Professionals at Work

First there are the IT professionals who write for the Journal. A very special mention goes to Brian Hitchcock, who has written dozens of book reviews over a 12-year period. The professional pictures on the front cover are supplied by Photos.com.

Next, the Journal is professionally copyedited and proofread by veteran copyeditor Karen Mead of Creative Solutions. Karen polishes phrasing and calls out misused words (such as “reminiscences” instead of “reminisces”). She dots every i, crosses every t, checks every quote, and verifies every URL.

Then, the Journal is expertly designed by graphics duo Kenneth Lockerbie and Richard Repas of San Francisco-based Giraffex.

And, finally, Jo Dziubek at Andover Printing Services deftly brings the Journal to life on an HP Indigo digital printer.

This is the 109th issue of the NoCOUG Journal. Enjoy! ▲

—NoCOUG Journal Editor

Table of Contents

<table>
<thead>
<tr>
<th>Interview .......................................................... 4</th>
<th>ADVERTISERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference Report ................................................ 7</td>
<td>Confio Software ........................................... 2</td>
</tr>
<tr>
<td>Book Excerpt ...................................................... 8</td>
<td>Delphix ......................................................... 2</td>
</tr>
<tr>
<td>Special Feature ..................................................... 15</td>
<td>Embarcadero Technologies ................................. 2</td>
</tr>
<tr>
<td>SQL Corner .......................................................... 22</td>
<td>Kaminario ......................................................... 2</td>
</tr>
<tr>
<td>Session Descriptions ............................................... 24</td>
<td>Quilogy Services ................................................ 21</td>
</tr>
<tr>
<td>Treasurer’s Report .................................................. 25</td>
<td>Database Specialists .......................................... 27</td>
</tr>
<tr>
<td>Conference Agenda .................................................. 28</td>
<td></td>
</tr>
</tbody>
</table>

Publication Notices and Submission Format

The NoCOUG Journal is published four times a year by the Northern California Oracle Users Group (NoCOUG) approximately two weeks prior to the quarterly educational conferences.

Please send your questions, feedback, and submissions to the NoCOUG Journal editor at journal@nocoug.org.

The submission deadline for each issue is eight weeks prior to the quarterly conference. Article submissions should be made in Microsoft Word format via email.

Copyright © by the Northern California Oracle Users Group except where otherwise indicated.

NoCOUG does not warrant the NoCOUG Journal to be error-free.
For the last ten years, Oracle Spatial and Graph has delivered the industry’s most advanced enterprise spatial and graph data management solution. Oracle Spatial and Graph is completely integrated with Oracle Database 12c, thereby providing developers with the enterprise-class performance, scalability, reliability, and security necessary for today's graph-based applications. It provides support for two graph data models, along with the industry's leading spatial data management.

**RDF Semantic Graph** supports the World Wide Web Consortium (W3C) Resource Description Framework (RDF) graph standard. This model supports the unique data management, querying, and inferencing commonly used in a variety of social network analysis and linked open data solutions.

**Network Data Model (NDM) Graph** is a property graph data model used to model and analyze physical and logical networks used in industries such as transportation, logistics, and utilities. NDM persists the network connectivity metadata in the database, while a Java API provides fast in-memory graph analytics, including shortest path, nearest neighbors, within cost, and reachability.

Spatial data management supports the complex requirements found in Geographic Information Systems (GISs), enterprise applications, and location-enabled business and web mapping applications. Capabilities include native support for spatial data models and types such as georaster (for geo-referenced imagery and gridded data); topology; 3D, including triangulated irregular networks (TINs) and point clouds (supporting LIDAR data); and linear referencing. In addition, a geocoding engine, a routing engine, and spatial web services conformant with Open Geospatial Consortium (OGC) and ISO standards are also included. These geospatial features provide a complete platform to deploy a range of enterprise GIS and web mapping solutions used in the public sector, utilities, telecommunications, retail, and logistics.

The following sections provide detailed overview of Oracle's two native graph capabilities: RDF Semantic Graph and the Network Data Model.

