Many Graphs for Many Uses:
An Introduction to RDF, Property, and Network Graphs in Oracle Database 12c

Xavier Lopez, Senior Director, Oracle
Zhe Wu, Architect
Safe Harbor Statement

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 decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Program Agenda

1. Oracle Spatial and Graph (RDF Graph)
2. Oracle Spatial and Graph (Spatial Graph)
3. Big Data Spatial and Graph (Property Graph)
Linked Data
RDF Graph Database
Graph Data Models

Property Graph Model
- Graph Data Management
- Social Network Analysis
- Entity analytics

Social Network Analysis

Network Data Model
- Network path analysis
- Transportation modeling

Spatial Network Analysis

RDF Data Model
- Data federation
- Knowledge representation
- Semantic Web

Linked Data / Metadata Layer

Use Case

Graph Model

Industry Domain

- Public Safety
- Social Media search
- Marketing - Sentiment

- Logistics
- Transportation
- Utilities
- Telcos

- Life Sciences
- Health Care
- Publishing
- Finance

- Life Sciences
- Health Care
- Publishing
- Finance

- Logistics
- Transportation
- Utilities
- Telcos

- Public Safety
- Social Media search
- Marketing - Sentiment
RDF Graph: A Purpose-built Graph Model

Linked Data

• Unified metadata model for distributed data sources
• Flexible metadata product vocabularies and ontologies
• Validates semantic and structural consistency
Graph-based Metadata Layer

- W3C standard, flexible model for sparse and evolving data
- Common vocabulary enables data integration & app development
- Relational data stays in place, apps don’t need to change
RDF graph is an *enterprise metadata framework*. The metadata graph associates underlying instance data to other data resources based on their semantics. This linking of resources enables interoperability between apps that exchange data.
Enterprise Information Harmonization

Industries

• Life Sciences
• Health Care
• Finance
• Media
• Networks & Communications
• Defense & Intelligence
Consolidated Knowledge Layer

Business Challenge
• 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
• Semantic integration layer using RDF graph.
• Rich domain-specific terminology (biology, chemistry, and medicine) 1.6 M terms.
• Terminology Hub: 8 GB of referential data (ontologies) that cross-reference various data repositories.
EU Publications Office
Linked Metadata Platform for European Union

**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
National Intelligence Agency

Information Extraction
Feature Extraction, Term Extraction

Extracted Entities & Relationships

Intelligence Ontologies

Search, Presentation, Report, Visualization, Query

SQL/SPARQL

National Intelligence Scenario

Data Sources
Contents Repository
Databases
Web resources
Blogs, Mails, news, RSS feeds

Enterprise Data
Spatial images Documents
Oracle Database 12c RDF Semantic Graph Database

• Exadata ready
• Compression & partitioning
• Parallel load, inference, query
• High availability
• Label security: triple-level
• W3C standards compliance
• Semantic Indexing of text
• Enterprise Manager

Load / Storage
• Native RDF graph data store
• Manages trillions of triples
• Optimized storage architecture

Query
• SPARQL-Jena/Joseki/Fuseki
• 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
• Integration with
  • OBIEE, Oracle R Enterprise
  • Oracle Data Mining
RDB to RDF Mapping

- RDF views on relational tables
- Enables SPARQL query on distributed resources
- Views: Automatic and custom
- Aligns with W3C RDB2RDF standard
- No duplication of data and storage
RDF Semantic Graph: Graphical Tools

Graph Visualization

- Cytoscape

Ontology Modeling

- Protégé

RDF Studio

- SQL Developer
Core Inferencing Features

- Forward-chaining based inference engine in the database
  -- Removes on-the-fly reasoning and results in fast query times
- Native rulebases: RDFS, OWL 2 RL, OWL 2 EL, SKOS
  - SNOMED (subset of OWL 2 EL)
- 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 (TrOWL)
Manageability of RDF Semantic Graph
Built in support from Oracle Database 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

**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
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

Oracle Database 12c can load, query and inference millions of RDF graph edges per second

Millions of triples per second

<table>
<thead>
<tr>
<th></th>
<th>Query</th>
<th>Load</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>1.13</td>
<td>1.42</td>
<td>1.52</td>
</tr>
</tbody>
</table>
What Sets Oracle RDF Triple Database Apart?

