Introduction to Graph Analytics and Oracle Cloud Service

Hans Viehmann
Product Manager EMEA
Oracle
@SpatialHannes

Jean Ihm
Product Manager US
Oracle
@JeanIhm

Korbi Schmid
Engineering Manager
Oracle

October 22, 2018
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, timing, and pricing of any features or functionality described for Oracle’s products may change and remains at the sole discretion of Oracle Corporation.
Program Agenda

1. Product Introduction
2. Use Cases
3. Feature Overview
4. Demo
Oracle’s Spatial and Graph Strategy

Enabling Spatial and Graph use cases on every platform

Oracle Database
Spatial and Graph

Oracle Big Data
Spatial and Graph

Spatial and Graph in Oracle Cloud

Database Cloud Service
Exadata Cloud Service
Two Graph Data Models

Property Graph Model
- Path Analytics
- Social Network Analysis
- Entity analytics

Linked Data Semantic Web
- Financial
- Retail, Marketing
- Social Media
- Smart Manufacturing

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

Linked Data Semantic Web
- Life Sciences
- Health Care
- Publishing
- Finance

Use Case Graph Model Industry Domain
Graph Database Features:

• Scalability, Performance, Security
• Graph Analytics
• Graph Query Language
• Graph Visualization
• Standard Interfaces
• Integration with Machine Learning tools
Oracle Products Supporting Property Graphs

**Oracle Big Data Spatial and Graph**

- Available for Big Data platform
  - Hadoop, HBase, Oracle NoSQL
- Supported both on BDA and commodity hardware
  - CDH and Hortonworks
- Database connectivity through Big Data Connectors or Big Data SQL
- Part of Big Data Cloud Service

**Oracle Spatial and Graph (DB option)**

- Available with Oracle 12.2/18c (EE)
- Using tables for graph persistence
- In-database graph analytics
  - Sparsification, shortest path, page rank, triangle counting, WCC, sub graph generation...
- SQL and PGQL queries possible
- Included in Database Cloud Services
Use Cases
Graph Analysis for Business Insight

Identify Influencers

Discover Graph Patterns in Big Data

Generate Recommendations

We think you might like this.

** We think you might like this.

*** We think you might like this.

**** We think you might like this.

***** We think you might like this.
Banco de Galicia

- Customer profitability analysis
  - Part of larger Hadoop/Big Data project
- Analysis of banking transactions
  - Focus on corporate customers
- Identification of undesired behavioural patterns, eg.
  - Customers using other banks to make large numbers of transactions
  - Many of which flow back to Banco Galicia
- Increase fees, terminate contracts, or move activities to Banco Galicia
- Implemented by Oracle Consulting
Romanian Police Force

• Creating Knowledge Graphs from all kinds of content
  – Social media networks, documents, images, audio, video, structured data
  – Using machine learning (text analysis, classification, entity extraction, face recognition, speech2text, ...)

• Enabling relationship analysis and semantic search

• bigCONNECT platform built by mWARE
  – Running on Big Data Appliance, Big Data Cloud Service or commodity Hadoop
Ministry of Finance, Eastern Europe

• Detecting relationships between people, accounts, companies
  – Similar to Paradise Papers

• Identifying suspicious patterns
  – Circular money transfers
  – Connections (existing path/shortest path) to companies in tax havens

• Ingesting accounting data in SAF-T format
  – Hadoop-based processing (Oozie, Spark, Hive)
  – Terabytes of data, rapidly growing

• Interactive graph analysis in Apex with Cytoscape.js

EU VAT fraud
Mazda

• Management of Bill-of-materials
  – Automotive manufacturing process
  – Supporting high variance and short innovation cycles
• Data coming from various sources
• Complex PGQL queries to associate parts and subcomponents
  – Performance as key requirement
  – Happy with response times and scaleability
Feature Overview
Oracle Graph Analytics Architecture

Graph Analytics
In-memory Analytic Engine

Graph Storage Management
Blueprints & SolrCloud / Lucene

Scalable and Persistent Storage
Property Graph Support on
Apache HBase, Oracle NoSQL or Oracle 12.2

Java APIs
Java APIs/JDBC/SQL/PLSQL

Cytoscape Plug-in
R Integration (OAAgraph)
Spark Integration

Python, Perl, PHP, Ruby,
Javascript, ...

