Oracle Spatial and Graph
Overview of New Graph Features
"THE FOLLOWING IS INTENDED TO OUTLINE OUR GENERAL PRODUCT DIRECTION. IT IS INTENDED FOR INFORMATION PURPOSES ONLY, AND MAY NOT BE INCORPORATED INTO ANY CONTRACT. IT IS NOT A COMMITMENT TO DELIVER ANY MATERIAL, CODE, OR FUNCTIONALITY, AND SHOULD NOT BE RELIED UPON IN MAKING PURCHASING DECISION. THE DEVELOPMENT, RELEASE, AND TIMING OF ANY FEATURES OR FUNCTIONALITY DESCRIBED FOR ORACLE'S PRODUCTS REMAINS AT THE SOLE DISCRETION OF ORACLE."
Program Agenda

- Background: Graph Concepts
- Use Cases
- Newest Graph Features in Oracle Database 12c
- Performance, Scalability, Manageability
Oracle Spatial and Graph
Mature, Proven Graph Database Capabilities

Graph Features

- Network Data Model graph
- W3C RDF Semantic graph
Network Data Model:
Graph model to represent physical and logical networks

Key Features
- Explicitly stores and maintains connectivity
- Attributes at link and node level
- Java API to perform Analysis in memory
- Multiple Cost Support in Path Analysis
- Traveling salesman, spanning tree, shortest path, sub-path, within cost, nearest neighbors
GRAPH CONCEPTS
APPLIED TO THE ENTERPRISE
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
Modeling Entity Relationships as graphs

- The basic unit of information (fact) is represented as <subject, predicate, object> “triple”
- Originally created to encode metadata such as ‘author’, ‘date’, etc. for web resources.
- Recently, it has become popular to relate things in the real-world such as people, places, concepts etc.
- Triples together form a “graph”, connecting pieces of data
Relational to Graph Modeling

Compartmental Relationships:
- Neuron
- Pathological Change
- Agent
- Drug
- Neuronal Property
- Pathological Agent
- Channel
- Compartment

Graph Modeling Table:

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Cell: Neuron/DB</th>
<th>Receptor</th>
<th>Channel</th>
<th>Pathological Agent (PA)</th>
<th>PA Action</th>
<th>Drug</th>
<th>Drug Action</th>
<th>Stage</th>
<th>Note</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soma</td>
<td>CA1 pyramidal neuron</td>
<td>GabaA</td>
<td>L A</td>
<td>beta Amyloid</td>
<td>Inhibits</td>
<td></td>
<td></td>
<td>Early</td>
<td>View</td>
<td>66240</td>
</tr>
<tr>
<td>Dendrite</td>
<td>CA1 pyramidal neuron</td>
<td>GabaA</td>
<td>L A</td>
<td>beta Amyloid</td>
<td>Inhibits</td>
<td></td>
<td></td>
<td>Early</td>
<td>View</td>
<td>66240</td>
</tr>
<tr>
<td>Unspecified</td>
<td>Oocyte</td>
<td>GabaA</td>
<td>L L high threshold</td>
<td>beta Amyloid</td>
<td>Inhibits</td>
<td></td>
<td></td>
<td>Early</td>
<td>View</td>
<td>66252</td>
</tr>
<tr>
<td></td>
<td>CA1 pyramidal neuron</td>
<td>NMDA</td>
<td>L Calcium</td>
<td>beta Amyloid</td>
<td>Inhibits</td>
<td></td>
<td></td>
<td>Early</td>
<td>View</td>
<td>66255</td>
</tr>
</tbody>
</table>

Courtesy: SenseLab, Yale University
Enterprise Metadata Modeling

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
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

Copyright © 2013, Oracle and/or its affiliates. All rights reserved.
Benefits of RDF Graph Metadata Management

- Open, flexible, unifying enterprise metadata model
- Incorporates ALL business content (RDBMS, ECM, Big Data, events)
- Highly expressive rules for managing data policies and governance
- Enables users to link, analyze, and discover information horizontally, as shared services, across a variety of enterprise and Web sources
USE CASES
Industries Have Already Adopted the Concept

- Life Sciences
- Finance
- Media
- Networks & Communications
- Defense & Intelligence
- Public Sector
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
Allied Nation Intelligence Service

Oracle Spatial and Graph: Social Analysis

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
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
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,....
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 output
- Graph visualization (Cytoscape)
- Map (GIS) Visualization
- Oracle Business Intelligence (OBIEE)
- Oracle BPM

