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

People who also follow sports purchased...

Do you know...
Knowledge Graph Applications

Who painted the Mona Lisa?

What museums should Alice visit while in Paris?

What artists have paintings in The Louvre?
“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

1. Unnatural for querying graph
2. Inefficient graph processing
3. 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

```
g.V("GraduateStudent").as("student")
.outE("GraduateStudent_undergraduateDegreeFrom_University").inV()
.inE("Department_subOrganizationOf_University").outV()
.inE("GraduateStudent_memberOf_Department").outV()
.where(eq("student"))
```
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:undergraduateDegreeFrom ?dept
}
```
The benefits of Linked Data

Linking across datasets by referencing globally unique URIs

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

Amazon Comprehend → 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
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)
Thank you!

Brad Bebee
beebs@amazon.com