REST Web Service

Copyright © 2018, Oracle and/or its affiliates. All rights reserved.
Interacting with the Graph

On-premise product geared towards data scientists and developers

• Access through APIs
  – Implementation of Apache Tinkerpop Blueprints APIs
  – Based on Java, REST plus SolR Cloud/Lucene support for text search

• Scripting
  – Groovy, Python, Javascript, ...
  – Apache Zeppelin integration, Javascript (Node.js) language binding

• Graphical UIs
  – Cytoscape, plug-in available
  – Commercial Tools such as TomSawyer Perspectives
Example: Betweenness Centrality in Big Data Graph

Code

```java
analyst.vertexBetweennessCentrality(pg).getTopKValues(15)
```
Pattern matching in Property Graphs using PGQL

- Finding a given pattern in graph
  - Fraud detection
  - Anomaly detection
  - Subgraph extraction
  - ...

- SQL-like syntax but with graph pattern description and property access
  - Interactive (real-time) analysis
  - Supporting aggregates, comparison, such as max, min, order by, group by

- Proposed for standardization by Oracle
  - Specification available on-line
  - Open-sourced front-end (i.e. parser)

https://github.com/oracle/pgql-lang
Basic graph pattern matching

• Find all instances of a given pattern/template in the data graph

```
SELECT v3.name, v3.age
FROM socialNetworkGraph
MATCH (v1:Person) -[:friendOf]-> (v2:Person) -[:knows]-> (v3:Person)
WHERE v1.name = 'Amber'
```

Query: Find all people who are known by friends of ‘Amber’.
Basic graph pattern matching

- Find all instances of a given pattern/template in the data graph

```sql
SELECT v3.name, v3.age
FROM socialNetworkGraph
MATCH (v1:Person) -[:friendOf]-(v2:Person) -[:knows]-> (v3:Person)
WHERE v1.name = 'Amber'
```

Query: Find all people who are known by friends of 'Amber'.
Regular path expressions

- Matching a pattern repeatedly
  - Define a **PATH** expression at the top of a query
  - Instantiate the expression in the **MATCH** clause
  - Match repeatedly, e.g. zero or more times (*) or one or more times (+)

```
PATH has_parent AS (child) -[:has_father|has_mother]-> (parent)
SELECT x.name, y.name, ancestor.name
FROM snGraph
MATCH (x:Person) -[:has_parent+]-> (ancestor)
    , (y) -[:has_parent+]-> (ancestor)
WHERE x.name = 'Peter' AND x <> y
```
Regular path expressions

- Matching a pattern repeatedly
  - Define a PATH expression at the top of a query
  - Instantiate the expression in the MATCH clause
  - Match repeatedly, e.g. zero or more times (*) or one or more times (+)

```sql
PATH has_parent AS (child) -[:has_father|has_mother]-> (parent)
SELECT x.name, y.name, ancestor.name
FROM snGraph
MATCH (x:Person) -[:has_parent]+/-:has_parent+/-> (ancestor)
, (y) - [:has_parent]+/-:has_parent+/-> (ancestor)
WHERE x.name = 'Peter' AND x <> y
```

```
<table>
<thead>
<tr>
<th>x.name</th>
<th>y.name</th>
<th>ancestor.name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter</td>
<td>Retta</td>
<td>Paul</td>
</tr>
<tr>
<td>Peter</td>
<td>Dwight</td>
<td>Paul</td>
</tr>
<tr>
<td>Peter</td>
<td>Dwight</td>
<td>Retta</td>
</tr>
</tbody>
</table>
```

Result set
Oracle Graph Analytics Architecture

Graph Analytics
In-memory Analytic Engine

Graph Storage Management
Blueprints & SolrCloud / Lucene

Scalable and Persistent Storage
Property Graph Support on
Apache HBase, Oracle NoSQL or Oracle 12.2

Java APIs
Java APIs/JDBC/SQL/PLSQL

Cytoscape Plug-in
R Integration (OAAgraph)
Spark Integration

Python, Perl, PHP, Ruby,
Javascript,...

