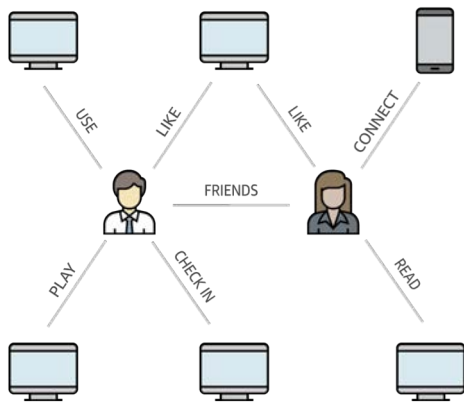


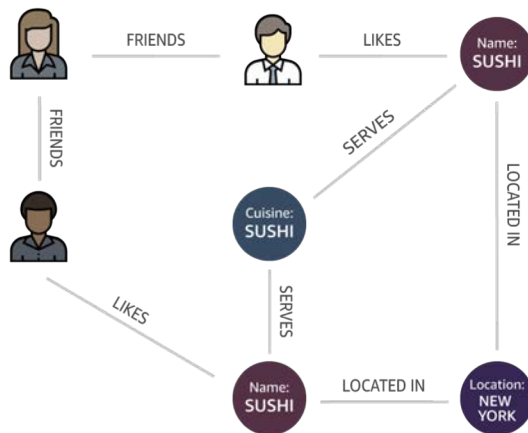
Graph Processing in the Cloud

Brad Bebee, Amazon Neptune, Principal Product Manager, AWS
July 11, 2018

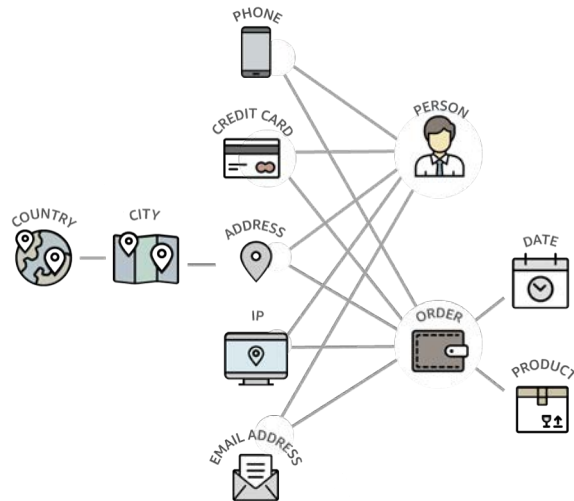
HIGHLY CONNECTED DATA



Social Networks



Restaurant Recommendations



Retail Fraud Detection

USE CASES FOR HIGHLY CONNECTED DATA



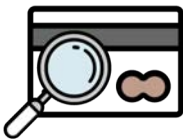
Social Networking



Recommendations



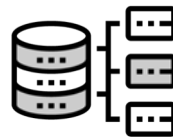
Knowledge Graphs



Fraud Detection

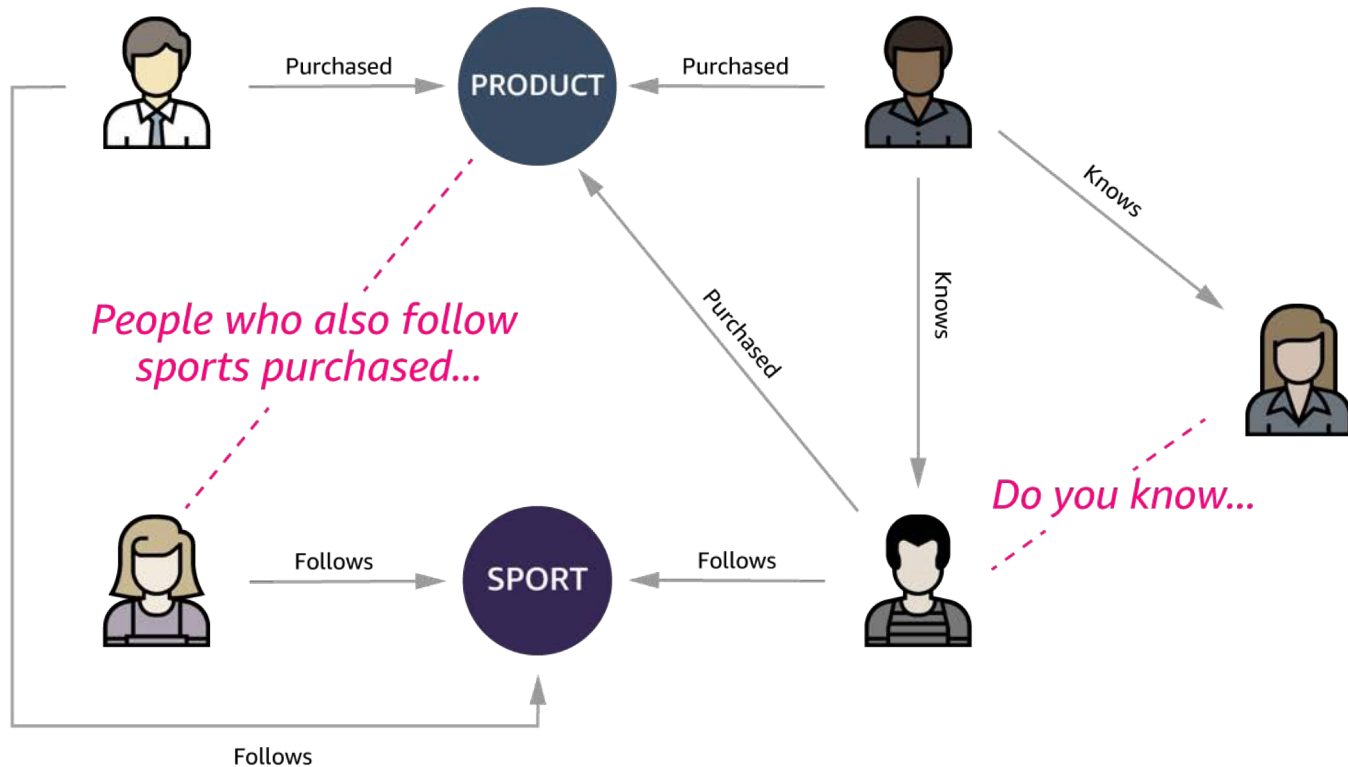


Life Sciences



Network & IT Operations

Recommendations based on relationships

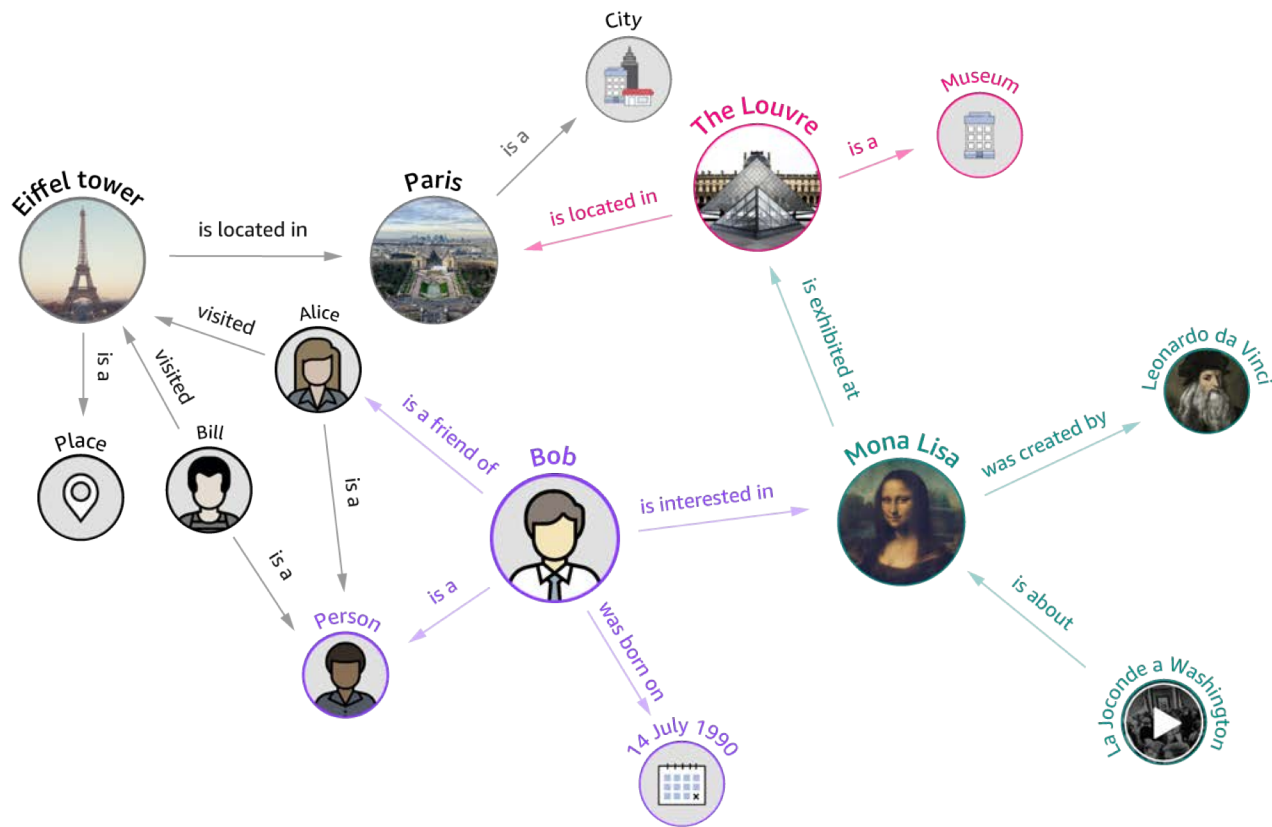


Knowledge Graph Applications

Who painted the Mona Lisa?

What museums should Alice visit while in Paris?

What artists have paintings in The Louvre?



Navigate a web of global tax policies



THOMSON REUTERS

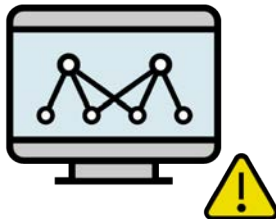
“Our customers are increasingly required to navigate a complex web of global tax policies and regulations. We need an approach to model the sophisticated corporate structures of our largest clients and deliver an end-to-end tax solution. We use a microservices architecture approach for our platforms and are beginning to leverage Amazon Neptune as a graph-based system to quickly create links within the data.”

said Tim Vanderham, chief technology officer, Thomson Reuters Tax & Accounting

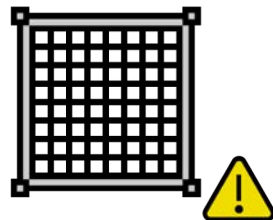
The challenges of building apps with highly connected data using a relational database



