Understanding Complex Spatial and Graph Relationships: On Database, Big Data, and Cloud

Jayant Sharma
Director, Product Management

BIWA and Spatial Summit, 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.
IDC 2016 Futurescape Top 10 Predictions

1. By the End of 2017, Two-Thirds of the CEOs of the G2000 Will Have Digital Transformation at the Center of Their Corporate Strategy

2. By 2017, 60% Companies with a DX Strategy Will Deem It Too Critical for Any One Functional Area and Create an Independent Corporate Executive Position to Oversee the Implementation

3. By 2018, 80% of B2C Companies Will Have Created Immersive, Authentic Omni-Experiences for Customers, Partners, and Employees; 60% of B2B-Centric Companies Will Have Done the Same

4. The Top New Investment Areas Through 2017 Will Be Contextual Understanding and Automated Next Best Action Capabilities

5. In 2016, 65% of Large Enterprises Will Have Committed to Become Information-Based Companies, Shifting the Organizational Focus to Relationships, People, and Intangible Capital

6. By 2018, 75% of the G2000 Will Have Deployed Full, Information-Based, and Economic Models or "Digital Twins" of Their Products/Services, Supply Network, Sales Channels, and Operations

7. By 2020, 60% of the G2000 Will Have Doubled Their Productivity by Digitally Transforming Many Processes from Human-Based to Software-Based Delivery

8. In 2016, the Level of Connectivity Related to Products, Assets, and Processes Will Increase 50% for All Industry Value Chains

9. The Sharing Economy Will Give Rise to the Networked Free Agent and Skill-Based Marketplaces, Resulting in More than 10% of Work Being Sourced in this Fashion in Mature Economies by 2019

10. By 2018, at Least 20% of All Workers Will Use Automated Assistance Technologies to Make Decisions and Get Work Done

http://www.idc.com/research/viewtoc.jsp?containerId=259616
“Being context aware is about knowing location, identity, activity and time.”

– “Location is a context, not THE context”, Helene Anderson.
   https://www.keypoint-tech.com/location-context-context/
Organizing Information


Illustration by Nancy Duarte.
http://www.duarte.com/blog/organizing-information-is-finite/
Smart Analytics for Everyone

ANALYSTS
• Faster time to insights
• Share, collaborate and maximize community’s wisdom

BUSINESS LEADERS
• Easy, timely, proactive business insights
• Empower to go beyond consuming others’ findings

DEVELOPERS
• Develop and deploy analytic applications for people to customize and personalize
• Embed analytics in any application context
“While hardcore GIS professionals may start their work in other applications, when they want to solve spatial problems in production and with web- and IoT-scale data, Oracle gives them the platform to do so.”

Analysts: Rowan Curran with Holger Kisker, Ph.D. and Emily Miller
September 1, 2016
Graph and Spatial Analysis – It is about relationships

• Are things in the same location? Who is the nearest? What tax zone is this in? Where can we deliver in 35 minutes? What is in my sales territory? Is this built in a flood zone?

• Which supplier am I most dependent upon? Who is the most influential customer? Do my products appeal to certain communities? What patterns are there in fraudulent behavior?
Oracle Spatial and Graph

Deployable Services
- Geocoding
- Routing
- Web Services (OGC)

Spatial & Graph

- Polygons
- Lines
- Points
- Networks
- RDF Graphs
- Property Graphs (new in 12.2 on Oracle Cloud)
- 3D / LIDAR
- Topologies

ORACLE

SQL

Java

REST
Oracle Big Data Spatial and Graph

Spatial Analysis Features for:
- Location Data Enrichment
- Proximity and containment analysis
- Spatial data preparation (Vector and raster data)

Property Graph for Analysis of:
- Social Media Relationships
- eCommerce Targeted Marketing
- Cyber-Security, Fraud Detection
- IoT, Industrial Engineering
Spatial and Graph 12.2
GeoJSON support