REST Web Service

Copyright © 2018, Oracle and/or its affiliates. All rights reserved.
Support for Graph Pattern Matching

Graph Analytics
- In-memory Analytic Engine
- PGQL in PGX

Graph Storage Management
- Blueprints & SolrCloud / Lucene
- PGQL-to-SQL

Scalable and Persistent Storage
- Property Graph Support on Apache HBase, Oracle NoSQL or Oracle 12.2

Java APIs
- Java APIs/JDBC/SQL/PLSQL

Integration
- Python, Perl, PHP, Ruby, Javascript, ...

Blueprints & SolrCloud / Lucene
- REST Web Service

PGQL in PGX
- Cytoscape Plug-in
- R Integration (OAAgraph)
- Spark integration
Path Query (Parallel Recursive With)

**PGQL:**

```sql
PATH knows_path := () -[:knows]-> ()
SELECT s1.fname, s2.fname
WHERE (s1) -/:knows_path*/-> (o) </:knows_path*/-(s2)
ORDER BY s1,s2
```

**SQL:**

```sql
SELECT T2.T AS "s1.fname$T",T2.V AS "s1.fname$V",T2.VN AS "s1.fname$VN",T2.VT AS "s1.fname$VT",
FROM (/*Path[*/SELECT DISTINCT SVID, DVID FROM ( SELECT VID AS SVID, VID AS DVID FROM "GRAPH1VT$" UNION ALL SELECT SVID,DVID
FROM (WITH RW (ROOT, SVID, DVID, LVL) AS ( SELECT ROOT, SVID, DVID, 1 LVL FROM (SELECT SVID ROOT, SVID, DVID, 1 LVL
FROM (SELECT T0.SVID AS SVID, T0.DVID AS DVID FROM "GRAPH1GT$" T0 WHERE (T0.EL = n'knows'))
) UNION ALL SELECT DISTINCT RW.ROOT, R.SVID, R.DVID, RW.LVL+1 FROM (SELECT T1.SVID AS SVID, T1.DVID AS DVID FROM "GRAPH1GT$" T1 WHERE (T1.EL = n'knows')) R, RW WHERE RW.DVID = R.SVID )
CYCLE SVID SET cycle_col TO 1 DEFAULT 0 SELECT ROOT SVID, DVID FROM RW ))/*Path]*) T6,
(/*Path[*/SELECT DISTINCT SVID, DVID FROM ( SELECT VID AS SVID, VID AS DVID FROM "GRAPH1VT$" UNION ALL SELECT SVID,DVID
FROM (WITH RW (ROOT, SVID, DVID, LVL) AS ( SELECT ROOT, SVID, DVID, 1 LVL FROM (SELECT SVID ROOT, SVID, DVID, 1 LVL
FROM (SELECT T4.SVID AS SVID, T4.DVID AS DVID FROM "GRAPH1GT$" T4 WHERE (T4.EL = n'knows'))
) UNION ALL SELECT DISTINCT RW.ROOT, R.SVID, R.DVID, RW.LVL+1 FROM (SELECT T5.SVID AS SVID, T5.DVID AS DVID FROM "GRAPH1GT$" T5 WHERE (T5.EL = n'knows')) R, RW WHERE RW.DVID = R.SVID )
CYCLE SVID SET cycle_col TO 1 DEFAULT 0 SELECT ROOT SVID, DVID FROM RW ))/*Path]*) T7,
"GRAPH1VT$" T2, "GRAPH1VT$" T3
ORDER BY T6.SVID ASC NULLS LAST, T7.SVID ASC NULLS LAST
```

Find the **pairs of people** who are connected to a common person through the “knows” relation.
Notebook integration

- Multi-purpose notebook for data analysis and visualization
  - Browser-based script and query execution
- For documentation and interactive analysis
  - Typically used by Data Scientist
- Interpreters for graph analysis and graph pattern matching
  - PGX, PGQL, Markdown
- Graph visualization
- Integrated with Graph Cloud Service
Combining Graph Analytics and Machine Learning