Unnatural for
querying graph



Inefficient
graph processing



Rigid schema inflexible
for changing data

Leading graph models and frameworks

PROPERTY GRAPH

Open Source Apache TinkerPop™

Gremlin Traversal Language



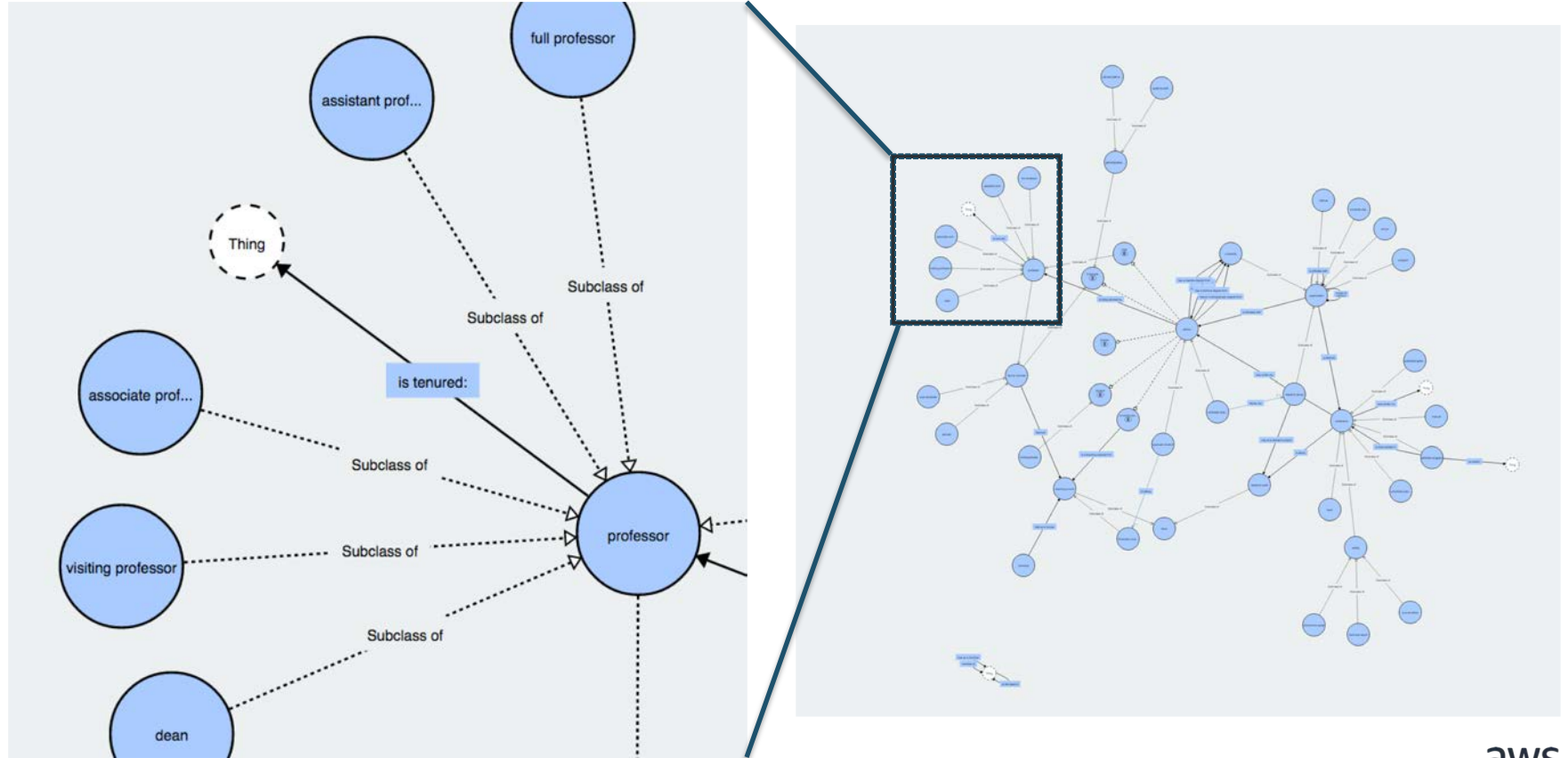
RESOURCE DESCRIPTION FRAMEWORK (RDF)

