Introduction to Graph Cloud Services, Database, and Analytics

Xavier Lopez, Senior Director Product Management, Oracle
Zhe Wu, Architect, Oracle
Masahiro Yoshioka, Principal Engineer, IT Solutions Division, Mazda

October 2, 2017
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. Product Introduction
2. Use Cases
3. Feature Overview
4. Demo
5. Mazda Example
Oracle’s Spatial and Graph Strategy

On Premise and Oracle Cloud

Oracle Database Spatial and Graph

Oracle Big Data Spatial and Graph

Spatial and Graph in Oracle Cloud
Two Graph Data Models

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

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

**Use Case**
- Graph Model
- Industry Domain

**Social Network Analysis**
- Financial
- Retail, Marketing
- Social Media
- Smart Manufacturing

**Linked Data**
- Life Sciences
- Health Care
- Publishing
- Finance
Graph Database Features:

• Scalability and Performance
• Graph analytics
• Graph Visualization
• Graph Query Language
• Standard interfaces
• Integration with Machine Learning tools
# Graph Product Options

## Oracle Big Data Spatial and Graph
- Available for Big Data platform/BDCS
  - Hadoop, HBase, Oracle NoSQL
- Supported both on BDA and commodity hardware
  - CDH and Hortonworks
- Database connectivity through Big Data Connectors or Big Data SQL
- Included in Big Data Cloud Service

## Oracle Spatial and Graph (DB option)
- Available with Oracle 12.2 / DBCS
- Using tables for graph persistence
- Graph views on relational data
- In-database graph analytics
  - Sparsification, shortest path, page rank, triangle counting, WCC, sub graphs
- SQL queries possible
- Included in Database Cloud Service
Use Cases
Graph Analysis for Business Insight

Identify Influencers

Discover Graph Patterns in Big Data

Generate Recommendations
Some Use Case Scenarios

• **Finance**
  – Customer 360, Fraud detection

• **Public Sector**
  – Tax Evasion, Crime network analysis

• **Retail**
  – Recommendation, sentiment analysis

• **Manufacturing**
  – Analyzing complex bill of materials (BoM)
Financial Services
Applying Graph Analysis To Improve Customer Service

• Model customer relationship to products, services, people, places.
• Analyze money customer’s flow between non-bank to bank accounts
• Combine internal CRM data with enterprise and social media content
• Identify high-value customers across business divisions
• Enhance new product/service opportunities
• Provide Real-time recommendations
Tax Fraud Analysis
Chinese Province Tax Office

Challenge:

– Modeling relationships between individuals and corporations
– Ingest documents, social media, web content, and publically available open data
– Create a ‘picture’ of the taxpayer network
  • Taxpayer relationship with other taxpayers
  • If a company structure, identify associated directors and shareholders in that company
  • Relationship between taxpayer’s and their associates’ financial affairs
  • Identify relevant intermediaries acting on behalf of taxpayer
– Explore tax evasion and fraud, trigger a formal case investigation
Analyzing Blockchain Ledger Transactions
Land Management, Banking, Public Services

- Distributed Ledgers being adopted in Finance, Public Sector
- Load and manage massive transactions from a distributed digital ledger
- Efficiently traverse a blockchain transaction graph
- Query and visualize – search for patterns of activity
Public Security: Analyzing Criminal Networks
Chinese Police Department

Business Requirement
– Model relationships between known and suspected criminals
– Ingest documents, social media, web content, chat rooms, flight records, hotel stay registries, and publically available open datasets.

How graph analysis solves the problem
• Search for known individuals in web of content
• Analyze relationship with other criminals, travel history, addresses, employers
• Relationship between suspects and their financial affairs
IT Network Modeling & Monitoring

• Model cyber network topology as a Graph
• Identify CyberNetwork intrusions
  – Combine deep learning with graph analytics
• Visualize real-time state of CyberNetwork
• Analyze impact of component failure on an IoT system?
  – Reachability analysis: understand which routines, libraries, servers, routers are affected by a modification
Automotive Manufacturing
Support high variance, short innovation cycles of complex autos

Graph View of Enterprise Data

• Unified graph representation of BoM, Configuration, CAE, Simulation...
• Generate “graph view” of relational data, or model instance data as graph
• Apply graph query and search across BoM and configuration models
• Apply graph analytics
• Scale to trillions of nodes and edges
Feature Overview
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
Relational Model vs. Graph Model

- Relational Model

- Graph Model

Courtesy: Tom Sawyer 2016
Architecture of Property Graph Support
Architecture of Property Graph Support

Graph Analytics

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

Java APIs

Graph Data Access Layer (DAL)