- select json_value('{{"type": "Point", "coordinates": [125.6, 10.1]}}', '$' returning sdo_geometry) from dual

- Support spatial index and spatial queries on JSON documents

- Run Spatial operators on JSON in Oracle or accessible through External Tables

- Extend JSON_VALUE() to support GeoJSON and SDO_GEOMETRY
  - Support the push down to Exadata smart scan

- SDO_GEOMETRY constructors extended to take JSON as input
Location tracking APIs
For Cloud-based tracking services

- Track millions of moving objects against thousands of regions
- Java API using JMX to capture and track location-based events using spatial attributes
- Parallel polygon analysis for multiple moving objects and tracking regions
Performance and Usability for Raster analysis

- J2EE Catalog and Console for managing Rasters in files and database
  - Perform raster operations on data stored both in database and in files
  - Create “virtual images” spanning database and files
  - Move file-based content into database
- Improve performance of raster algebra operations
  - 30 % to 40%
- Web Coverage Server for Rasters in DB
  - Extend support for OGC standards
- JPEG 2000 compression in the DB
Property Graph

• Massively-Scalable Graph Database
  – Scales to **trillions** edges

• Memory-based Graph Analytics
  – More than 35 graph analysis algorithms

• Simple Standard interfaces
  – SQL, Java
  – Tinkerpop: Blueprints, Gremlin, Rexster
  – Groovy, Python

Detecting Components and Communities

Ranking/Walking

Evaluating Communities

Path-Finding
Big Data Spatial and Graph

Oracle Big Data Spatial and Graph

NoSQL

Spark
Big Data Spatial and Graph

Big Data often needs an Organizing principle – Data Harmonization

Big Data analysis is often about relationships not aggregation

Big Data platform is economically compelling for working with massive data sets found in spatial and graph workflows
Motivation: Emergence of Hadoop for Spatial Processing in the Insurance Industry

- Actuarial and Demographic data
- Accident data
- Call data
- Customer data
- Enrich with Postal Code
- Categorize by Region

86% Of Insurance companies agree that analyzing multiple data sources together is crucial to making accurate predictions.

88% Agree that linking information by location is key to combining disparate sources of Big Data.

Source: “The big data: How data analytics can yield underwriting gold. Survey conducted by Ordnance Survey and Chartered Insurance Institute, 25 April 2013.”
Oracle Big Data Spatial and Graph (BDSG)

Spatial Analysis Features

Property Graph Database
What problems can Big Data Spatial analysis address?

1. Data Harmonization using any location attribute (address, postal code, lat/long, placename, etc).
2. Preparation, validation and cleansing of Spatial and Raster data.
3. Categorization and filtering based on location and proximity.
4. Visualizing and displaying results on a map.
# Big Data Spatial Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data enrichment service API using GeoNames and geometry hierarchy data</td>
<td>MapReduce routines for distance calculations, PointInPolygon, buffer creation, Categorization, KMeansClustering, Binning</td>
</tr>
<tr>
<td>Spatial processing of data stored in HDFS or NoSQL. Raster processing operations: Mosaic and sub-set operations. Geodetic and Cartesian data</td>
<td>HTML5 Map Visualization API</td>
</tr>
<tr>
<td></td>
<td>Hive SQL API Query from Oracle DB with Big Data SQL &amp; Oracle SQL Connectors for Hadoop</td>
</tr>
</tbody>
</table>
Data Harmonization: Linking information by location

Are these data points related?

• Tweet: sailing by #goldengate
• Instagram image subtitle: 골든게이트 교*
• Text message: Driving on 101 North, just reached border between Marin County and San Francisco County
• GPS Sensor: N 37°49′11″ W 122°28′44″
• Now find all data points around Golden Gate Bridge ...

