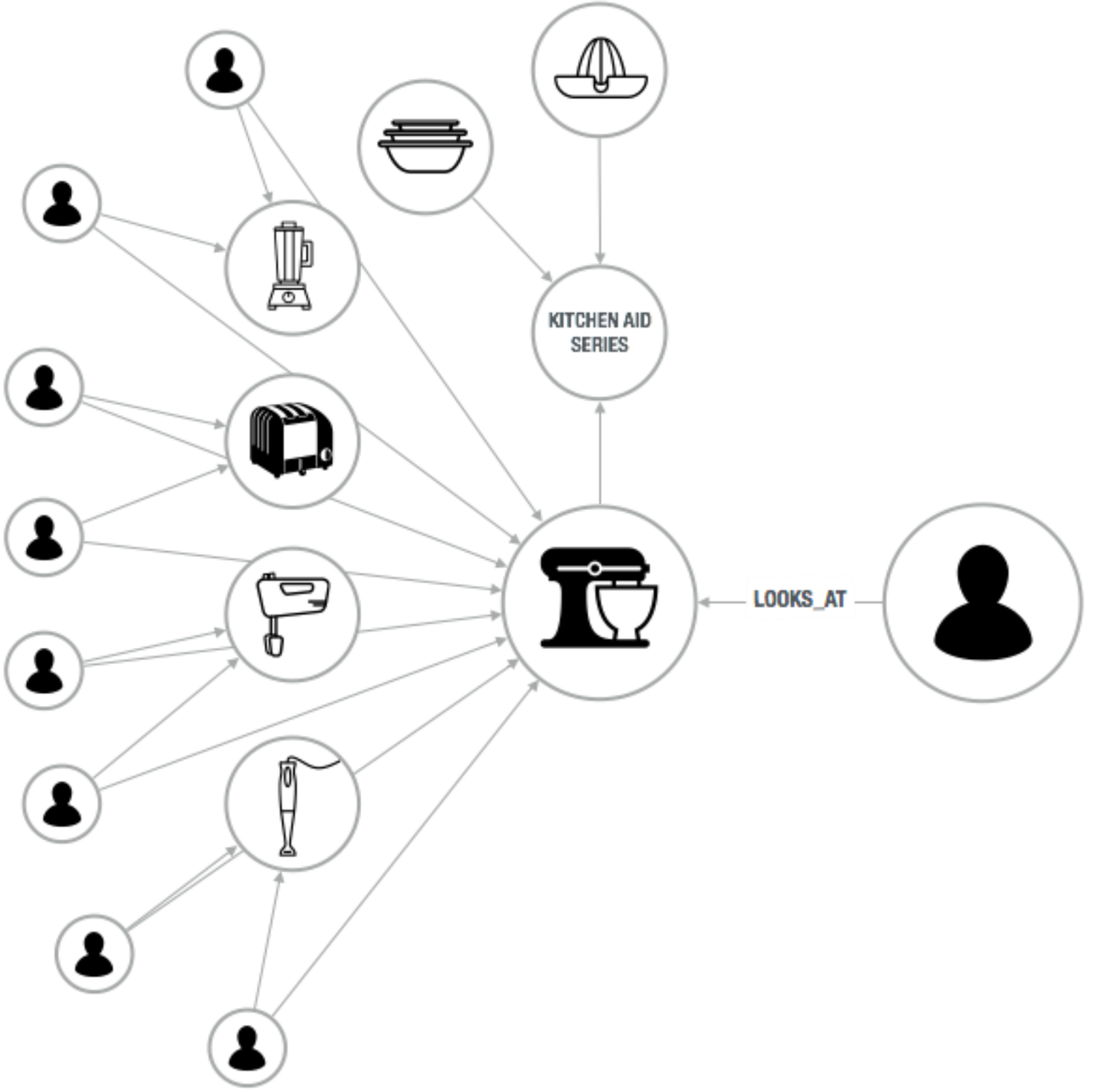
LDA Model from Blei (2011)

Topic models are based upon the idea that documents are mixtures of topics, where a topic is a probability distribution over words.