• **Scalability**: Trillions of triples
• **Transactional**: Concurrent loading and updates with ACID properties
• **Security**: OLS security labels at “triple” level (OLS).
• **Standards based**: W3C, SPARQL, RDF, OWL, REST, JSON
• **Manageable**: Use existing DB tools, utilities and expertise
• **Multi-type support**: graph, relational, geospatial ...
Upcoming RDF Graph Features

• SPARQL 1.1 completion and performance
• Major enhancements in Jena adapter
• Performance improvements in querying RDF views of relational data
• Improved visualization and management of RDF using graphical tools
Graph Data Models

Social Network Analysis

Property Graph Model
- Graph Data Management
- Social Network Analysis
- Entity analytics

Network Data Model
- Network path analysis
- Transportation modeling

Spatial Network Analysis

Linked Data / Metadata Layer

RDF Data Model
- Data federation
- Knowledge representation
- Semantic Web

Use Case   Graph Model   Industry Domain

- Public Safety
- Social Media search
- Marketing - Sentiment
- Logistics
- Transportation
- Utilities
- Telcos
- Life Sciences
- Health Care
- Publishing
- Finance

Copyright © 2015 Oracle and/or its affiliates. All rights reserved.
Oracle Spatial Network Graph
Oracle Spatial and Graph

Complete
Open
Integrated
Most Widely Used
Graph Features – Network Data Model Graph

- A storage model to represent graphs and networks
- Graph tables consist of links and nodes
- Explicitly stores and maintains connectivity of the network graph
- Attributes at link and node level
- Logical or spatial graphs
- Directed and undirected graphs with or without cost
- Can logically partition the network graph

- Java API to perform Analysis in memory
- Loads and retains only the partitions needed
- Dynamic costs
- Multi-level / priority routing
- Shortest path, within cost, nearest neighbors
- Traveling salesman, spanning tree, ...
- Multiple Cost Support in Path/Subpath Analysis
Network Data Model

A purpose-built graph model

- Road and Multimodal Networks
- Drive Time Polygon Analysis
- Trade Area Management
- Service Delivery Optimization
- Water, Gas, Electric Utility, Network Applications
COTRAL, larger Rome region, Italy

**City Service**
Proactive info on route changes, optimal route planning

**City Operation**
Geographical Motion Monitoring and Management with MapViewer

**City Infrastructure**
Oracle Spatial Database stores real time bus info

**Sustainable City**
Bus position, speed, engine status, # passengers, fuel...

Illustrations © COTRAL
Istanbul Municipality

• Diverse Transport Modes
  – Buses
  – Tramways
  – Metro
  – Trains

• But also
  – Ferries
  – Private mini-buses
  – Shared taxis
Garmin

**Connect** Fitness Data Portal

**Objectives**
- Match user’s fitness activities to popular routes
- Create leader boards for popular routes

**Solution**
- Stores and simplifies processing of more than 5 billion miles of user activities
- Utilizes parallel processing, DB partitioning and pipelined table functions to analyze the data on Oracle Exadata
- Matches user’s activity to a segment using LRS
Consistent Vision

• Use the unique qualities of Oracle technology to create a Geospatial Data Platform to address challenges being faced by “GIS” systems & users

• Make geospatial information available to every operational system, every application, every analysis tool in the way that is appropriate to the developer or the application

• So users get the most value from their investments
Oracle Big Data Spatial and Graph Property Graph
Graph Data Models

Property Graph Model
- Graph Data Management
- Social Network Analysis
- Entity analytics