* Golden Gate Bridge (in Korean)
Store any business data with spatial information in HDFS

Oracle JGeometry

ANY BUSINESS DATA

USER-PROVIDED InputFormat/RecordReader Class

LOAD ANY FORMAT

HDFS
Vector Data Processing API Functions

Single Geometry

- Length
- Area
- Buffer
- Simplify

Geometry Pairs

- Range Queries
  - Point in Polygon
  - Touch, Overlap, Intersect, Contains, Any Interaction
- Join Queries
  - Interactions on sets of data
  - E.g.: Find all the dropped cell calls in all coverage areas

Categorization and Enrichment

- Associate a data set with a known geometry or named hierarchy
  - Process all Tweets for a period of time and count how many are associated with each city, county, state, etc.
Data Categorization Services

Any hierarchical geometry data set for reference

Customers choose a set of layers. For example, they can select (continents, countries, cities) or (countries, states, counties) as the hierarchy.

Big Data Spatial map-reduce job processes the customer data and produces a result file.
Big Data Spatial and Graph
Spatial Vector Processing Framework

- Mapper and Reducer Classes
- S&G Java API
  - Spatial Operators, Functions
  - Spatial Enrichment, Categorization API
- GeoJSON, JGeometry format
- RecordReader class
- HDFS
  - Spatial Content (any format)
  - Enrichments, Categorizations results
- MapReduce Framework, templates
- Customer Application
- Sample Console Application

- Generated data
- Oracle Provided
- Customer data
- Customer code

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
Image Server

• HDFS storage for the image or raster files
  – We can support dozens of file formats (GDAL supported formats)
  – Images are geo-referenced
  – Images can be in different coordinate systems and resolutions

• Three main capabilities
  – Loader to load raster data from NFS to HDFS
  – Mosaic and subset operations based on a virtual mosaic
  – Image processing framework for raster analysis
Spatial in Spark

• Spatial Resilient Distributed Datasets
• Distributed Spatial Indexes
• Spatial Hive (UDF) functions can be used with Spark SQL
Oracle Big Data Spatial and Graph (BDSG)

Spatial Analysis Features

Property Graph Database
Relational Model vs. Property Graph Model

- Relational Model

- Graph Model

Courtesy: Tom Sawyer 2016
Who is most important? There Are Lots of Answers.

• Answers from **Aggregation**
  – Who spends the most?
  – Who buys the highest margin goods?
  – Who is most consistently a top contributor?

• Answers from **Connectivity**
  – Who’s most influential?
  – Which supplier do I depend on the most?
  – What is the right product mix for millennials?

**Tabular questions:**
Well-suited to SQL-like tools

**Graph questions:**
We need something different!
Graph Analytics: New Capabilities Using Graphs / Networks: Do Pattern Recognition, Identify New Relationships

Recommend the most similar item purchased by similar people

Find out people that are central in the given network – e.g. influencer marketing

Identify group of people that are close to each other – e.g. target group marketing

Find out all the sets of entities that match to the given pattern – e.g. fraud detection

Product Recommendation

Influencer Identification

Community Detection

Graph Pattern Matching
Detecting Fraud in Financial Transactions

• Transaction Data (from a Tax dept) in CSV
  – Converted into Oracle flat files for property graph

• Pattern Matching and Visualization
Oracle BDSG Analyzes Panama Papers

Source: community.oracle.com/docs/DOC-1006400
Supply Chain / CyberNetwork Modeling

Impact Analysis on Business and IT Networks

• What is impact of given supplier on supply chain?
  – Reachability analysis to understand which products (and $ value) affected by each supplier

• How does the impact change with a recent acquisition?

• What is impact of a component outage on a CyberNetwork system?
  – Reachability analysis to understand which routines, libraries, servers, routers are affected by a modification
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 model
Architecture of Property Graph Support in Big Data Spatial and Graph v2.0+

Graph Analytics
Parallel In-Memory Graph Analytics (PGX)

Data Access Layer
Apache Blueprints & Lucene/SolrCloud

Oracle Big Data Spatial and Graph
Apache HBase
Oracle NoSQL Database

Property graph formats supported
GraphML
GML
Graph-SON
Flat Files
CSV
Relational Data Sources

Java, Groovy, Python, ...