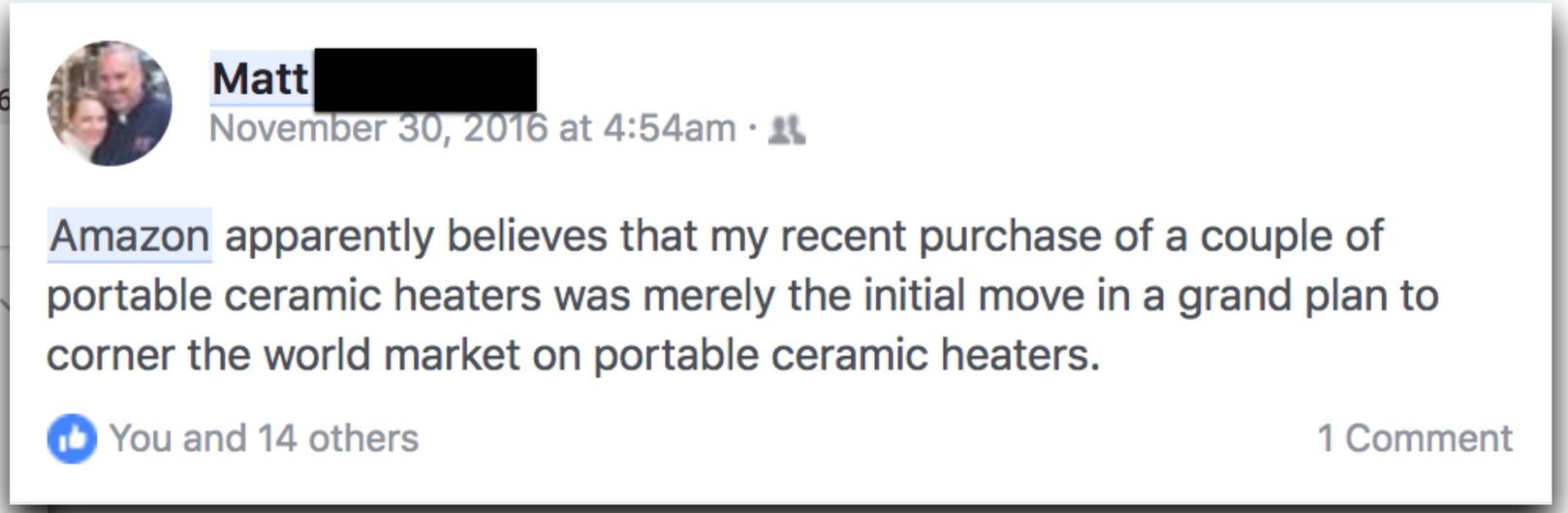
Blei, David M. 2011. "Introduction to Probabilistic Topic Models." *Communications of the ACM*. David Blei homepage - http://www.cs.columbia.edu/~blei/topicmodeling.html



