Network Graph Databases, RDF, SPARQL, and SNA

NoCOUG Summer Conference
August 16 2012 at Chevron in San Ramon, CA

David Abercrombie
Data Analytics Engineer, Tapjoy
david.abercrombie@tapjoy.com
About me

Former (Oracle)
  Performance
  Diagnostics
  Capacity planning

Current (Vertica, column store, parallel, BI)
  Data structure design & SQL
  DBA

Graph databases are a hobby
Big Data SF Bay Area
(a LinkedIn Group)

6:30 to 9:00pm **TONIGHT**!
Santa Clara

Dave Rubin, Oracle
Oracle’s NoSQL

Brian Clark, Objectivity
InfiniteGraph
NoSQL Now!
August 21-23, San Jose

Keynote: Creating Powerful New Applications with Graphs
Tutorial: Introduction to Graph Databases
Intro to Graph Databases 101
Transform Big Data into Actionable Intelligence with Graphs
Use Graph Databases to Analyze Relationships, Risks, and Business Opportunities – A Case Study
Lunch ‘N Learn with Neo4j
Oracle Database 11g

Spatial and Graph RDF Semantic Graph

RDF management platform
Graph data model
RDF data (triples) are persisted, indexed and queried

June 2010 Oracle workshop presentation PDFs
An Enterprise Inference Engine Inside Oracle...
... How to Install, Load, Query and Inference
Resource Description Framework

RDF
Resource Description Framework

A "language for representing information about resources in the World Wide Web"

Generalized:

- From directly retrieved on the Web
- To identified on the Web

Applications, not humans
Exchange
Subject – Object - Predicate

Simple properties and property values

Must be well defined

Uniform Resource Identifier (URI)
http://www.example.org/index.html has a creator whose value is John Smith

Subject
http://www.example.org/index.html

Predicate
http://purl.org/dc/elements/1.1/creator

Object
http://www.example.org/staffid/85740

This and the next several slides based on http://www.w3.org/TR/rdf-primer/
As a directed graph

http://www.example.org/index.html

http://purl.org/dc/elements/1.1/creator

http://www.example.org/staffid/85740
More Details

http://www.example.org/staffid/85740

http://www.example.org/terms/address

http://www.example.org/addressid/85740

http://www.example.org/terms/city

Bedford

http://www.example.org/terms/street

1501 Grant Avenue

http://www.example.org/terms/postalCode

01730

http://www.example.org/terms/state

Massachusetts
## Triples

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>staff:85740</td>
<td>terms:address</td>
<td>addressid:85740</td>
</tr>
<tr>
<td>addressid:85740</td>
<td>terms:street</td>
<td>&quot;1501 Grant Avenue&quot;</td>
</tr>
<tr>
<td>addressid:85740</td>
<td>terms:city</td>
<td>&quot;Bedford&quot;</td>
</tr>
<tr>
<td>addressid:85740</td>
<td>terms:state</td>
<td>&quot;Massachusetts&quot;</td>
</tr>
<tr>
<td>addressid:85740</td>
<td>terms:postalCode</td>
<td>&quot;01730&quot;</td>
</tr>
</tbody>
</table>

![Triples Diagram](image-url)
Serialization

Representing as a file
  RDF/XML
  Notation3 (N3)
  Turtle
  N-Triples
Turtle Serialization

@prefix staff: <http://www.example.org/staffid> .
@prefix terms: <http://www.example.org/terms> .
@prefix addressid: <http://www.example.org/addressid> .

staff:85740 terms:address addressid:85740 .
staff:85740 terms:address addressid:85740 .
addressid:85740 terms:street "1501 Grant Avenue" .
addressid:85740 terms:city "Bedford" .
addressid:85740 terms:state "Massachusetts" .
addressid:85740 terms:postalCode "01730" .
SPARQL Protocol And RDF Query Language

SPARQL
SELECT ?who
WHERE { :JamesDean :playedIn ?what .
         ?what :directedBy ?who .}
James Dean

Rebel Without a Cause
  directedBy Nicolas Rey

Giant
  directedBy George Stevens
  directedBy Fred Guiol

East of Eden
  directedBy Elia Kazan

James Dean played in what directed by who

SPARQL query as a directed graph
SELECT ?who
WHERE { :JamesDean :playedIn ?what .
?what :directedBy ?who .}
Properties and Schema

What kinds of things do we know about James Dean?

```
SELECT DISTINCT ?property
WHERE { :JamesDean ?property ?value }
```

