Oracle Spatial and Graph Graph Overview

Bill Beauregard
Senior Principal Product Manager

Matt Perry
Principal Member of Technical Staff
Agenda

• Oracle’s Graph Database Strategy
• Introduction to RDF Graph
• Use Cases
• Feature Overview & Performance
• Summary
Oracle’s Graph Database Strategy

Support Graph Data Types…

- Add graph analytics to applications, tools, and information technology platforms
- Deliver a scalable, secure, and high performing product
- Simplify development with integrated graph analysis, APIs and services

…On all enterprise platforms

- Oracle Database
- Cloudera with Apache Hadoop
- Oracle NoSQL Database
- Oracle Big Data Appliance
- Oracle Exadata Database Machine
- Oracle Cloud
# 3 Graph Models / 3 Domain Use Cases

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Graph Model</th>
<th>Industry Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Network Analysis</td>
<td><strong>Network Data Model</strong></td>
<td>Logistics, Transportation, Utilities, Telcoms</td>
</tr>
<tr>
<td></td>
<td>• Network path analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Multi-model modeling</td>
<td></td>
</tr>
<tr>
<td>Linked Data / Semantic Mediation</td>
<td><strong>RDF Data Model</strong></td>
<td>Life Sciences, Finance, Publishing, Public Sector</td>
</tr>
<tr>
<td></td>
<td>• Data federation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Knowledge representation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Master Metadata Mgmt</td>
<td></td>
</tr>
<tr>
<td>Social Network Analysis</td>
<td><strong>Property Graph Model</strong></td>
<td>National Intelligence, Public Safety, Social Media search, Marketing - Sentiment</td>
</tr>
<tr>
<td></td>
<td>• Graph Search &amp; Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Big Data analytics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Entity analytics</td>
<td></td>
</tr>
</tbody>
</table>

- Copyright © 2015, Oracle and/or its affiliates. All rights reserved.
RDF Graph Use Cases

- **Semantic Metadata Layer**
  - Unified content metadata for federated resources
  - Validate semantic and structural consistency

- **Text Mining & Entity Analytics**
  - Find related content & relations by navigating connected entities
  - “Reason” across entities

- **Social Media Analysis**
  - Analyze social relations using curated metadata
    - Blogs, wikis, tweets, video
    - Calendars, IM, voice
Graph Metadata Mapping

**Benefits**
- Existing relational data stays in place and corresponding applications do not need to change
- Use of virtual mapping eliminates synchronization issues
- Common vocabulary helps with data integration issues

Diagram:
- Mid-Tier Server
  - Application 1
  - Application 2
  - Application 3
- Database Server
  - HR Database
  - Inventory Database
  - Sales Database
- SQL
- RDF Graph
- SPARQL
- HR Schema
- Inventory Schema
- Sales Schema
- Shared Ontologies
Linked Data in Enterprise

Access & Presentation Layer

Semantic Graph model
(W3C RDF Metadata Model)

Data Servers

Data Sources / Types

Machine Generated Data
Social Media
Human Sourced Information
Subscription Services
Transaction Systems

Event Server
Hadoop Appliance
Content Mgmt
BI Server
Data Warehouse
Italian Statistical Agency (ISTAT)
Italian National Institute of Statistics

Linked Open Data (LOD)

OVERVIEW

• Istat is the Italian National Institute of Statistics, producing Official Statistics serving the government, the economy and the public with population and housing data characterizing the economic, demographic, social and environmental situation

CHALLENGES / OPPORTUNITIES

• Make Census data broadly available to non-technical users
• Enable schema flexibility and simplify data management for a 402,903 Census Sections data warehouse
• Complex data production workflow
• Multiple data models: DDI, Neuchâtel, SDMX, GSIM