Blueprints & Lucene/SolrCloud

Java APIs/JDBC/SQL/PLSQL

Scalable and Persistent Storage Management

Oracle RDBMS

Apache HBase

Oracle NoSQL Database

Property Graph formats

GraphML

GML

Graph-SON

Flat Files

REST/Web Service/Notebooks

Java, Groovy, Python, …
Architecture of Property Graph Support

- **Graph Data Access Layer (DAL)**
  - **Graph Analytics**
    - Parallel In-Memory Graph Analytics/Graph Query (PGX)
  - **Apache Spark**
  - **Java APIs**
  - **Blueprints & Lucene/SolrCloud**
  - **Java APIs/JDBC/SQL/PLSQL**

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

- **Property Graph formats**
  - GraphML
  - GML
  - Graph-SON
  - Flat Files

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

- **Oracle NoSQL Database**
  - **Oracle RDBMS**
  - Apache HBase
  - **Scala**

**Copyright © 2017, Oracle and/or its affiliates. All rights reserved.**
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

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

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)

... and parallel graph mutation operations

The original graph

Left Set: “a,b,e”

Create Bipartite Graph

Sort-By-Degree (Renumbering)

Filtered Subgraph

Filter-Expression

Create Undirected Graph

Simplify Graph

Create Directed Graph
Graph Analysis Algorithms can be very hard to code ...

BDSG and OSG Property Graph comes with 40+ pre-built algorithms

• Example: Find the size of the 2-hop network of vertices (Gremlin+Python)

```python
sum([v.query() \n    .direction(blueprints.Direction.OUT).count() \n    for v in OPGIterator(v0.query() \n    .direction(blueprints.Direction.OUT) \n    .vertices().iterator())])
```

• Single API call instead
  – Analysis in memory, in parallel

• Results can be persisted in Graph store and accessed from Oracle Database
  – Big Data SQL, Connectors
Text Search through Apache Lucene/SolrCloud

Why?

– Contribute to the performance of graph traversal queries
– Constrained to be uniform in type among the indexed elements (vertices or edges)

Automatic Indexes

– Automatic update based on a subset of property keys
– Avoid linear scan to access an element by key/value

Manual Indexes

– Maintained by users
– Fasten up text searches by a particular key/value pair
– Sub-graphs based on a set of (existing or temporary) properties
Visualizing Property Graphs (with Cytoscape)

- Cytoscape supports Property Graph
- Connects to Oracle Database, Oracle NoSQL Database, or Apache HBase
- Runs Page Rank, Clustering, Shortest Path, etc
- Alternative to command-line for in-memory analytics once base graph created
Additional Graph Visualization Partners
Tom Sawyer, Cambridge Intelligence, Linkurios, Vis.js,...
Pattern matching using PGQL

• 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

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

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

https://github.com/oracle/pgql-lang
Zeppelin Frontend

• Apache Zeppelin
  – **Multi-purpose notebook** for data analysis and visualization
  – Enables to embed interactive execution inside Browsers
  – Renders execution results with plots and tables within Browsers

• PGX provides a hook (interpreter) for Zeppelin integration
Interacting with the Graph

• Access through APIs
  – Implementation of Apache Tinkerpop Blueprints APIs
  – Based on Java, REST plus SolR Cloud/Lucene support for text search
  – SQL/PLSQL for property graph functions in Oracle Database

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

• Graphical UIs
  – Cytoscape, plug-in available for BDSG
  – Commercial Tools such as TomSawyer Perspectives, Ogma
Enhancing ML and Data Analytics with Graphs

• Graph analysis can enhance the quality of ML and data analytics

• Graph representation helps discover hidden information about the data
  – Multi-hop relationship between data entities

• This can be used to further improve predictive models in R, Advanced Analytics, machine learning

<table>
<thead>
<tr>
<th></th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>Feature 3</th>
<th>Feature 4</th>
<th>Feature 5</th>
<th>Feature 6</th>
<th>Feature 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Machine Learning Predictive Models (R, Advanced Analytics)
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
Demo
We Have Many Property Graph Demos
Demo booth at Moscone West SOA 127 (Oracle’s Graph Database)

- Fraud Detection
- Graph Construction
- Notebooks
- Deep Learning Integration
- Graph Studio
- Network Intrusion Detection
- Bitcoin/Blockchain
- Recommender System
- Graph Visualization
Mazda Example
Who Is MAZDA...?

1920  Founded as 『Toyo Cork Kogyo Co., Ltd
1927  Renamed as 『Toyo Kogyo Co., Ltd
1929  Started the production of motorcycle
1984  Renamed as Mazda Motor Corporation
2020  Centennial anniversary

Sales price was around $3.5 ~ $3.8 then.