**RDF Semantic Graph Features Overview**

Graphs are becoming central to a new category of social network and linked data applications common in public sector, health sciences, finance, media and intelligence communities. The graph data structure is an ideal *metadata layer* to support the integration of various data sources.
Storing, Loading, and Data Manipulation

RDF is the W3C standard for semantic data. An RDF data element is a “resource.” A unique benefit of RDF graphs is that resources enable data integration between different and even disparate data sets. Integration is possible because each resource has a globally unique universal resource identifier (URI), like a Social Security number. Resources form simple “subject-predicate-object” statements. The predicate is a property of the subject and the type of relationship between the subject and the object. The object is either a value for the property (a literal) or the URI of another subject that connects triples to form a graph. Ultimately, RDF graphs comprise nodes (subjects and objects) connected by directed, labeled edges (predicates) as illustrated by the labeled arrow in the figure below.

Figure 4: An RDF triple has a directed, labeled edge.

RDF graphs are flexible. It is easy to expand and combine RDF graphs because business information about the data (metadata) is stored in the graph as relationships rather than as column headers in a table.

Relational data can easily be converted or mapped to an RDF graph.

The example above illustrates another benefit of RDF graphs; sparse data is stored flexibly and efficiently. Only existing relationships for a given subject are stored. Add new relationships by simply inserting a new triple for a particular subject.

The unit of storage in RDF Semantic Graph is a model. Models are user-defined and application-specific. In addition, virtual model capability provides a view-like feature to combine models for querying. RDF Semantic Graph leverages the standard Oracle Database 12c loading, storing, and data manipulation operations on RDF/OWL models. Space-efficient storage saves up to 60% disk space for scalable and performant loading, querying, and inferencing. It has proven scalability beyond tens of billions of triples (LUBM 200K benchmark) and can readily scale into the tens of petabytes of triples.

Querying RDF Graphs in Oracle Database 12c

RDF has its own graph query language, SPARQL. It is a simpler way to write query patterns that are joined together. For example, the following SPARQL query finds pairs of siblings, people with the same parents.

```
SELECT ?x ?y
FROM <rdf_graph>
WHERE
{
?x hasFather ?f .
?x hasMother ?m .
?y hasFather ?f .
?y hasMother ?m .
FILTER( ?x != ?y)
}
```

The same query in SQL is more complex:

```
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.object != g3.object
```

RDF Semantic Graph supports SPARQL 1.1, the latest W3C standard and a range of deployment approaches, including a web service endpoint and Java APIs. Unique to Oracle, SQL can query RDF graphs by embedding SPARQL queries in SQL using the SEM_MATCH table function.

Here is an example of a SPARQL query:

```
PREFIX foaf: <http://...>
SELECT ?n1 ?n2
FROM <http://g1>
WHERE
{
?p foaf:name ?n1
OPTIONAL
{
?f foaf:name ?n2
}
FILTER (REGEX(?n1, "^A"))
}
```

Here is an example of a SQL query with embedded SPARQL:

```
SELECT n1 n2
FROM TABLE(SEM_MATCH (```
```
{
?p foaf:name ?n1
OPTIONAL
{
?f foaf:name ?n2
}
FILTER (REGEX(?n1, "^A"))
}
SEM_MODELS( g1 )
...
SEM_ALIASES ( SEM_ALIAS ('foaf', 'http://...') )
)
```
```
The SEM_MATCH table function in Oracle Spatial and Graph supports SPARQL 1.1. Additionally, RDF Semantic Graph supports the OGC GeoSPARQL standard that allows spatial querying and filtering on location data stored in the graph.

**Viewing Relational Data as RDF Triples**

It is possible to apply RDF views on relational tables using Oracle Spatial and Graph. RDF views present relational data in RDF triple format. SPARQL queries on these views enable powerful data integration across federated data sources, a key requirement for Linked Data and enterprise integration applications. Development is simplified; graph pattern-matching relationship queries on relational tables do not require converting the tables to RDF triples.

RDF Semantic Graph supports three types of view definitions: a) **automatic mapping** (also referred to as Direct Mapping); b) **custom mapping**, using the W3C R2RML language; and c) **materialized views**. The simplest way to create a mapping of relational data to RDF data is by calling the SEM_APIS.CREATE_RDFVIEW_MODEL procedure to create an RDF view. It takes as input the list of tables or views you would like to view as RDF. This provides a direct mapping of those relational tables or views.