Network Data Model
- Network path analysis
- Transportation modeling

RDF Data Model
- Data federation
- Knowledge representation
- Semantic Web

Social Network Analysis
- Public Safety
- Social Media search
- Marketing - Sentiment

Spatial Network Analysis
- Logistics
- Transportation
- Utilities
- Teloms

Linked Data / Metadata Layer
- Life Sciences
- Health Care
- Publishing
- Finance

Use Case
- Graph Model
- Industry Domain
The Property Graph Data Model

- A set of vertices (or nodes)
  - each vertex has a unique identifier.
  - each vertex has a set of in/out edges.
  - each vertex has a collection of key-value properties.

- A set of edges (or links)
  - each edge has a unique identifier.
  - each edge has a head/tail vertex.
  - each edge has a label denoting type of relationship between two vertices.
  - each edge has a collection of key-value properties.

https://github.com/tinkerpop/blueprints/wiki/Property-Graph-Model
Modeling and Analyzing The Internet of Things

• Cyber-Security
  Critical / Alternate Path

• Analysis
  Community Detection
  Network Monitoring
  Predictive Analysis

• Multiple System Impact Analysis
  Transportation
  Utilities
  Finance
Property Graph: Usage Scenarios

- Insurance fraud detection
  - Find parties in insurance data who are on both sides of multiple claims, who live near each other

- Software Code Analysis
  - Score the risk of individual change and determine need for review, probability of affecting schedule

- Border Control
  - Analyze flight histories of a suspicious passenger. Identify his co-travelers, co-traveler’s co-travelers, ...

- Electrical Grid
  - Determine the effect of an outage across network

- Network intrusion forensics
  - Find entry points and affected machines
Common Graph Analysis Use Cases

- **Product Recommendation**: Recommend the most similar item purchased by similar people.
- **Influencer Identification**: Find out people that are central in the given network – e.g. influencer marketing.
- **Community Detection**: Identify group of people that are close to each other – e.g. target group marketing.
- **Graph Pattern Matching**: Find out all the sets of entities that match to the given pattern – e.g. fraud detection.
Property Graph Workflow

• Graph Data Management
  – Raw business data is converted to a graph schema
  – In Database graph queries using SQL (useful for breath first search)

• Analysis and Exploration (in-memory analysis engine)
  – Data scientists try different ideas (algorithms) on the data
  – Flexible, interactive, iterative, small-scale (sampled), ....
Architecture of Property Graph Support

Graph Data Access Layer (DAL)
- Blueprints & Lucene/SolrCloud

Graph Analytics
- Parallel In-Memory Graph Analytics (PGX)

Scalable and Persistent Storage Management
- Apache HBase
- Oracle NoSQL Database

Java APIs

REST/Web Service
- Java, Groovy, Python, ...

Property Graph formats
- GraphML
- GML
- Graph-SON
- Flat Files
In-Memory Parallel Graph Analytics

**Graph Analyst**

- An in-memory, parallel framework for fast graph analytics
  - Read a graph from Oracle Database (through data access layer)
    - SQL parallel query support on PG data
  - Handles *analytic* workloads while the data access layer handles *transactional* workloads
  - Supports concurrent sessions and multiple users/graphs

---

**Data Storage (delta update)**

**Graph Analyst**

**Analytic Request**

**Transactional Request**

**SQL**
Two Kinds of Graph Workloads

**Computational Graph Analytics**

- **Connected Components**
- **Modularity**
- **Conductance**
- **Shortest Path**
- **Pagerank**
- **Clustering Coefficient**
- **Centrality**
- **Coloring**

Compute certain values on nodes and edges

While (repeatedly) **traversing** or **iterating** on the graph

In certain **procedural** ways

**Graph Pattern Matching**

Given a **description** of a pattern

Find every sub-graph that **matches** it

**In-Memory Analyst** supports **both** kinds, as well as **combinations of the two**
In-Memory Parallel Graph Analytics: APIs, Formats, Analytics, Search