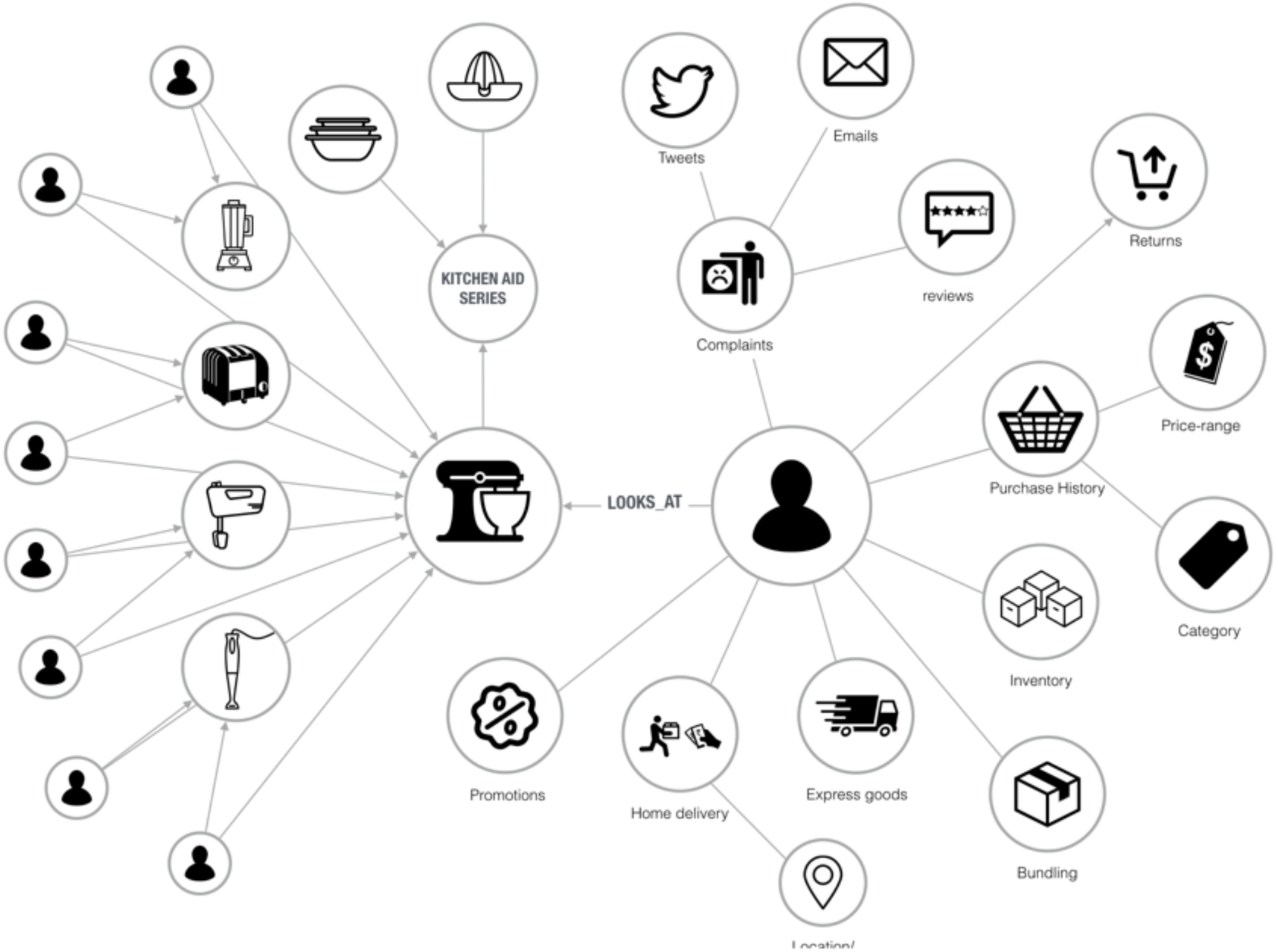
Typical Recommendation Engine



4:46
Tweet
Justin Shanes
@justinshanes
Amazon thinks my recent humidifier purchase was merely the inaugural move in a newfound hobby of humidifier collecting.
3:18 pm · 29 Nov. 16
10.4K Retweets
29.4K Likes


Matt [redacted]
November 30, 2016 at 4:54am ·
Amazon apparently believes that my recent purchase of a couple of portable ceramic heaters was merely the initial move in a grand plan to corner the world market on portable ceramic heaters.
You and 14 others
1 Comment

Recommendations with Neo4j



NEO4j USE CASES

Real Time Recommendations

Master Data Management

Fraud Detection

Graph Based Search

Network & IT-Operations

Identity & Access Management

"Graph databases offer new methods of uncovering fraud rings and other sophisticated scams with a high-level of accuracy, and are capable of stopping advanced fraud scenarios in real-time."

Large Intelligence Agency
Cyber Security Expert

Augmented Fraud Detection

1.

Endpoint-Centric

Analysis of users and their end-points

2.

Navigation Centric

Analysis of navigation behavior and suspect patterns

3.

Account-Centric

Analysis of anomaly behavior by channel

4.

Cross Channel

Analysis of anomaly behavior correlated across channels

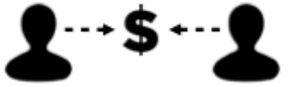
5.