The following example illustrates Direct Mapping:

```sql
-- Use the following relational tables.
CREATE TABLE dept (  
deptno NUMBER CONSTRAINT pk_DeptTab_deptno PRIMARY KEY,  
dname VARCHAR2(30),  
loc VARCHAR2(30)  );
CREATE TABLE emp (  
empno NUMBER PRIMARY KEY,  
ename VARCHAR2(30),  
job VARCHAR2(20),  
deptno NUMBER REFERENCES dept (deptno)  );

-- Insert some data.
INSERT INTO dept (deptno, dname, loc)  
VALUES (1, 'Sales', 'Boston');
INSERT INTO dept (deptno, dname, loc)  
VALUES (2, 'Manufacturing', 'Chicago');
INSERT INTO dept (deptno, dname, loc)  
VALUES (3, 'Marketing', 'Boston');

INSERT INTO emp (empno, name, job, deptno)  
VALUES (1, 'Alvarez', 'SalesRep', 1);  
INSERT INTO emp (empno, name, job, deptno)  
VALUES (2, 'Baxter', 'Supervisor', 2);  
INSERT INTO emp (empno, name, job, deptno)  
VALUES (3, 'Chen', 'Writer', 3);  
INSERT INTO emp (empno, name, job, deptno)  
VALUES (4, 'Davis', 'Technician', 2);

-- Create an RDF view model using direct mapping of two tables, EMP and DEPT,  
-- with a base prefix of http://empdb/.  
-- Specify KEY_BASED_REFPROPERTY=T for the options parameter.
BEGIN  
  sem_api.create_rdfview_model(  
    model_name => 'empdb_model',  
    tables => SYS.ODCIVarchar2List('EMP', 'DEPT'),  
    prefix => 'http://empdb/',  
    options => 'KEY_BASED_REFPROPERTY=T'  );
END;
```

This produces the following output:

```sql
EMP
http://empdb/TESTUSER.EMP/EMPNO=1
http://empdb/TESTUSER.EMP/EMPNO=3
```

The preceding query is functionally comparable to this:

```sql
SELECT e.empno FROM emp e, dept d  
WHERE e.deptno = d.deptno AND d.loc = 'Boston';
```

Alternatively, an R2RML mapping document could customize the mapping between the relational tables and the RDF graph in the example above.

**Native Inferencing in Oracle Database 12c**

Inferencing allows machine-driven discovery of implicit facts from explicit RDF data. It gives application developers the advantage of deriving more application knowledge with no coding by using standards-defined rules and semantics. Inferencing adds new facts to the graph.

```
1) Graph Data
California
USA
NorthAmerica
```

```
2) Ontology
California::partOf::USA::partOf::NorthAmerica
```

```
3) Inference
California::partOf::USA::partOf::NorthAmerica
```

**Figure 6: Inferencing adds the fact that California is part of North America.**

RDF Semantic Graph provides native, forward-chaining, persistent inferencing using any combination of the built-in RDF, RDFS, and OWL 2 RL and EL profiles, as well as user-defined rules for specialized inference capabilities. Because inferencing is persistent and performed in advance of queries, it enhances query performance. Recent performance enhancements include optimized large owl:sameAs sets, incremental inference to update entailments after triple inserts, and parallel inference on multi-core or multi-CPU architectures.
For high-security applications, Oracle Spatial and Graph supports ladder-based inferencing. This feature applies appropriate security labels to newly inferred triples. RDF Semantic Graph supports additional reasoning requirements with a plug-in framework to integrate third-party special-purpose reasoners.

**XML, JSON and Relational Interoperability**

Oracle uniquely allows a SQL query to consult an ontology to provide more complete SQL results. Ontology-assisted SQL querying allows SQL queries to extract more semantically complete results from relational tables. By aligning relational data with a pre-defined ontology (or enterprise vocabulary) developers can better organize their instance data with a rich domain context.

For example, the following SQL query to find patients with upper extremity fractures on the table below returns no results:

```
SELECT p_id, diagnosis
FROM Patients
WHERE diagnosis = 'Upper_Extremity_Fracture';
```

However, the National Cancer Institute (NCI) medical ontology indicates that a hand fracture is a type of upper extremity fracture.

```
-- Ontology assisted SQL query
SELECT p_id, diagnosis
FROM Patients
WHERE SEM_RELATED (diagnosis, 'rdfs:subClassOf', 'Upper_Extremity_Fracture';) = 1
AND SEM_DISTANCE() <= 2;
```

To support integration with external business intelligence (BI) and mid-tier tools, Oracle Spatial and Graph also provides a SPARQL Gateway feature that presents SPARQL query results in XML format. This is essential for supporting enterprise software tools that ingest XML data sources, such as Oracle Business Intelligence (OBIEE). RDF Semantic Graph can return RDF query results in JSON interoperability format as well.