SOLUTION

• Oracle Database 12c (12.1) Enterprise Edition
• Oracle Spatial and Graph 12c RDF Semantic Graph
• Vocabularies (ontologies) to unify disparate datasets, define data provenance, and enable machine driven discovery of inferred relationships
• Front-end open source access points
• SPARQL Web Services endpoint for advanced users & machine-to-machine communications
• Faceted/Graph browser for non-technical users
• GUI for ad-hoc dataset downloads by non-technical users

RESULTS:

• A “5 star” W3C OpenData model Endpoint service
• Open: Linked Open Data system EU/Italian Gov Agency compliance
• Scalable:
  • 1.5B Triples: 700M asserted, 885 inferred
  • Ontology: 95 classes & 200 relationship types
• Interoperable: with Italian federated Linked Open Data Endpoints
• Extensible: Endpoint is a central element of Istat dissemination architecture
• Pioneering Linked Open Data in the European Statistical System (ESS)
• W3C Standards-based: RDF, SPARQL, OWL2 profiles
Objectives

- Common metadata model supports:
  - Search and discovery of EU Publications
  - Multiple domains and languages

Solution

- Validate and tag EU law, tenders, and publicity to standardized vocabularies
- Unified RDF graph metadata model
- Supports discovery of content through user’s terminology and language
- Provides variety of dissemination modes

Benefits

- Evolving data model that flexibly supports a variety of business use cases
- Scalability:
  - Over billion RDF triples in Oracle Graph DB
  - 2.5 TB of compressed data in Oracle DB
  - Links to 3.9 TB (60M) files of EU pubs
- Reliability and maintainability
  - Oracle ASM (Automatic Storage Management)
  - Two failover systems
Eli Lilly and Company

Oracle Spatial and Graph: RDF Graph Metadata Repository

Objectives

- Unified vocabulary for scientific investigation
- Easier, more complete investigations

Solution

- Integrate patient records, chemical structures, biological sequences & pathways, images, scientific papers…
- View related data as a graph
- Traverse graphs to discover relationships, search for a term, or browse ontologies

“[This technology…] provides improved insight into our business by bringing together related information from diverse data sources,”

J. Phil Brooks
Information Consultant, Eli Lilly and Company
Novartis Institutes for BioMedical Research (NIBR) Project Metastore

BRIEF ORGANIZATIONAL OVERVIEW

NIBR is the global pharmaceutical organization for Novartis committed to discovering innovative medicines to treat diseases with high unmet medical need.

6000+ scientists, physicians, business professionals worldwide.

BUSINESS CHALLENGES / OPPORTUNITIES

Scientific knowledge portal does not provide ability to:

- Link database information on genes, proteins, metabolic pathways, compounds, ligands, etc. to original sources.
- Increase productivity for accessing, sharing, searching, navigating, cross-linking, analyzing internal/external data.

SOLUTION

Provide a semantic integration layer on existing relational tables:

- Rich domain-specific terminology (biology, chemistry, and medicine) containing 1.6 M terms.
- Terminology Hub: 8 GB of referential data that cross-references between data repositories.

BUSINESS BENEFITS REALIZED BY ORACLE SOLUTION

- Performance: met design goals for comparable performance to legacy application with enhanced usability.
- Better Analysis: enables discovery of unknown relationships based on the meaning (the semantics) of the data.
- Flexible Data Modeling: supports discovery and allows easy changes to incorporate new kinds of data and relationships.
- Manageability: RDF triple store benefits from Oracle Database functionality: backup, security, replication, …
Cisco WebEx Social
Graph for Enterprise Collaboration

Objectives
- Social connectivity and collaboration through semantic enablement
- Connect knowledge silos

Solution
- Persistent unified graph metadata model
- Concepts tagged with unique meaning
- Find related content & groups by navigating connected entities, recommendations

Benefits
- Unifies metadata model - forum, blog, wiki, etc.
- Tagging media documents, pictures, blogs, etc. to user-defined and/or enterprise vocabularies.
- Validates tag semantic/structural consistency
Objectives

- Profile suspects through telephone, email and social network communications
- Produce “data products” for analysts