W3C Standard

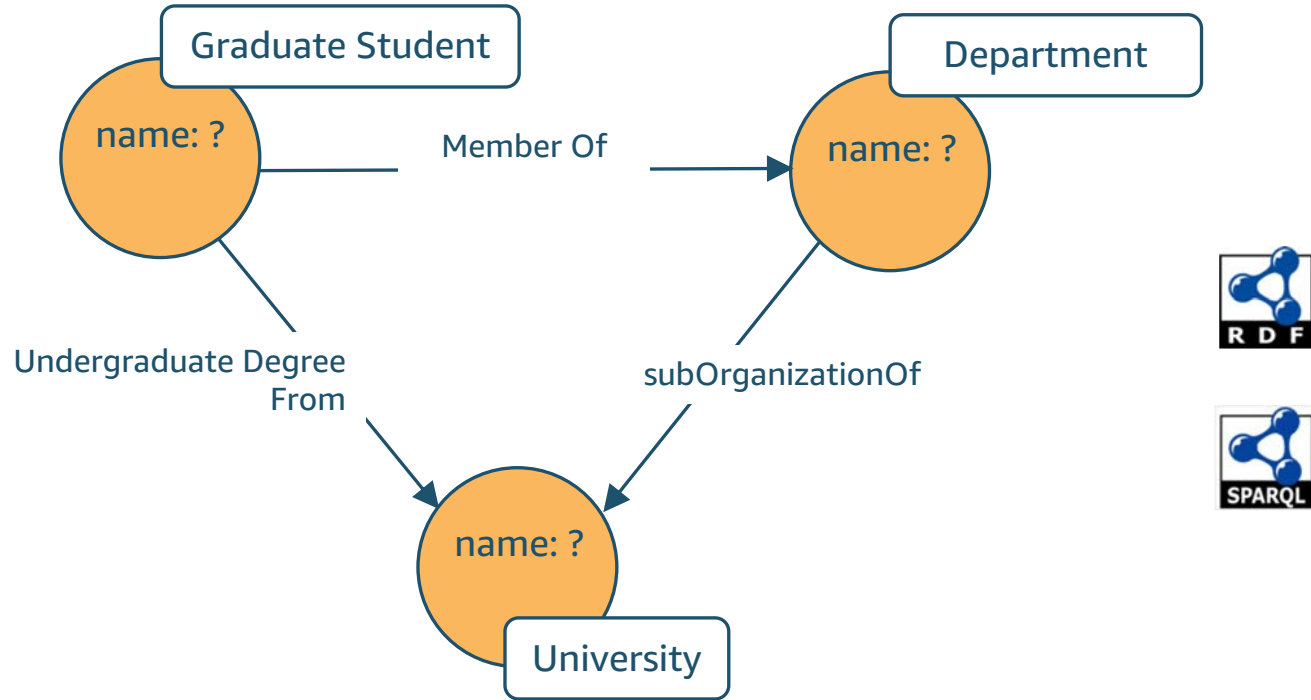
SPARQL Query Language



A highly connected university example



Find all of the graduate students who received an undergraduate degree from the same university



Find all of the graduate students who received an undergraduate degree from the same university

Graduate Student

Department

```
g.V("GraduateStudent").as("student")
  .outE("GraduateStudent_undergraduateDegreeFrom_University").inV()
  .inE("Department_subOrganizationOf_University").outV()
  .inE("GraduateStudent_memberOf_Department").outV()