Entity Linking

Analysis of relationships to detect organized crime and collusion

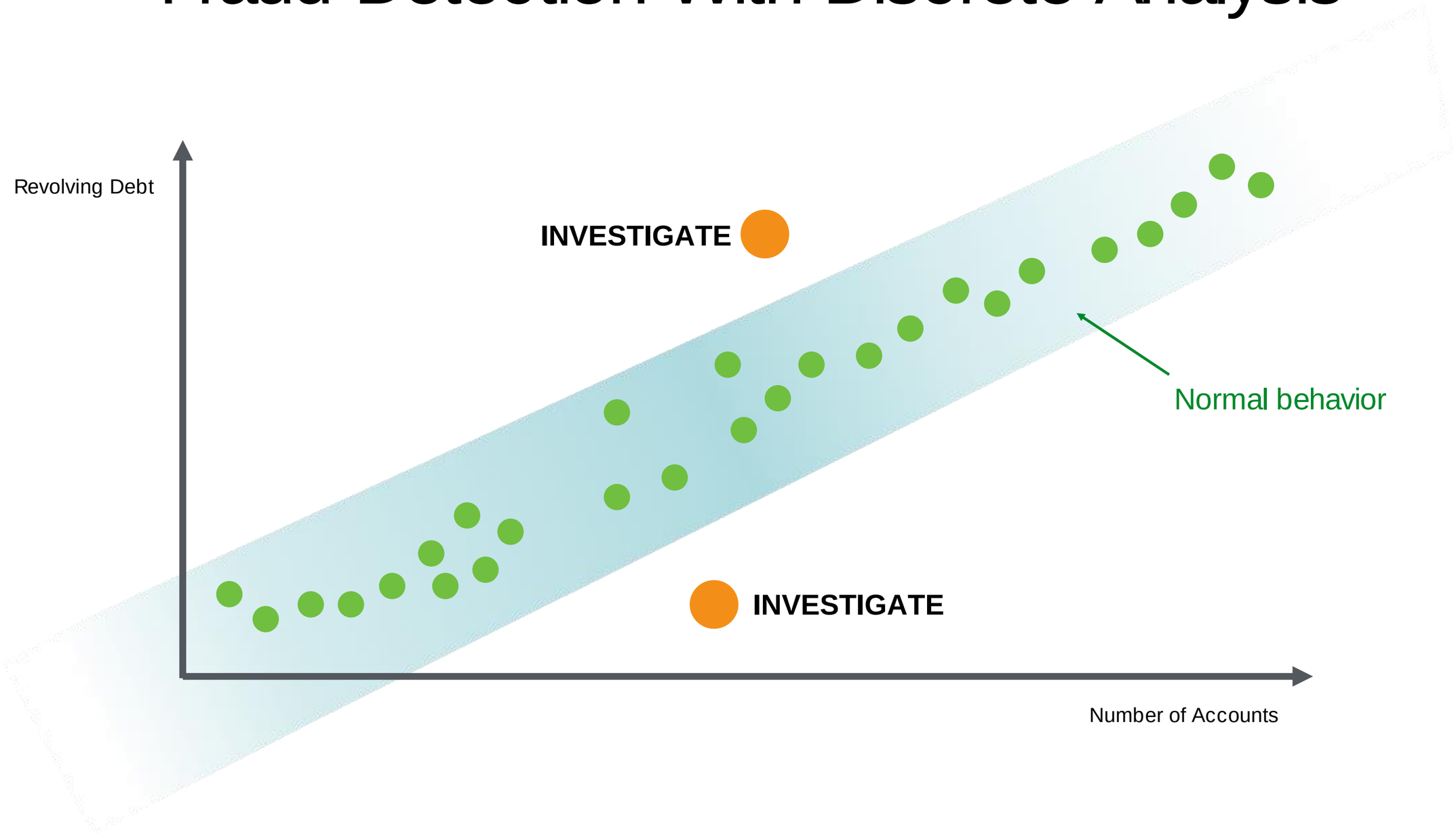
DISCRETE ANALYSIS

CONNECTED ANALYSIS

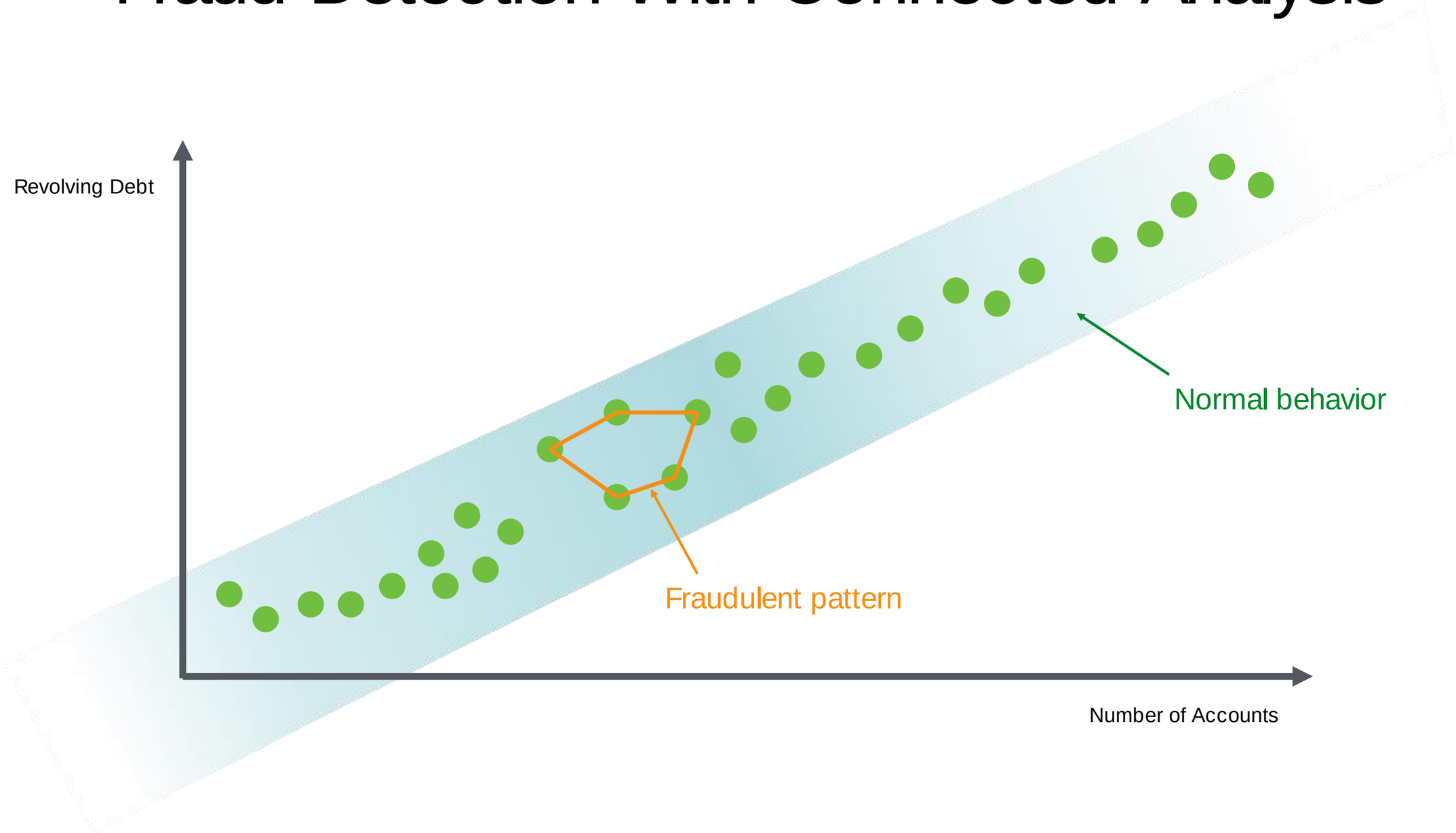


neo4j