Solution

- RDF Graph modeling of the social network: people, groups and places of interest
- Inferencing & graph analytics discover relationships among individuals & meaning of pseudonyms, aliases, codes, terminology

Benefits

- Standards-based tools: W3C RDF & SPARQL
- Semantic tagging for 600 TB / 10b triples graph
- Top-secret, compartmented security for data
- New discovery on ~100 million triples / month
- Find & label “same-as” relationships
Industries Have Already Adopted the Concept

- Life Sciences
- Finance
- Media
- Networks & Communications
- Defense & Intelligence
- Public Sector
RDF GRAPH DATABASE FEATURES
Oracle Database 12c Spatial and Graph Tooling

Transform & Modeling Tools
- R2RML
- Protégé for Oracle
- Jena for Oracle
- Sesame for Oracle
- NLP Extraction (partners)

Load, Query & Inference
- RDF/OWL Data Management
- SQL & SPARQL Query
- OWL Inferencing
- Semantic Rules
- Scalability & Security
- Semantic Indexing

Applications & Analysis Tools
- Java, HTTP access
- JSON, XML output
- Graph visualization (Cytoscape)
- Oracle Advanced Analytics (R, Mining)
- Oracle Business Intelligence (OBIEE)
- Map (GIS) Visualization
Oracle Spatial & Graph 12c

RDF Semantic Graph
Leverages Oracle Manageability:
• RAC & Exadata scalability
• Compression & partitioning
• SQL*Loader direct path load
• Parallel load, inference, query
• High Availability
• Triple-level label security
  • Ladder based inference
• Choice of SPARQL, SQL, or Java
• Native inference engine
• Enterprise Manager

Load / Storage
• Native RDF graph data store
• Manages 10s of billions of triples
• Relational to RDF mapping

Query
• SPARQL-Jena/Joseki, Sesame
• SQL/graph query, B-tree indexing
• Ontology assisted SQL query

Reasoning
• RDFS, OWL2 RL, EL+, SKOS
• User-defined rules
• Incremental, parallel reasoning
• User-defined inferencing
• Plug-in architecture

Analytics
• Semantic indexing framework
• OBIEE
• Oracle Advanced Analytics
• SPARQL Property Paths
World’s Fastest Big Data Graph Benchmark
1 Trillion Triple RDF Benchmark with Oracle Spatial and Graph

- World’s fastest data loading performance
- World’s fastest query performance
- Worlds fastest inference performance
- Massive scalability: 1.08 trillion edges
- Platform: Oracle Exadata X4-2 Database Machine
- Source: w3.org/wiki/LargeTripleStores, 9/26/2014
Trillion Triple Benchmark Set-up

**Exadata X4-2** High capacity full rack
ZS3-2 with 2 controllers, 8 trays of disk
8 database nodes (total 192 cores)
14 storage nodes (total 168 cores)
Oracle 12.1.0.1 DB standard install of Exadata

Open cursors = 1000
Processes = 1000
SGA = 132GB, PGA = 100GB
32K blocksize for tablespaces
A mix of DOP used: 296, 256, 192
TEMP group created with 3 bigfile tablespaces

* Test performed in Aug/Sept 2014

<table>
<thead>
<tr>
<th>Degrees of Parallelism</th>
<th>Data set</th>
<th>Load (B triples/hr)</th>
<th>OWL Inference (B triples/hr)</th>
<th>Query (B answers/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>256*</td>
<td>LUBM 4400K</td>
<td>605.4B / 115.2hrs</td>
<td>475.6+ B / 86hrs 30m</td>
<td>92.5B / 22.5 hrs</td>
</tr>
</tbody>
</table>
How is Semantic Data Stored in Oracle Database?