REST/Web Service

Java APIs
35+ Graph Functions

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

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

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

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

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

Other Classics
- Vertex Cover
- Minimum Spanning-Tree (Prim’s)
Oracle geospatial platform
Database and Big Data Platform support, On-premise and in the Cloud

Oracle Big Data Spatial and Graph

Oracle Database Spatial and Graph

Spatial and Graph in Oracle Cloud
Database Cloud Services for Enterprise

<table>
<thead>
<tr>
<th>Service</th>
<th>Schema CS</th>
<th>Express CS</th>
<th>Database CS</th>
<th>Exadata CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database Development</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>SMB &amp; Departmental Applications</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Enterprise Applications</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Highest performance</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Customer at Cloud Choices</td>
<td>✔️</td>
<td>✔️</td>
<td>Cloud Machine</td>
<td>Exadata Cloud Machine</td>
</tr>
</tbody>
</table>
Database Cloud Services for Enterprise

Spatial and Graph

Schema CS
High Performance and Extreme Performance editions
HTTP access only

Express CS
Coming soon...

Database CS
Elastic Compute or Bare Metal
High Performance and Extreme Performance editions

Exadata CS
Included
Oracle Big Data Cloud Service

Key Features
- Dedicated Compute Shapes with Direct Attached Storage
- Hadoop, Spark delivered as an automated Cloud Service
  - Cloudera Enterprise – Data Hub Edition 5.x
  - Oracle Big Data Connectors
  - **Oracle Big Data Spatial and Graph**
  - Oracle Data Integrator Enterprise Edition
- Platform for new Big Data Services
  - Big Data Discovery
  - Big Data SQL (Coming Soon)

Benefits
- Consistently high performance
- Secure by Default
- Comprehensive Software Stack
Spatial and Graph Sessions

• 30+ Spatial and Graph related sessions
  – See yellow track on agenda plus crossover sessions (see small colored boxes)
  – Room 103 for most sessions

• Tuesday: Spatial technical sessions & intro to Graph
• Wednesday: Spatial industry use cases
  --Insurance, retail, airports, Smart 3D City modeling & more
• Thursday: Graph technical sessions & use cases
Selected Spatial and Graph Sessions

• Maps & BI (Spatial)
  – Wayne Van Sluys: “Getting Started with Maps in OBIEE, BICS and Data Visualization Desktop” -- Wednesday 11:10 am
  – Hands-on Lab: Arthur Dayton: “Deploy Custom Maps in OBIEE” -- Wednesday 2:20 pm

• Graph
  – Mark Rittman: “Understanding how a Tweet goes Viral using Oracle Big Data Spatial and Graph” -- Wednesday 9:00 am
  – Zhe Wu/Hans Viehmann: A Shortest Path to Using Graph Technologies - Best Practices in Graph Construction, Indexing, Analytics and Visualization -- Tuesday 4:30 pm

• Hands On Labs on Big Data location analysis, graph social network analysis
• Many more!
Engage with the Spatial and Graph SIG

Promotes interaction and communication to drive the market for spatial technology and data

Members connect and exchange knowledge via online communities and at conferences and events

- Talk with us at the Summit!
  - Morning Arrivals
    - Tues & Wed
    - 7:45-8:30 a.m.
    - Registration Area
  - Receptions
    - Tues & Wed evenings
    - Spatial Table, lobby
  - Birds of a Feather Lunch
    - Wednesday
    - 12-1pm
    - Auditorium

- tinyurl.com/oraclespatialcommunity
- Search for “Oracle Spatial and Graph Community”
- Contact us:
  - oraclespatialsig@gmail.com
  - @oraspawtialsig

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.