Fraud Detection With Discrete Analysis



Fraud Detection With Connected Analysis



ebay

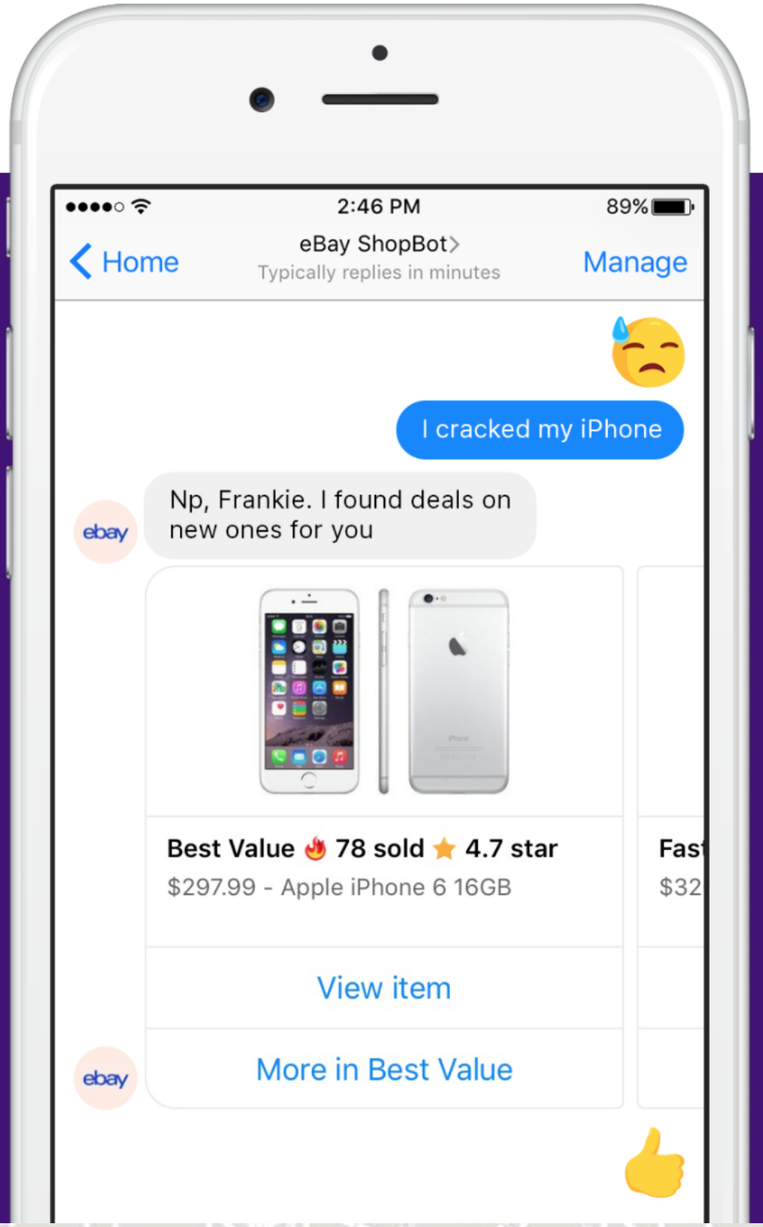
Chat to save

The new, effortless way to find deals

ShopBot is the new way to shop.
It's as effortless as chatting with a friend in Facebook Messenger –
Only this is a friend who works for you,
a friend who knows your sizes,
a friend who saves you money,
gets your taste,
finds you the best deals.

That's a hell of a friend.