Semantic (RDF) data is mainly stored in two tables in MDSYS schema

**RDF_Value$ Table (compressed)**
- Stores the canonical & user value of each S, P and O
- Globally unique Hash id for each data element
- Reconciles non-differences like numbers 25 & 0025
- Canonical values enable more accurate queries & joins

**RDF_Link$ Table**
- Stores the triples (quads) using the hashIds in Value$
- Partitioned w/ local indexing to accelerate loading, querying and inferencing
## Manageability of RDF Semantic Graph
Integration with Oracle Database 11g/12c utilities and tools

<table>
<thead>
<tr>
<th>Ingest / Replicate / Recover</th>
<th>Tune / Analyze</th>
<th>Manage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bulk load:</strong>&lt;br&gt;• Apache Jena bulk loader&lt;br&gt;• Oracle external tables &amp;&lt;br&gt;• SQL*Loader (Direct Path) w/ PL/SQL Bulk Load API</td>
<td><strong>Tune load/ query/ inference:</strong>&lt;br&gt;• Parallelism&lt;br&gt;• Btree indexing triple/quad&lt;br&gt;• Typed literals indexing&lt;br&gt;• SPARQL query hints&lt;br&gt;• Statistics gathering&lt;br&gt;• Dynamic Sampling</td>
<td><strong>Control query execution:</strong>&lt;br&gt;• in database &amp; Jena client&lt;br&gt;<strong>Create &amp; monitor graph:</strong>&lt;br&gt;• Semantic Network&lt;br&gt;• Models, virtual models&lt;br&gt;• Btree indexes&lt;br&gt;• Rule bases&lt;br&gt;• Entailments&lt;br&gt;• Security data labels&lt;br&gt;• Semantic index policies</td>
</tr>
<tr>
<td><strong>Replicate &amp; recover:</strong>&lt;br&gt;• Data Guard: physical standby&lt;br&gt;• Data Pump: staging tables&lt;br&gt;• Recovery Manager: RMAN</td>
<td><strong>Analyze performance:</strong>&lt;br&gt;• Enterprise Manager: view optimizer plans, monitor execution / resource usage</td>
<td></td>
</tr>
</tbody>
</table>
How Does Semantics Address Finding Patterns in Data?

- Has its own graph query language - W3C SPARQL
- It’s a simpler way to write query patterns that need to be joined together
  e.g., Find pairs of siblings (same parents)

<table>
<thead>
<tr>
<th>SPARQL</th>
<th>SQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELECT ?x ?y FROM &lt;rdf_graph&gt; WHERE {</td>
<td>SELECT g1.subject x, g3.subject y FROM rdf_graph g1, rdf_graph g2, rdf_graph g3, rdf_graph g4 WHERE g1.predicate = 'hasFather' AND g2.predicate = 'hasMother' AND g3.predicate = 'hasFather' AND g4.predicate = 'hasMother' AND g1.subject = g2.subject AND g3.subject = g4.subject AND g1.object = g3.object AND g2.object = g4.object AND g1.subject != g3.subject</td>
</tr>
<tr>
<td>?x hasFather ?f .</td>
<td></td>
</tr>
<tr>
<td>?x hasMother ?m .</td>
<td></td>
</tr>
<tr>
<td>?y hasFather ?f .</td>
<td></td>
</tr>
<tr>
<td>?y hasMother ?m .</td>
<td></td>
</tr>
<tr>
<td>FILTER( ?x != ?y )</td>
<td></td>
</tr>
</tbody>
</table>
Does this mean I can’t use SQL?

• Not at all, Oracle SQL extended to include graph queries
• SEM_MATCH table function is based on SPARQL
• Uses patented SQL table function rewrite
  • Converts graph query to SQL, the whole SQL query is optimized
  • Returns a whole set of results rather than one result per table function call
SEM_MATCH: Adding SPARQL to SQL

Query example: “List everyone’s name and anyone else they know”
(Analogous to an outer join)

```sql
SELECT n1, n2
FROM TABLE(
    SEM_MATCH(
        'PREFIX foaf: <http://...>
        SELECT ?n1 ?n2
        FROM <http://g1>
        WHERE {?p foaf:name ?n1
            OPTIONAL {?p foaf:knows ?f .
                ?f foaf:name ?n2 }
        }
        ORDER BY ?n1 ?n2',
    SEM_MODELS('M1'),...));
```