**Fine-Grained Security**

As mentioned earlier, certain applications require fine-grained security. Model-level access control is the default for RDF graph data. Oracle Label Security provides triple-level security for the most stringent security levels. It defines sensitivity labels on individual triples and users that conditionally restrict a user's access to individual triples stored in an RDF model. Oracle Spatial and Graph is unique in providing such a fine-grained security model.

**Graph Analytics**

To support social network analysis (SNA), Oracle Spatial and Graph supports SPARQL 1.1 property path expressions that find relationships within a large social graph. Oracle Advanced Analytics data mining and R statistical analysis can operate on graph query results. The Network Data Model provides in-memory graph analytics for shortest path, reachability, within cost, and nearest neighbor analysis on RDF data.

**SPARQL Property Path Expressions**

Support for SPARQL property path expressions is included. An expression determines whether there is any connectivity in the graph between two nodes in the graph, such as between Tim and Sam in the example below.

Oracle Advanced Analytics provides data mining and R statistical analysis capabilities for Oracle Database 12c. These features can also analyze the results of SPARQL pattern-matching queries on graphs in Oracle Database 12c.
Oracle R Enterprise

➤ Open source language and environment

➤ Statistical computing and graphics

➤ Easily produces publication-quality plots

➤ Highly extensible with open-source R packages

Semantic Indexing for Documents

Semantic indexing for documents goes beyond key word searching to find documents that have semantically related content. SPARQL graph pattern queries interrogate the semantic index to find relevant documents. For instance, referring to our earlier SPARQL example, one could find all documents that mention pairs of siblings. The results could include the names of the siblings and the document(s) that mention them, as well as other attributes.

Indexing occurs automatically by inserting documents, text, and URLs into a table that has a semantic index defined on it. An associated third-party natural language processing (NLP) engine or annotator analyzes the text and generates RDF triples or XML from the concepts and terms found in the text for the semantic index. Inferencing on the semantic index can discover new relationships and expand search results.

Enterprise Performance and Scalability

RDF Semantic Graph supports parallelism, compression, partitioning, Oracle Real Applications Clusters, and the Oracle Exadata Database Machine for enterprise-level performance, scalability, and reliability.

Partitioning Support for Spatial Indexes

Models subdivide a graph in Oracle Database 12c and reside in separate partitions. Partitioning offers significant performance, scalability, and manageability benefits, including the following:

➤ Reduce response times for long-running queries; partitioning can reduce disk I/O operations.

➤ Reduce response times for concurrent queries; I/O operations run concurrently on each partition.

➤ Maintain indexes more easily with partition-level operations to create and rebuild an index.

➤ Rebuild indexes on partitions without affecting the queries on other partitions.

➤ Change storage parameters for each local index independent of other partitions.

Parallel Support

Support for parallelism is an important differentiating feature of Oracle Spatial and Graph. Parallelism enhances the performance of graph data loading, queries, and inferencing operations. Parallelism also speeds index creation; it subdivides graph B-tree index creation into multiple tasks performed in parallel.

Unique Enterprise Manageability

RDF Semantic Graph supports Oracle Database 12c utilities and tools, including:

➤ Bulk loading, including the Jena Adapter bulk loader, Oracle external tables, and SQL*Loader Direct Path loading

➤ Replication and recovery, including Data Guard: physical standby, Data Pump staging tables, and Recovery Manager

➤ Tuning with parallelism, Btree indexing of triples and quads, typed literals indexing, SPARQL query hints, statistics gathering, and Dynamic Sampling

Open Standards

Oracle is a World Wide Web Consortium (W3C) member and active contributor and/or editor in various technical working groups.

RDF Semantic Graph supports the latest W3C specifications, including RDF, RDF Schema (RDFS), SPARQL 1.1 query language, OWL 2 (RL and EL profiles) knowledge representation languages for authoring ontologies, the Simple Knowledge Organization System (SKOS), and the RDB2RDF specifications for creating RDF views on relational tables—Direct Mapping (DM) and Mapping Language (R2RML). In addition, Oracle was instrumental in defining and supporting the Open Geospatial Consortium’s (OGC’s) GeoSPARQL 1.0 query specification.