**Graph Analytics**
- Compute graph metric(s)
- Explore graph or compute new metrics using ML result

**Machine Learning**
- Build predictive model using graph metric
- Build model(s) and score or classify data

Add to graph
Add to structured data
OAAgraph integration with R

- OAAgraph integrates in-memory engine into ORE and ORAAH
- Adds powerful graph analytics and querying capabilities to existing analytical portfolio of ORE and ORAAH
- Built in algorithms of PGX available as R functions
- PGQL pattern matching
- Concept of “cursor” allows browsing of in-memory analytical results using R data structures (R data frame), allows further client-side processing in R
- Exporting data back to Database / Spark allows persistence of results and further processing using existing ORE and ORAAH analytical functions
Graph Analytics on SPARK

- Use SPARK for conventional tabular data processing (RDD, Dataframe, -set)
- Define graph view of the data
  - View it as node table and edge table
- Load into PGX
- Execute graph algorithms in PGX
  - Orders of magnitude faster than GraphX
  - More scaleable
- Push analysis results back into SPARK as additional tables
- Continue SPARK analysis

SPARK data structure and communication mechanism not optimized for graph analysis workloads
Graph Visualization – Cytoscape, Vis.js, ...
Graph Visualization – Commercial Tools

Dedicated session on Tom Sawyer Perspectives integration
Distributed Graph Analysis Engine

Handling extremely large graphs

• Oracle Big Data Spatial and Graph uses very compact graph representation
  – Can fit graph with ~23bn edges into one BDA node

• Distributed implementation scales beyond this
  – Processing even larger graphs with several machines in a cluster (scale-out)
  – Interconnected through fast network (Ethernet or, ideally, Infiniband)

• Integrated with YARN for resource management
  – Same client interface, but not all APIs implemented yet

• Again, much faster than other implementations
  – Comprehensive performance comparison with GraphX, GraphLab
Graph Cloud Service

**Fully managed graph cloud service**

- “One-click” deployment: no installation, zero configuration
  - Automated failure detection and recovery
- Automated graph modeler
  - Easily convert your relational data into property graphs
- Pre-built Algorithms, Flows and SQL-like graph query language
  - Java, Groovy
  - Rest APIs
- Rich User Interface
  - Low code / zero code features
  - Notebook support and powerful data visualization features
Demo
Summary

Graph capabilities in Oracle Big Data Spatial and Graph

• Graph databases are powerful tools, complementing relational databases
  – Especially strong for analysis of graph topology and multi-hop relationships

• Graph analytics offer new insight
  – Especially relationships, dependencies and behavioural patterns

• Oracle Property Graph technology offers
  – Comprehensive analytics through various APIs, integration with relational database
  – Scaleable, parallel in-memory processing
  – Secure and scaleable graph storage using Oracle NoSQL, HBase or Oracle Database

• Available both on-premise or in the Cloud
## Spatial and Graph at OOW and Code One 2018

*View this list at [tinyurl.com/SpatialGraphOOW18]*

### Sessions

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, Oct. 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:00 am - 9:45 am</td>
<td><strong>Graph Query Language For Navigating Complex Data [DEV5447]</strong></td>
<td>Moscone West - 2016</td>
</tr>
<tr>
<td>10:30 am – 11:15 am</td>
<td><strong>When Graphs Meet Machine Learning [DEV5420]</strong></td>
<td>Moscone West - 2022</td>
</tr>
<tr>
<td>11:30 am – 12:15 pm</td>
<td><strong>Automate Anomaly Detection with Graph Analytics [DEV5397]</strong></td>
<td>Moscone West - 2022</td>
</tr>
<tr>
<td>1:30 pm – 2:15 pm</td>
<td><strong>How to Build Geospatial Analytics with Python and Oracle Database [DEV5185]</strong></td>
<td>Moscone West - 2003</td>
</tr>
<tr>
<td>2:30 pm – 3:15 pm</td>
<td><strong>Location-Based Tracking of Moving Objects with Apache Spark [DEV5355]</strong></td>
<td>Moscone West – 2016</td>
</tr>
<tr>
<td>4:45 pm – 5:30 pm</td>
<td><strong>Introduction to Graph Analytics and Oracle Graph Cloud Service [TRN4098]</strong></td>
<td>Moscone West – 3004</td>
</tr>
<tr>
<td>5:45 pm – 6:30 pm</td>
<td><strong>How to Analyze Data Warehouse Data as a Graph [TRN4099]</strong></td>
<td>Moscone West – 3004</td>
</tr>
</tbody>
</table>
## Spatial and Graph at OOW and Code One 2018