<table>
<thead>
<tr>
<th>n1</th>
<th>n2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>Jerry</td>
</tr>
<tr>
<td>Alex</td>
<td>Tom</td>
</tr>
<tr>
<td>Alice</td>
<td>Jerry</td>
</tr>
<tr>
<td>Alice</td>
<td>Tom</td>
</tr>
<tr>
<td>Alice</td>
<td>John</td>
</tr>
<tr>
<td>Jerry</td>
<td>Tom</td>
</tr>
<tr>
<td>Tom</td>
<td>Jerry</td>
</tr>
<tr>
<td>Zack</td>
<td>-</td>
</tr>
</tbody>
</table>
SPARQL and “SPARQL in SQL” Architecture

HTTP

Standard SPARQL Endpoint
Enhanced with query management control

Java

Jena API
Jena Adapter

SQL

SEM_MATCH
rewritable table function

ORACLE® DATABASE

SPARQL-to-SQL Translation Logic
### SPARQL 1.1 Query Support

- 40+ new query functions/operators: IF, COALESCE, STRBEFORE, REPLACE, ABS,
- Aggregates: COUNT, SUM, MIN, MAX, AVG, GROUP_CONCAT, SAMPLE
- Subqueries
- Value Assignment: BIND, GROUP BY Expressions, SELECT Expressions
- Negation: NOT EXISTS, MINUS

#### On the fly inference: Transitivity of rdf:subClassOf

```
SELECT ?c
WHERE {
  ?x rdf:type ?sc
  ?sc rdfs:subClassOf * ?c }
```

#### Social Networking: Find all of John’s friends

```
SELECT ?name
WHERE {
  ?x foaf:name “John”
  ?x (foaf:knows | foaff:friendsOf)* ?f
  ?f foaf:name ?name }
```
SPARQL 1.1 Update

Capabilities of SPARQL Update

- Insert triples into an RDF Graph
- Delete triples from an RDF Graph
- Load an RDF Graph
- Clear an RDF Graph
- Create a new RDF Graph
- Drop an RDF Graph
- Copy, move or add the content of one RDF Graph to another
- Perform a group of update operations as a single action
SPARQL 1.1 Property Paths

A property path is a possible route through a graph between two graph nodes.

W3C Property Paths

- Answers question, “Yes or No: does a path exist from Tim to Sam?”
- Extends matching of a triple pattern to any length path
- A more succinct way to write parts of basic graph patterns
RDB2RDF: Modeling Relational Data as a Graph

Relational to RDF Modeling

- W3C Standard Specification
- Oracle Spatial and Graph 12c can represent relational schema as graph view
- Integrate content from distributed sources
- Federate distributed databases
- Apply SPARQL queries on tables, views, SQL query results
- No duplication of data and storage
How Does Semantics Enable Better SQL Results?

Query: “Find all entries in patient diagnosis column related to ‘Upper_Extremity_Fracture’”

Patients diagnosis table

<table>
<thead>
<tr>
<th>P_ID</th>
<th>DIAGNOSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hand_Fracture</td>
</tr>
<tr>
<td>2</td>
<td>Rheumatoid_Arthritis</td>
</tr>
<tr>
<td>3</td>
<td>Finger_Fracture</td>
</tr>
</tbody>
</table>

No results w/ the usual SELECT

```
SELECT p_id, diagnosis
FROM Patients
WHERE diagnosis = 'Upper_Extremity_Fracture';
```
How Does Semantics Enable Better SQL Results?