+ = Shopping Reinvented

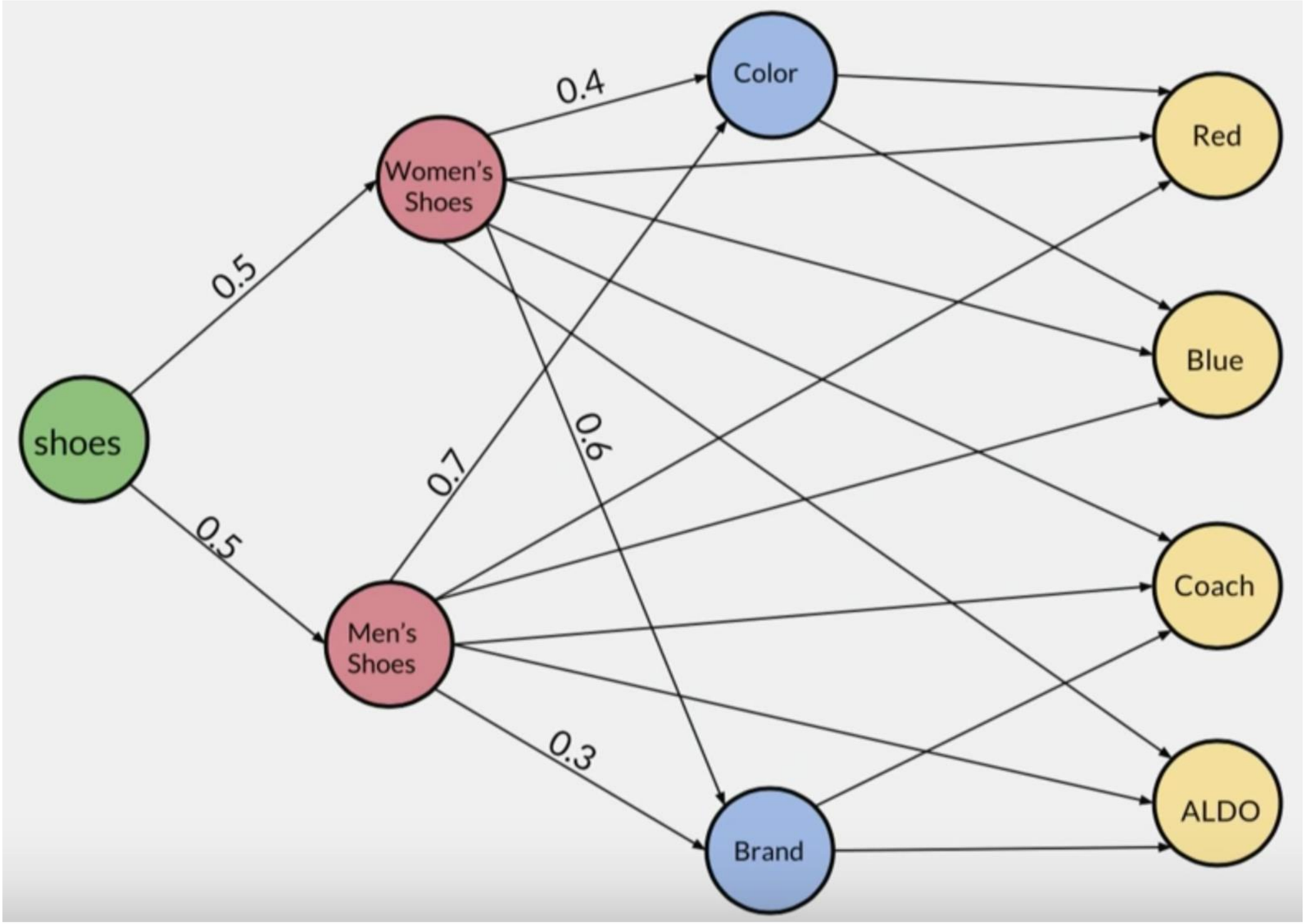


eBay Chat Bot

- There are many Bot Frameworks based on rules engines available
- Encoding rules engines with if-then-else rules may work fine for smaller data sets, but quickly becomes impractical at scale
- A probabilistic Knowledge Graph for inferencing (what question should the bot ask to narrow the recommendation?)
- Data on probabilities come from customer behavioral patterns (tons of data)
- Augment data with additional external datasets
- Machine Learning models to augment data
- ML model as a cache
- Curation to augment data

Natural Language Understanding

Inferencing



Mining the Knowledge Graph Reveals…

I am looking for the **eggplant** **foamposites**.