• Optimized built-in graph analytics and search
  – 30+ parallel algorithm implementations
  – Clustering, ranking, path finding, recommendation, and more
  – Graph pattern search

• Client Shell

• Groovy integration

• Graph format (in addition to those supported by the data access layer)
  – EBin, Adjacency list, Edge List

• J2EE container support (WLS, Tomcat)
Computational Analytics: Built-in Package

Rich set of built-in parallel graph algorithms

Detecting Components and Communities
- Tarjan’s, Kosaraju’s, Weakly Connected Components, Label Propagation (w/ variants), Soman and Narang’s Spacification

Evaluating Community Structures
- Conductance, Modularity, Clustering Coefficient (Triangle Counting), Adamic-Adar

Ranking and Walking
- Pagerank, Personalized Pagerank, Betweenness Centrality (w/ variants), Closeness Centrality, Degree Centrality, Eigenvector Centrality, HITS, Random walking and sampling (w/ variants)

Path-Finding
- Hop-Distance (BFS), Dijkstra’s, Bi-directional Dijkstra’s, Bellman-Ford’s

Other Classics
- Vertex Cover, Minimum Spanning-Tree (Prim’s)

Link Prediction
- SALSA (Twitter’s Who-to-follow)

... and parallel graph mutation operations

The original graph
- Create Undirected Graph
- Create Bipartite Graph
- Sort-By-Degree (Renumbering)
- Filtered Subgraph
- Simplify Graph

Left Set: “a,b,e”
Text Search through Apache Lucene/Solr Cloud

- Integration with Apache Lucene/Solr
- Support manual and auto indexing of Graph elements
  - Manual index:
    - `oraclePropertyGraph.createIndex("my_index", Vertex.class);`
    - `indexVertices = oraclePropertyGraph.getIndex("my_index", Vertex.class);`
    - `indexVertices.put("key", "value", myVertex);`
  - Auto Index
    - `oraclePropertyGraph.createKeyIndex("name", Edge.class);`
    - `oraclePropertyGraph.getEdges("name", "*hello*world");`
- Enables queries to use syntax like "*oracle* or *graph*"
The Big Picture – Oracle Big Data Management System

DATA RESERVOIR
- Cloudera Hadoop
- Oracle Big Data SQL
- Oracle NoSQL
- Oracle R Distribution
- Oracle Big Data Spatial and Graph
- Big Data Appliance

DATA WAREHOUSE
- Oracle Database
- In-Memory, Multi-tenant
- Oracle Industry Models
- Oracle Advanced Analytics
- Oracle Spatial and Graph
- Exadata

SOURCES
- Oracle Event Processing
- Apache Flume
- Oracle GoldenGate

Oracle Big Data Connectors
Oracle Data Integrator

Oracle Data Integrator
Oracle GoldenGate
Oracle Event Processing
Property Graph Summary

• Complete platform:
  • Available on BDA, or off BDA + Text indexing/search + Built-in analytics + Open source Java APIs & scripting languages for developers + Groovy console

• Scalable:
  • Distributed database and text indexing/search; parallel in-memory analytics are concurrent & multiuser; filter queries refine graph data read into memory for analysis

• Fast: Parallel everywhere - load, query and in-memory analytics

• Flexible:
  • Deploy on-premise or in the Cloud, Hbase & NoSQL, text search w/ Lucene & SolrCloud, 3 ways to deploy in-memory analytics, extensible analytics architecture

• Open Source-based: Apache, Java, Tinkerpop APIs; Groovy, Python... scripting languages
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
- Hadoop
- NoSQL
- Oracle Cloud
Q&A

Oracle Spatial and Graph:
www.oracle.com/technetwork/database/options/spatialandgraph
www.oracle.com/technetwork/database/database-technologies/bigdata-spatialandgraph