- Ontology-Assisted SQL Query

National Cancer Institute (NCI) medical ontology

Patients diagnosis table

<table>
<thead>
<tr>
<th>P_ID</th>
<th>DIAGNOSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hand_Fracture</td>
</tr>
<tr>
<td>2</td>
<td>Rheumatoid_Arthritis</td>
</tr>
<tr>
<td>3</td>
<td>Finger_Fracture</td>
</tr>
</tbody>
</table>

```
SELECT p_id, diagnosis
FROM Patients
WHERE diagnosis = "Upper_Extremity_Fracture"
```
How Does Semantics Address Better SQL Results

SEM_RELATED SQL operator expands the SQL WHERE clause with related terms from the Ontology

**Results:** Hand_Fracture, Finger Fracture

```
SELECT p_id, diagnosis
FROM Patients
WHERE SEM_RELATED (diagnosis,
'rdf:subClassOf',
':Upper_Extremity_Fracture',
'Medical_ontology' = 1)
```

### Patients diagnosis table

<table>
<thead>
<tr>
<th>P_ID</th>
<th>DIAGNOSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hand_Fracture</td>
</tr>
<tr>
<td>2</td>
<td>Rheumatoid_Arthritis</td>
</tr>
<tr>
<td>3</td>
<td>Finger_Fracture</td>
</tr>
</tbody>
</table>

NCI medical ontology:
Finger_Fracture ➔
Hand_Fracture ➔ Arm_Fracture ➔ Upper_Extremity_Fracture
Oracle Label Security Data Classification

• Fine grained security through integration with Oracle Label Security
• Oracle Label Security - mandatory access control
  • Labels assigned to both users and data
  • Data labels determine the sensitivity of the rows or the rights a person must posses in order to read or write the data.
  • User labels indicate their access rights to the data records.
• Model level security through GRANT/REVOKE privileges
How Does Semantics Address Discovery: Inferencing

• Business Constraint: An applicant can have only 1 line of credit
• Finds implicit (unstated) relationships in the data
• Allows machine-driven discovery based on W3C standard

Data:
Bank Data: John T Smith
  hasSocialSecNum ‘123’
Credit bureau report: JT Smith
  hasSocialSecNum ‘123’
Credit bureau report: JT Smith
  hasHomeEquityLoan ’789’

Machine adds the inference:
  John T Smith sameAs JT Smith

User-defined rule:
  if PersonA hasSocialSecNum SS1
    and PersonB hasSocialSecNum SS1
    then PersonA sameAs PersonB

Using OWL Construct:
  hasSocialSecNum rdf:type
    owl:InverseFunctionalProperty

Therefore: John T Smith
  hasHomeEquityLoan ’789’ and isn’t
  qualified for another line of credit
Core Inferencing Features

- Forward-chaining based inference engine in the database
- Native rulebases: RDFS, OWL 2 RL, OWL 2 EL, SKOS
- Validation of inferred data
- Proof generation
- User defined inferencing
  - Temporal reasoning, Spatial reasoning
- Ladder Based Inference
  - Fine grained security for inference graph
- Integration with external OWL 2 reasoners (e.g., TrOWL, Pellet)
Vocabulary Support in Oracle Spatial and Graph 12c

- W3C Simple Knowledge Organization System (SKOS)
  - New rulebase supporting the emerging SKOS standard on RDF
  - Enforces integrity constraints
- Dublin Core
- SNOWMED
- NCI (National Cancer Institute, Gene Ontology)
- FOAF (Friend of a Friend)
- GeoRSS
- SIOC (Semantically-Interlinked Online Communities)
- GoodRelations (eCommerce Product Ontology)
What about Java Programmers?