1931  Three-wheeler truck

1960 Mazda R360
(The very first passenger vehicle)
## Corporate Profile

<table>
<thead>
<tr>
<th>Company name</th>
<th>Mazda Motor Corporation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Founded</td>
<td>January 30, 1920</td>
</tr>
<tr>
<td>Headquarters</td>
<td>Hiroshima / Japan</td>
</tr>
<tr>
<td>Revenue</td>
<td>$30 Billion (FYE Mar 2017)</td>
</tr>
<tr>
<td>Retail Volume</td>
<td>1.5 million units (same FY as above)</td>
</tr>
<tr>
<td>Number of employees</td>
<td>48,749 (consolidated) (same FY as above)</td>
</tr>
<tr>
<td>R&amp;D center</td>
<td>5 locations (Hiroshima, Yokohama, US, Germany, China)</td>
</tr>
<tr>
<td>Production Site</td>
<td>3 factories in Japan (Hiroshima Plant, Hofu Plant, overseas)</td>
</tr>
<tr>
<td></td>
<td>Hiroshima Plant (Head Office, Ujina), Hofu Plant (Nishinoura, Nakanoseki), China, Thailand, Mexico, Vietnam, Malaysia, Russia</td>
</tr>
</tbody>
</table>
Mazda Plant
Mazda Plant
Mazda’s Problem

Imagine
Auto Manufacturer
Vehicle

Vehicle

Parts
(constructed by small parts)
### Mazda’s Problem

**Data Structure**

<table>
<thead>
<tr>
<th>Relational ?</th>
<th>Graph ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many Business Domain</td>
<td>Which Data Structure is better for each Data ?</td>
</tr>
<tr>
<td>Finance</td>
<td></td>
</tr>
<tr>
<td>Sale / Marketing</td>
<td></td>
</tr>
<tr>
<td>Production</td>
<td></td>
</tr>
<tr>
<td>Bill Of Materials</td>
<td></td>
</tr>
</tbody>
</table>
...
# Mazda’s PoC

<table>
<thead>
<tr>
<th></th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ite0</td>
<td>8</td>
<td>15</td>
<td>22</td>
<td>29</td>
<td>5</td>
<td>12</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>Ite1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ite2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ite3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ite4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ite5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ite6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ite7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2017, Oracle and/or its affiliates. All rights reserved.
Mazda’s PoC (4th Stage)

Total number of Edges : 53,993,161
Total number of Nodes : 7,099,473
Mazda’s PoC (4th Stage)

Number of Nodes are shown in blue color
Number of Edges are shown in black color

Number of Nodes:
- N1: 2,086
- N2: 39,213
- N3: 39,213
- N4: 8,395,290
- Na: 23,385
- Ne: 23,727
- Ni: 6,027
- Nm: 553,773
- Ns: 2,291,840
- Nu: 2,800,750
- Np: 2,798,431
- Nf: 3,385,933
- Ns: 709,030

Number of Edges:
- 4: 2,798,431
- 7: 8,395,290
- 8: 2,800,750
- 2017, Copyright © Oracle and/or its affiliates. All rights reserved.
Performance (PGQL Query)

<table>
<thead>
<tr>
<th>Nm</th>
<th>Num</th>
<th>Query time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaaaaaaa</td>
<td>62</td>
<td>43</td>
</tr>
<tr>
<td>bbbbbbbb</td>
<td>66</td>
<td>51</td>
</tr>
<tr>
<td>cccccccc</td>
<td>78</td>
<td>46</td>
</tr>
</tbody>
</table>
Summary (Current Result)

• Performance is Good!

• Issues: Refinement of complex PGQL queries

• Next Step: On going collaboration with Oracle Team
  • Oracle Japan, US Development, Oracle Labs
Overview:
Complete Graph Solution

• Distributed graph database
• Distributed in-memory analytics
• Graph Visualization
• Graph Query Language (PGQL)
• Standard interfaces
• Available on premise and Oracle Cloud
## Spatial and Graph Sessions

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, Oct. 2</td>
<td>2:15 pm - 3:00 pm Leveraging the Power of Graph Analytics to Fight Financial Crimes [CON2495]</td>
<td>Park Central (Floor 2) – Metropolitan III</td>
</tr>
<tr>
<td>Tuesday, Oct. 3</td>
<td>5:45 pm – 6:30 pm Fake News, Trolls, Bots, and Money Laundering: Find the Truth with Graphs [CON6683]</td>
<td>Park Central - Franciscan I</td>
</tr>
</tbody>
</table>

## Spatial and Graph Demos

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday - Wednesday</td>
<td>Oracle’s Spatial Technologies for Database, Big Data, and the Cloud</td>
<td>Moscone West Exhibit Hall 1&lt;sup&gt;st&lt;/sup&gt; floor Oracle Cloud Platform &gt; Analytics &amp; Big Data, pod SOA 131</td>
</tr>
<tr>
<td>Monday - Wednesday</td>
<td>Oracle’s Graph Database and Analytics for Database, Big Data, and the Cloud</td>
<td>Moscone West Exhibit Hall 1&lt;sup&gt;st&lt;/sup&gt; floor Oracle Cloud Platform &gt; Analytics &amp; Big Data, pod SOA 127</td>
</tr>
</tbody>
</table>
www.AnalyticsandDataSummit.org

Call for speakers is now open with rolling acceptances.