Color: **Purple**

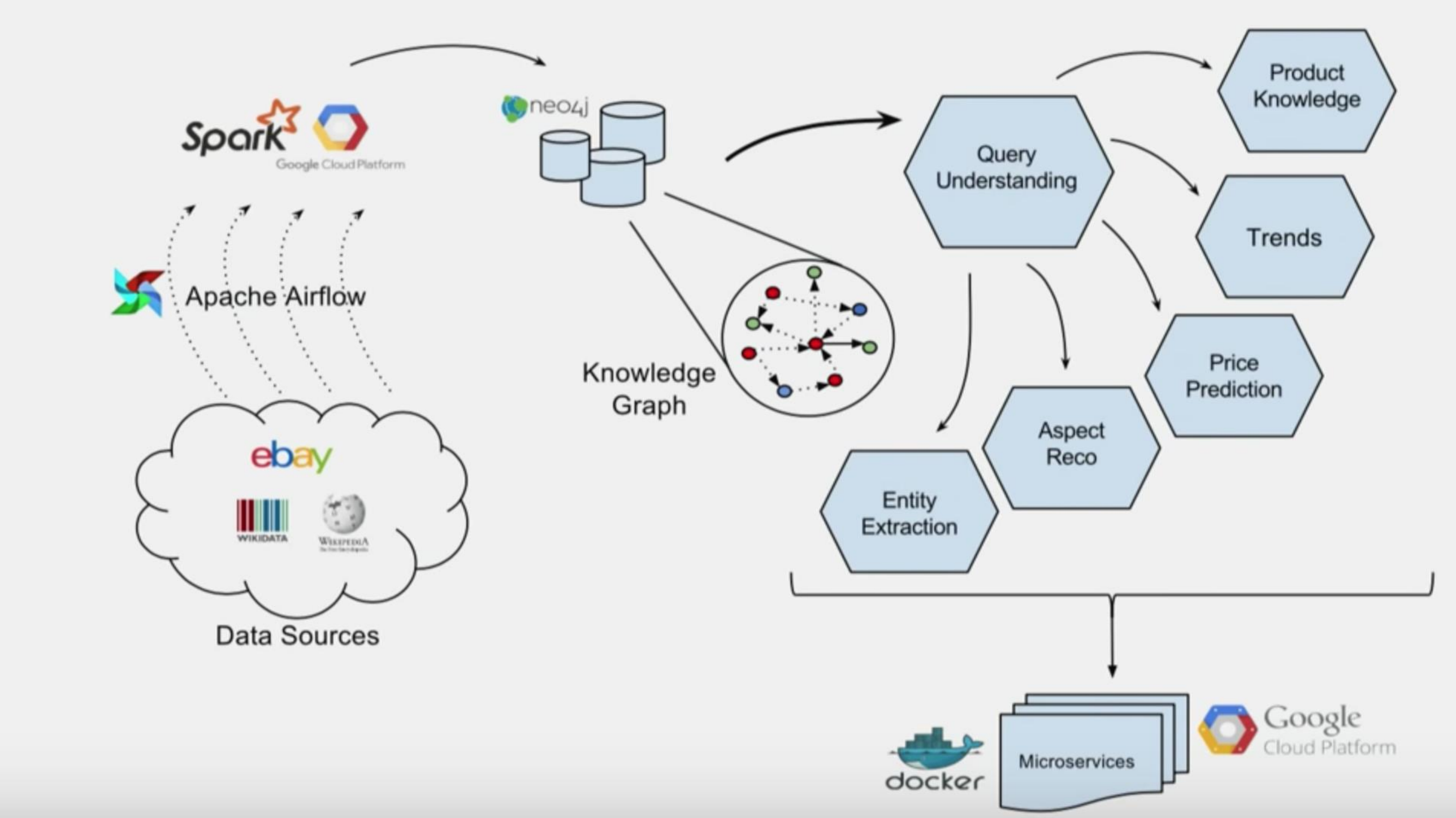
Material: **Foamposite**

Category: **Athletic Shoes**
Product: **Eggplant Foamposite Sneakers**
Brand: **Nike**
Style: **Basketball Shoes**
Release Date: **2009**

Apply insights from Knowledge Graph to better infer intent

I am looking for the **eggplant** **iphone case.**

Color: **Purple** Product: **Phone Case**


Spark
Google Cloud Platform
neo4j
Apache Airflow
Knowledge Graph
Query Understanding
Product Knowledge
Trends
Price Prediction
Aspect Reco
Entity Extraction
ebay
WIKIDATA
Wikipedia
Data Sources
docker
Microservices
Google Cloud Platform