  .where(eq("student"))
```



name: ?

University

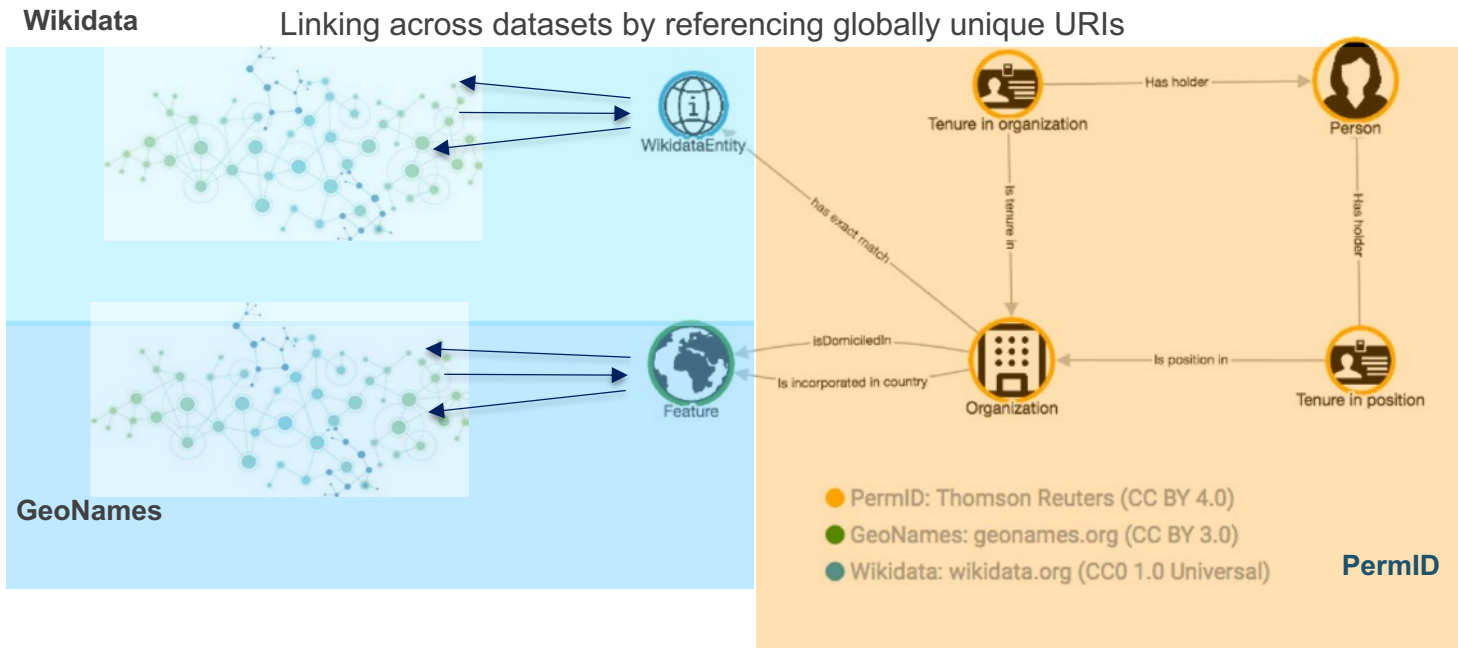
Find all of the graduate students who received an undergraduate degree from the same university

```
PREFIX rdf:http://www.w3.org/1999/02/22-rdf-syntax-ns#
PREFIX ub:http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#

SELECT ?student WHERE {
    ?student rdf:type ub:GraduateStudent .
    ?univ rdf:type ub:University .
    ?dept rdf:type ub:Department .
    ?student ub:memberOf ?dept .
    ?dept ub:subOrganizationOf ?univ .
    ?student ub:undergraduateDegreeFrom ?dept
}
```