• Support for Apache Jena (Adapter) provides the following features:
  • A set of easy-to-use and performant Java APIs to access Oracle database
  • SPARQL 1.1, Named graphs, web service endpoint, REST APIs, JSON output
    • SPARQL Protocol, Federated SPARQL, SPARQL update
  • Various data loading (RDF/XML, N-TRIPLES, Turtle, …) with long literal support
  • Oracle-specific extensions for query execution control and management
    • Timeout, abort, S2S, hints in SPARQL syntax, property path, result cache, mid-tier cache, user-defined functions…
  • Integration with open source and commercial tools and reasoners
Can I Apply Semantics to Text Data - Semantic Indexing

Auto maintained like a B-tree index
CREATE INDEX ArticleIndex
ON Newsfeed (Article)
INDEXTYPE IS SemContext
PARAMETERS ('my_policy')

Copy
Batch extractor
SemContext index on Article column
SELECT Sem_Contains_Select(1)
FROM Newsfeed
WHERE Sem_Contains (Article,
  '{?x rdf:type rc:Person .
   ?x :hasAge ?age .
   FILTER(?age >= 35)}',1)=1
AND Source = 'CNN'

Analytical Queries
On Graph Data

Triples table with rowid references
<table>
<thead>
<tr>
<th>Subject</th>
<th>Property</th>
<th>Object</th>
<th>graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>p:Marcus</td>
<td>rdf:type</td>
<td>rc::Person</td>
<td>&lt;…/r1&gt;</td>
</tr>
<tr>
<td>p:Marcus</td>
<td>:hasName</td>
<td>&quot;Marcus&quot;</td>
<td>&lt;…/r1&gt;</td>
</tr>
<tr>
<td>p:Marcus</td>
<td>:hasAge</td>
<td>&quot;38&quot;^xsd:...</td>
<td>&lt;…/r1&gt;</td>
</tr>
</tbody>
</table>

Newsfeed table
<table>
<thead>
<tr>
<th>Rowid</th>
<th>docId</th>
<th>Article</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td>Indiana authorities filed felony charges and a court issued an arrest warrant for a financial manager who apparently tried to fake his death by crashing his airplane in a Florida swamp. Marcus Schrenker, 38…</td>
<td>CNN</td>
</tr>
<tr>
<td>r2</td>
<td>2</td>
<td>Major dealers and investors …</td>
<td>NW</td>
</tr>
</tbody>
</table>

Content type:
- Text
- File (path)
- URL
RDF Semantic Graph: Graph Visualization & Modeling Support

Graph Visualization

Cytoscape

Semantic Modeling

Protégé
Reporting RDF Data with Oracle BI EE

• Powerful BI dashboards
  – Visually appealing
  – 100% thin client

• Across all styles of analysis
  – R-OLAP, M-OLAP, Scorecards, Reporting, Collaboration, Actions

• Across all data sources
  – Federated data access
  – Share, collaborate, & publish
## Problem Classification

<table>
<thead>
<tr>
<th>Problem</th>
<th>Sample Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomaly Detection</td>
<td>Given demographic data about a set of customers, identify customer purchasing behavior that is significantly different from the norm</td>
</tr>
<tr>
<td>Association Rules</td>
<td>Find the items that tend to be purchased together and specify their relationship – market basket analysis</td>
</tr>
<tr>
<td>Clustering</td>
<td>Segment demographic data into clusters and rank the probability that an individual will belong to a given cluster</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>Given demographic data about a set of customers, group the attributes into general characteristics of the customers</td>
</tr>
</tbody>
</table>
Performing Statistical Graph Analytics

Oracle R Enterprise

- Open source language
- Statistical computing and chart for graph data
- Produces publication quality plots
- Highly extensible with open source R packages
Oracle Partner Tools: (IO Informatics)
Oracle Partner Tools: Tom Sawyer Social Network Analysis
Summary

• Oracle brings enterprise-class RDF semantic graph data management
• Scalable, Secure, and High Performance: load, query, inference features
• Supports W3C Semantic Standards
• Works with structured and unstructured data
• Enterprise-class Oracle tools can now mine insight from semantic data
  – OBIEE
  – Oracle Data Mining
  – Oracle R Enterprise
• Graph DB now available on Oracle NoSQL Database EE
For More Information

Oracle RDF

RDF/Graph Product Management
William.Beauregard@oracle.com
Xavier.Lopez@oracle.com

oracle.com
Hardware and Software

Engineered to Work Together