Rich Ecosystem of Tools

Leading third-party commercial graph tools and applications include:

➤ Split, merge, and exchange partitions.
support RDF Semantic Graph. Products include ontology modeling and engineering, vocabulary alignment, and graph visualization tools. In addition, Oracle Business Intelligence and Oracle Advanced Analytics products can analyze RDF data. Finally, Oracle’s RDF graph database supports the leading open-source application development frameworks, including Apache Jena, Sesame, Protégé, and associated open-source tools as well as linked open data ontologies.

**Big Data and NoSQL Interoperability**

As graph data management becomes increasingly important to big data and NoSQL platforms, Oracle now supports RDF graphs in Oracle NoSQL Database. By incorporating W3C standards-based RDF support and open-source Apache Jena support (a Jena Adapter) into Oracle NoSQL Database, users benefit from high performance, low latency, and a no schema design ideal for low-latency retrieval and high-volume capture on simple data. This key-value store is ideally suited for exploiting the cost/performance benefits of horizontally scaled cloud-based and Linked Open Data architectures.

**Network Data Model (NDM) Graph Features Overview**

The NDM graph is a feature of Oracle Spatial and Graph that supports the network modeling and analysis requirements of leading telecommunications, electrical utilities, transportation, and logistics companies worldwide.

NDM stores general network (or property graph) data structures persistently in Oracle Database 12c. It explicitly stores and maintains network connectivity and provides network analysis capability, including shortest path, nearest neighbors, within cost, and reachability.

NDM has a PL/SQL API for managing network data in the database and a Java API for performing network analysis in mid-tier memory. The Java API is also instrumental in creating and applying network constraints.

To support analysis of very large graph models that exceed the platform’s memory limits, the NDM supports partitioning of large networks into manageable sub-networks that automatically load graph partitions into memory as needed for efficient in-memory analysis. This enables application developers to benefit from the speed of in-memory analytics on networks that are larger than available memory.

**Geocoding and Route Analysis**

NDM is central to support a range of web mapping, point-to-point routing, and geocoding services common in most geospatial applications. It uses the features of geocoding and the routing engine in Oracle Spatial and Graph. NDM supports commercial street network data from data providers Here (formerly Navteq) and TomTom in their respective Oracle formats.

**Modeling Capabilities**

- Model and represent any point along a link for all analysis functions, such as specific addresses in street networks with any number of properties on the nodes and links.
- Model partial-link paths (sub-paths).
- Customize link and node properties (e.g., costs).
- Perform path analysis with multiple link and node properties (e.g., distance/time/hops costs).
- Perform partitioning of logical networks (e.g., social and biochemical pathway networks) based on metrics appropriate to the application.
- Compute the shortest route connecting a given set of nodes (the traveling salesperson problem).
- Generate a polygon representing the region that is reachable from a given node with a specific cost. A typical application is the generation of drive time and drive distance polygons.
Generate the shortest path on a hierarchical network, where links are prioritized by property (e.g., highways, local roads), to support queries such as finding the route between two addresses that favors highways over local roads as much as possible.

Compute a buffer based on network cost; the buffer representation contains coverage and cost information.

Compute K shortest paths between two nodes.

**Real-World Feature Modeling and Analysis**

Oracle Spatial and Graph NDM simplifies feature editing and analysis by providing a feature analysis function that associates feature representations with network elements.

Feature modeling simplifies application development by associating real-world objects with network elements. For example, if a utility network application needs to find affected households when a substation experiences a power failure, it is necessary to associate the application features (substations, power lines, and transformers) with network elements (links and nodes). Feature modeling maintains these relationships through feature metadata, simplifying application development and maintenance.

**Network Modeling with Time; Multimodal Transportation Routing**

Most real-world networks have a time element. Travel times on road segments vary with the time of day. Utility networks experience different demand loads based on seasonal demand and the time of day. To support temporal analysis of such networks, Oracle Spatial and Graph supports the modeling of networks containing a time dimension. NDM supports queries such as finding the fastest travel route for a specified time of day. NDM supports modeling and analysis of multimodal transportation networks, and computing the fastest paths on multimodal transportation networks.

**Summary**

Graphs are central to a new category of social media and linked data applications emerging in the life sciences, finance, media, and public sectors. In addition, graphs have been a fundamental data model to support a range of enterprise applications that require the modeling and analysis of network models common in telecommunications, utilities, and transportation. The availability of two graph models in Oracle Spatial and Graph ensures that applications developers can choose the appropriate model for their application. In doing so, they can leverage the industry’s most scalable and secure platform for enterprise-scale graph applications.


Bill Beauregard is a senior principal product manager with Oracle Corporation.