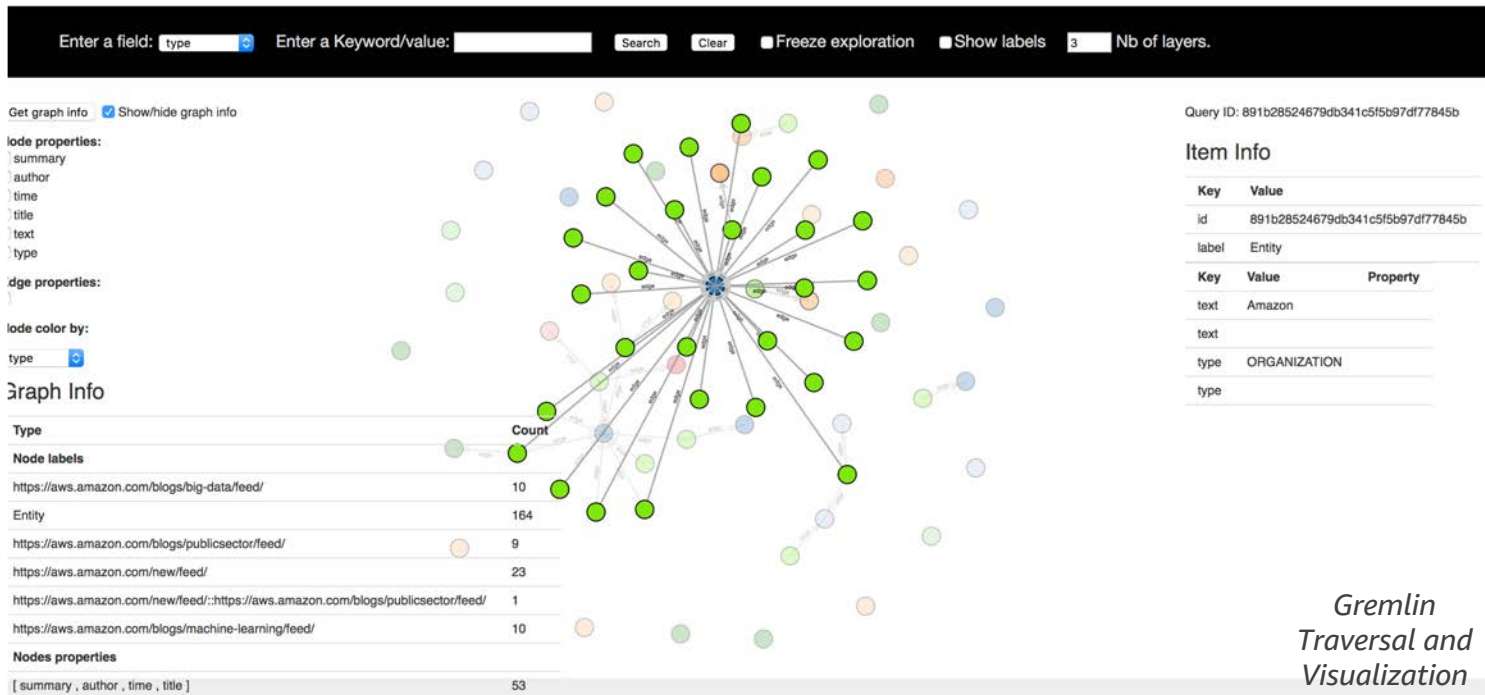


The benefits of Linked Data



Example: PerMID (re)uses **<<http://sws.geonames.org/6252001/>>**
as a global Identifier for the USA, which is an identifier rooted in GeoNames.