*View this list at* [tinyurl.com/SpatialGraphOOW18](http://tinyurl.com/SpatialGraphOOW18)

### Sessions

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tuesday, Oct. 23</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:45 pm – 2:30 pm</td>
<td>Machine Learning with R and Zeppelin on Oracle Big Data Solutions [PRO4054]</td>
<td>Moscone West - 3005</td>
</tr>
<tr>
<td>4:00 pm – 4:45 pm</td>
<td>Build Serverless Big Data and Graph Viz Web Apps with Spring Data and Core Java [DEV5479]</td>
<td>Moscone West - 2001</td>
</tr>
<tr>
<td>5:45 pm – 6:30 pm</td>
<td>Geo-Tagging, Geo-Enrichment, Geo-Fencing, and Tracking for Location-Enabled Apps [TRN4095]</td>
<td>Moscone West - 3003</td>
</tr>
<tr>
<td><strong>Thursday, Oct. 25</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00 am – 11:45 am</td>
<td>Using Location in Cloud Applications with Python, Node.js, and More [TRN4089]</td>
<td>Moscone West – 3001</td>
</tr>
<tr>
<td>2:00 pm – 2:45 pm</td>
<td>Analyzing Blockchain and Bitcoin Transaction Data as Graphs [DEV4835]</td>
<td>Moscone West – 2018</td>
</tr>
</tbody>
</table>
### Meet the Experts

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tuesday, Oct. 23</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00 am - 11:50 am</td>
<td>Location-Based Tracking and Geospatial Analytics for Database and Big Data Platforms [MTE6755]</td>
<td>Moscone West – The Hub – Lounge B</td>
</tr>
<tr>
<td>12:00 pm – 12:50 pm</td>
<td>Graph Analysis and Database Technologies [MTE6744]</td>
<td>Moscone West – The Hub – Lounge B</td>
</tr>
<tr>
<td>3:00 pm – 3:50 pm</td>
<td>Graph Analysis and Machine Learning [MTE6746]</td>
<td>Moscone West – The Hub – Lounge B</td>
</tr>
<tr>
<td><strong>Wednesday, Oct. 24</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00 am – 11:50 am</td>
<td>Graph Queries and Analysis [MTE6746]</td>
<td>Moscone West – The Hub – Lounge A</td>
</tr>
<tr>
<td>12:00 pm – 12:50 pm</td>
<td>Location-Based Tracking and Geospatial Analytics for Database and Big Data Platforms [MTE6755]</td>
<td>Moscone West – The Hub – Lounge B</td>
</tr>
</tbody>
</table>
### Demos

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday 9:45 am – 5:45 pm</td>
<td>Oracle Spatial and Graph Database, Analytics, and Cloud Services</td>
<td>Moscone South Exhibit Hall (‘The Exchange’)</td>
</tr>
<tr>
<td>Tuesday 10:30 am – 5:45 pm</td>
<td></td>
<td>• Oracle Demogrounds: Cloud Platform &gt; Application Development area</td>
</tr>
<tr>
<td>Wednesday 10:30 am – 4:45 pm</td>
<td></td>
<td>• Kiosk # APD-WU3</td>
</tr>
</tbody>
</table>
Moscone South Exhibit Hall / Demogrounds (“The Exchange”)

Oracle Spatial and Graph demopod (behind the Keynote Arena)
Kiosk #: APD-WU3

Intelligent Business Applications

Application Development

Cloud Platform
The Spatial & Graph SIG User Community

We are a vibrant community of customers and partners that connects and exchanges knowledge online, and at conferences and events.

Join us!

tinyurl.com/oraclespatialcommunity

LinkedIn Google+ IOUG SIG

@oraspatailsg

oraclespatialsg@gmail.com
Call for Speakers now open!
Submit an abstract to share your use case or technical session

Analytics and Data Summit
All Analytics. All Data. No Nonsense.
March 12 – 14, 2019

Formerly called the BIWA Summit with the Spatial and Graph Summit
Same great technical content...new name!
www.AnalyticsandDataSummit.org