

September 2012

Oracle Spatial and Graph: Checklist for RDF Semantic Graph Project Planning

Introduction

Oracle Spatial and Graph includes advanced features for spatial data and analysis and for social and semantic graph, and physical, logical, network applications. The graph features in Oracle Spatial and Graph includes RDF semantic graphs used in social networks and social interactions and in linking disparate data sets to address requirements from the research, health sciences, finance, media and intelligence communities. In addition, Network Data Model (NDM) graphs are used in traditional network applications in major transportation, telcos, utilities and energy organizations. These are proven, robust graph database technologies.

Oracle would like your RDF semantic graph application to be successful. This white paper has a checklist of items to consider when planning your project and discussing it with Oracle.

Planning an RDF Semantic Graph Project

There are four main categories of information about the project that are important to consider. The categories include the business requirements and benefits, project overview, application functionality, and design considerations. These areas are outlined for consideration in the checklist below.

Summary Checklist of Project Questions

1. Business requirements:

1. Business problem being addressed?
2. Desired business outcome?
3. Current approach to the problem, if any?
4. Problem with the current approach?
5. How is a RDF semantic graph approach expected to solve the problem?
6. What is the benefit: How will the business operation be easier, faster or better? Is there an estimate of time, money, or resources saved? Incremental revenue and/or discoveries expected?

2. Project overview

1. Project objectives?
2. Milestones and phases?
3. Resources (developers, implementers, maintainers)?
4. Training: W3C standards for RDF, SPARQL, RDF-S, OWL, SKOS, RDFa, ontology engineering (TopBraid Composer, Protege), RDF semantic graph features of Oracle Spatial and Graph?
5. Documentation requirements?

3. Functional requirements

1. Workflow?
2. Workload?
 1. concurrent and total users?
 2. data set size/growth?
 3. throughput and response time?
3. Solution features?

4. Design requirements

1. Programming languages and frameworks to be used?

2. Third party and Open Source requirements
 1. Is the Jena or Sesame open source application framework required? (Jena Adapter & Sesame Adapter for Oracle Database are available on the OTN software download page:
http://www.oracle.com/technology/tech/semantic_technologies/)
 2. Is open source or 3rd party software required for ontology engineering, natural language processing, visualization, etc.?
3. Technology stack
 1. Are there plans to use a distributed system, cloud service, or a single system?
 1. How will the data management, application logic and user interface layers be deployed (SOA, client/server, etc.)?
 2. What are the hardware, software, and network platform requirements?
4. Data requirements - integration
 1. Will source data (the authoritative data) come from more than one source?
 2. Is data expected to be accessed in place or replicated?
 1. Will the RDF graph be the primary data store of record?
 2. Will the source data and/or its metadata be extracted, transformed and loaded (ETL) into an RDF graph store?
5. Data requirements - sources, formats, modeling and loading:
 1. Source data format(s) – RDF, relational, XML, unstructured text, spreadsheets, etc.?
 2. Size of the data set overall (in bytes) and number of triples (or quads) or number of rows & columns?
 3. Number of graphs or data model(s)?
 4. Are the graphs static or dynamic?
 5. If the graphs are dynamic, how frequent are changes and how big are they?
 6. If the graphs are dynamic, must the inference graph be maintained?
 7. Is the data loaded in bulk or incrementally? How much is loaded at a time?
 8. How fast is the data expected to grow?
 9. Is there a need to track provenance (who/when/where/confidence, etc.)
6. Discovery requirements – query, visualization and analytics:
 1. Will SPARQL be the query language?
 2. Is SQL with embedded SPARQL graph queries (SEM_MATCH) be used?
 3. Examples of typical queries? Are they OLTP, or more complex & analytical?
 4. How are query results expected to be returned, reported, visualized?
 5. Is SPARQL query/update and a SPARQL endpoint required?
 6. Are data mining algorithms, or statistical and graphical analysis needed?
7. Discovery requirements – Inferencing:
 1. Is there a need for RDFS? OWL2 – RL, EL? SKOS? User-defined rules?
 2. Will a particular ontology be used? What ontology modeling tool will be used?
 3. Is full OWL 2 DL expressivity required?
8. Security requirements:
 1. Is triple or model level security needed for this application?
 2. How is security implemented today on the source data?
9. Text mining requirements:
 1. Is there a need to extract content from documents, web pages etc. to query and/or find documents of interest?

2. Is a natural language extraction tool being used or considered?

Conclusion

The RDF semantic graph feature of Oracles Spatial and Graph provides a robust and standards-based platform on which to build semantic solutions. Identifying the business requirements and benefits, project requirements, application functionality, and design considerations helps in planning and discussing the project with Oracle. This information can help Oracle provide recommendations and best practices to facilitate a successful project.



Oracle Spatial and Graph: RDF Semantic Graph
Project Planning
September 2012
Author: Oracle

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2012, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. UNIX is a registered trademark licensed through X/Open Company, Ltd. 0410

SOFTWARE. HARDWARE. COMPLETE.