Graph is complementary to ML and analytics



Entity Extraction
from RSS Feeds



Amazon Simple
Storage Service
(S3)

Load from S3 into
Neptune



Amazon
Neptune

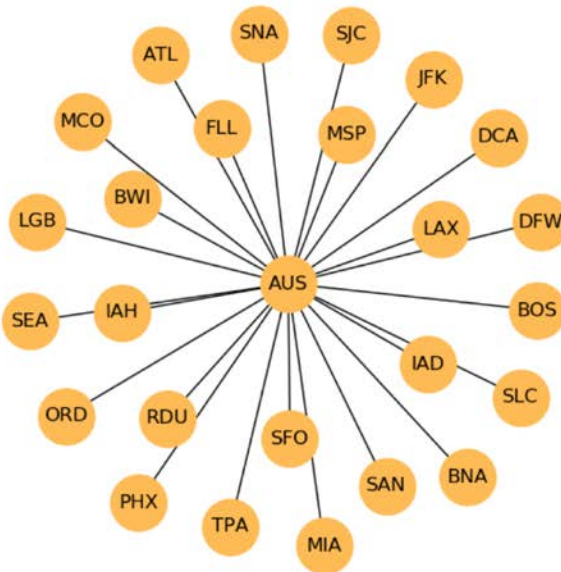
Graphexp



Using TinkerPop Gremlin Traversals with Jupyter Notebooks

```
In [337]: import matplotlib.pyplot as plt; plt.rcParams['figure.figsize'] = (5,5)
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import networkx as nx

G=nx.Graph()
#G.add_node('AUS')
G.add_nodes_from(verts)
for s in verts:
    G.add_edge('AUS',s)
#print(verts)
plt.figure(figsize=(5,5))
nx.draw(G, node_color="#ffbb55", node_size=1200,with_labels=True)
plt.show()
```

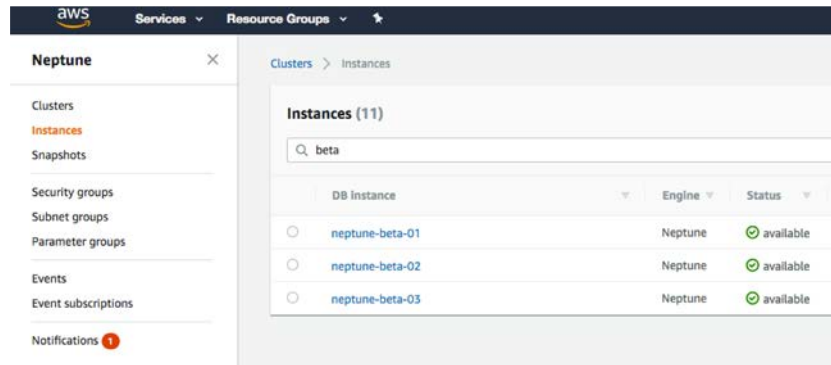
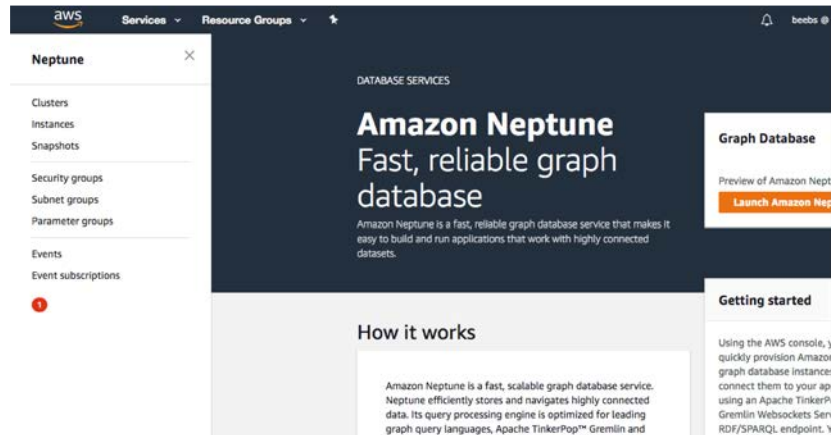


Neptune GA Customers



Neptune General Availability

- Announced on 5/30/2018
- RDF and Property Graph
- Cloud-native
- Four regions
 - US East (No. Virginia), US East (Ohio), US West (Oregon), EU West (Ireland)
- <https://aws.amazon.com/about-aws/whats-new/2018/05/amazon-neptune-is-now-generally-available/>



Thank you!

Brad Bebee

beeb@amazon.com