What kinds of things do we know about actors?

```
SELECT DISTINCT ?property
WHERE { ?q0 a :Actor .
    ?q0 ?property ?object . }
```

See Working Ontologist 2012, pp. 78-83
Other SPARQL Features

FILTER
OPTIONAL
UNSAID
ASK
CONSTRUCT

Rules: when you see this, conclude that (define properties)

Aggregates
Subqueries
Transitive (like Oracle’s CONNECT)

ORDER BY
LIMIT
UNION

See Working Ontologist 2012, pp. 83-112
RDF Schema, Ontologies, and Inferencing

RDFS
Inference

Given some useful information, we can derive other related information

Asserted triples

Inferred triples

Inference rules

Inference engine
Men’s Trail Running shoes are a type of Fitness Footwear. Fitness Footwear is a type of Fitness product.
rdfs:subClassOf

:FitnessProduct rdf:type rdfs:Class .

:FitnessFootwear rdfs:subClassOf :FitnessProduct .

:MenTrailRunningShoe rdfs:subClassOf :FitnessFootwear .
rdfs:subPropertyOf

:freeLancesTo rdfs:subPropertyOf contractsTo .
:indirectlyContractsTo rdfs:subPropertyOf :contractsTo .
:isEmployedBy rdfs:subPropertyOf :worksFor .
:contractsTo rdfs:subPropertyOf :worksFor .

Example from Working Ontologist 2012, pp. 128-9
rdfs:subPropertyOf


Inferred: :Dave :contractsTo :TheFirm .

Example from Working Ontologist 2012, pp. 128-9
RDFS domains
(properties and classes)

Domain of square root is positive real numbers

For properties "P" and classes "D", we have

```
CONSTRUCT { ?x rdf:type ?D . }
   ?x ?P ?y . }
```
\texttt{MarriedWoman} \texttt{rdfs:subClassOf Woman}.

\texttt{hasMaidenName} \texttt{rdfs:domain MarriedWoman} \texttt{rdfs:subClassOf Woman}.

\texttt{:MarriedWoman rdfs:subClassOf :Woman}.
\texttt{:hasMaidenName :rdfs:domain :MarriedWoman}.

Example from \textit{Working Ontologist} 2012, pp. 130-2
If we assert:

`:Karen :hasMaidenName "Abercrombie" .

We can infer:

`:Karen rdf:type :MarriedWoman .
`:Karen rdf:type :Woman .

Example from Working Ontologist 2012, pp. 130-2
Social Network Analysis

SNA
Ding, Finin, and Joshi found these basic patterns in Friend-of-a-Friend networks.
Indirect connections

Person: Alice

Concept: Surfing

Person: Bob

Kinsella, Harth, and Breslin studied networks of objects and people
Ontologies

**FOAF**: Friend of a Friend
Knows, based_near, interest

**SIOC**: Semantically-Interlinked Online Communities
Site, Post, Thread, Forum

**SCOT**: Social Semantic Cloud of Tags
Tagging, Web 2.0
Social Network Analysis (SNA)

Bridges
Centrality
Clustering
Cohesion
Density
Visualization
AllegroGraph lisp API examples

Graphs Measures
  nodal-degree
  nodal-neighbors
  ego-group
Centrality Measures
  graph-density
  actor-degree-centrality
  group-degree-centrality
  actor-closeness-centrality
  group-betweenness-centrality
Cliques
  cliquep
  cliques
  map-cliques
The End

REFERENCES
Graph Databases

Oracle **Spatial and Graph RDF Semantic Graph**
(Formerly Oracle Database Semantic Technologies)

**InfiniteGraph** (Objectivity)

W3C **LargeTripleStores**
  AllegroGraph (1+Trillion)
  OpenLink Virtuoso v6.1 - 15.4B+
  BigOWLIM (12B explicit, 20B total)
  Garlik 4store (15B)
  Bigdata(R) (12.7B)
  YARS2 (7B)
  Jena TDB (1.7B)
  Jena SDB (650M) ...
References

RDF Primer (HTML)
Semantic Social Network Analysis (PDF)
Network Analysis of Semantic Connections in Heterogeneous Social Spaces (PDF)
Analyzing Social Networks on the Semantic Web (PDF)
Network Analysis of Semantic Web Ontologies (PDF)
AllegroGraph 4.6 Overview (HTML)
AllegroGraph on Amazon EC2 (HTML)
What did we talk about?

Upcoming meetings
RDF
SPARQL
RDFS
SNA