Data Sources
- Transaction Systems
- Unstructured Content
- RSS, email
- Other Data Formats

Oracle Database 12c Spatial and Graph Tooling
Oracle Database 12c RDF Triple Store

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

- Native RDF graph data store
- Manages tens of billions of triples
- Relational to RDF mapping

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

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

- Semantic indexing framework
- Integration with
  • OBIEE, Oracle R Enterprise
  • Oracle Data Mining
SPARQL and “SPARQL in SQL” Architecture

HTTP

Java

Standard SPARQL Endpoint
Enhanced with query management control

Jena API
Jena Adapter

Sesame API
Sesame Adapter

SQL

SEM_MATCH
rewritable table function

SPARQL-to-SQL
Translation Logic
NEWEST GRAPH FEATURES IN ORACLE SPATIAL AND GRAPH FOR ORACLE DATABASE 12c
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

<table>
<thead>
<tr>
<th>On the fly inference: transitivity of rdfs:subClassOf</th>
<th>Social Networking: find all of John’s friends</th>
</tr>
</thead>
</table>
| SELECT ?c  
WHERE {  
?x rdf:type ?sc .  
?sc rdfs:subClassOf* ?c } | SELECT ?c  
WHERE {  
?x foaf:name “John” .  
?f (foaf:knows|foaf:friendOf)+ ?f .  
?f foaf:name ?name } |
RDB2RDF: Modeling Relational Data as a Graph

- 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
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.

- 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
Oracle Spatial and Graph Inference

- Native OWL 2 EL inference support
- User defined inferencing
  - Allows generation of new RDF resources
  - Temporal reasoning, Spatial reasoning
  - Web service callouts
- Ladder Based Inference
  - Fine grained security for inference graph
- Performance optimization for user defined rules
- Integration with TrOWL*, an external OWL 2 reasoner
  - TrOWL is a transformation based, tractable reasoner for OWL 2

* http://trowl.eu/
Jena and Sesame Adapters

Preconfigured, low query cost, ease of install & management

- Easy-to-use Java APIs to access Oracle database
- A standard-compliant SPARQL web service endpoint
- Data loading (RDF/XML, N-TRIPLES, N-QUADS, TriG, Turtle) w/ long literals
- JSON output
- Oracle-specific extensions for query execution control and management
- Integration with OBIEE, RDF browser
Graph Support on Oracle NoSQL
Available on Oracle NoSQL Database (Enterprise Edition)

Graph Feature for NoSQL

- RDF Graph support in Oracle NoSQL Database Enterprise Edition
- High performance Key Value store
- Standard access to graph data: SPARQL 1.1
- Jena & Joseki SPARQL endpoint Web Services
- Massive horizontal scalability – petabytes of triples
- Support for World Wide Web Consortium (W3C) Semantic Web standards
### 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>
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
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
PERFORMANCE, SCALABILITY, MANAGEABILITY
Performance and Scalability

• Scales to 100s of billions of triples (petabytes) and more
  - Scales linearly with Oracle database and hardware
  - No limitations as with other in-memory approaches

• Fast loading of triples
  – Incremental and bulk loading

• Parallelism is exploited
  – Load, Query, Inference

• Comparable to or faster than competing RDF graph databases
Manageability of RDF Semantic Graph
Integration with Oracle Database 11g/12c utilities and tools

**Ingest / Replicate / Recover**
- **Bulk load:**
  - Apache Jena bulk loader
  - Oracle external tables & SQL*Loader (Direct Path) w/ PL/SQL Bulk Load API
- **Replicate & recover:**
  - Data Guard: physical standby
  - Data Pump: staging tables
  - Recovery Manager: RMAN

**Tune / Analyze**
- **Tune load/ query/ inference:**
  - Parallelism
  - Btree indexing triple/quad
  - Typed literals indexing
  - SPARQL query hints
  - Statistics gathering
  - Dynamic Sampling

**Analyze performance:**
- Enterprise Manager: view optimizer plans, monitor execution / resource usage

**Manage**
- **Control query execution:**
  - in database & Jena client
- **Create & monitor graph w/ SQL Developer:**
  - Semantic Network
  - Models, virtual models
  - Btree indexes
  - Rule bases
  - Entailments
  - Security data labels
  - Semantic index policies
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
Hardware and Software
Engineered to Work Together