How to Analyze Data Warehouse Data as a Graph

How graph databases can complement relational technology

Hans Viehmann
Product Manager EMEA
ORACLE Corporation
HOUG Conference 2019

@SpatialHannes
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.
Graph – an important growth area for data & analytics

Gartner Identifies Top 10 Data and Analytics Technology Trends for 2019

Trend No. 5: Graph

Graph analytics is a set of analytic techniques that allows for the exploration of relationships between entities of interest such as organizations, people and transactions.

The application of graph processing and graph DBMSs will grow at 100 percent annually through 2022 to continuously accelerate data preparation and enable more complex and adaptive data science.

Graph data stores can efficiently model, explore and query data with complex interrelationships across data silos, but the need for specialized skills has limited their adoption to date, according to Gartner.

Graph analytics will grow in the next few years due to the need to ask complex questions across complex data, which is not always practical or even possible at scale using SQL queries.

A social network generated from Game of Thrones.

The color of a vertex indicates its community.

The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality.

An edge’s thickness represents its weight.

https://www.macalester.edu/~abeverid/thrones.html
Graph Data Model

• What is a graph?
  – Data model representing entities as vertices and relationships as edges
  – Optionally including attributes
  – Also known as „linked data“

• What are typical graphs?
  – Social Networks
    • LinkedIn, facebook, Google+, ...
  – IP Networks, physical networks, ...
  – Knowledge Graphs
    • Apple SIRI, Google Knowledge Graph, ...
Why are graphs popular?

– Easy data modeling
  • „whiteboard friendly”
– Flexible data model
  • No predefined schema, easily extensible
  • Particularly useful for sparse data
– Insight from graphical representation
  • Intuitive visualization
– Enabling new kinds of analysis
  • Overcoming some limitations in relational technology
  • Basis for Machine Learning (Neural Networks)
Oracle’s Spatial and Graph Strategy
Enabling Spatial and Graph use cases on every platform

Oracle Database
Spatial and Graph Option

Oracle Big Data
Spatial and Graph

Cloud
Services

Exadata
Non-Engineered Systems

Big Data Appliance
Commodity Hadoop
Spark

Database Cloud Service
Exadata Cloud Service
Categories of Graph Analysis

Computational Graph Analytics

- Compute values on vertices and edges
- Traversing graph or iterating over graph (usually repeatedly)
- Procedural logic
- Examples:
  - Shortest Path, PageRank, Weakly Connected Components, Centrality, ...

Graph Pattern Matching

- Based on description of pattern
- Find all matching sub-graphs

![Graph Pattern Example]

- :Person{100} name = 'Amber' age = 25
- :Person{200} name = 'Paul' age = 30
- :Person{300} name = 'Heather' age = 27
- :Company{777} name = 'Oracle' location = 'Redwood City'
- :worksAt{1831} startDate = '09/01/2015'
- :friendOf{1173} since = '08/01/2014'
- :Person{2513} friendOf since = '08/01/2014'
- :Person{2200} knows since = '08/01/2014'
- :Person{300} friendOf {2513} since = '08/01/2014'
Example: Fraud Detection in Healthcare

• Finding anomalies in healthcare billing data
  – Public domain dataset
  – Medical providers and their operations

• Question
  – Are there any medical providers that are suspicious
  ➔ medical providers that perform different operations than their fellows
  (e.g. eye doctors doing plastic surgery)

• Approach
  – Create graph between doctors and operations
  – Apply personalized pagerank (a.k.a equivalent to random walking)
  – Identify doctors that are far from their fellows
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’.
Oracle Graph Analytics Architecture

Graph Analytics
In-memory Analytic Engine

Graph Storage Management
Blueprints & SolrCloud / Lucene

Scalable and Persistent Storage

Visualization
R Integration (OAAgraph)
Spark Integration

Java APIs

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

REST Web Service

Java APIs/JDBC/SQL/PLSQL
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

Visualization
- R Integration (OAAgraph)
- Spark integration

Java APIs
- Java APIs
- Java APIs/JDBC/SQL/PLSQL

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

Spark integration

PGQL in PGX
PATH knows_path := () -[:knows]-> ()
SELECT s1.fname, s2.fname
WHERE (s1) -[:knows_path*]-> (o) <[:knows_path*]-(s2)
ORDER BY s1,s2

Find the pairs of people who are connected to a common person through the “knows” relation
<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Description</th>
<th>Owner</th>
<th>Status</th>
<th>Last Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Financial Demo Dataset</td>
<td>Financial Demo Dataset</td>
<td>ogcs-dev-user</td>
<td>Tue, 16 Oct 2018 20:18:04 GMT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>slcail607</td>
<td>Sales Demo Dataset</td>
<td>ogcs-dev-user</td>
<td>Fri, 12 Oct 2018 21:30:25 GMT</td>
<td></td>
</tr>
</tbody>
</table>
Summary
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 connectedness

• 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 Hadoop platform or Oracle Database

• Available both on-premise or in the Cloud already today
More information

  - White papers, software downloads, documentation and videos
- Hands On Lab included in `/opt/oracle/oracle-spatial-graph/`
  - Content also available on GITHUB under http://github.com/oracle/BigDataLite/
- Blog – examples, tips & tricks: [blogs.oracle.com/bigdataspatialgraph](http://blogs.oracle.com/bigdataspatialgraph)
- @OracleBigData, @SpatialHannes, @JeanIhm
- Oracle Spatial and Graph Group
Introduction to Graph analytics

Youtube videos

• What is Oracle Big Data Spatial and Graph?
  https://youtu.be/t9pJJhzZKOE

How can graph analytics help my business?
  https://youtu.be/0dJNzBi7B-k

Detecting anomalies with Oracle Big Data Spatial and Graph
  https://youtu.be/nfP6HDOIImjY

Generating recommendations with Oracle Big Data Spatial and Graph
  https://youtu.be/9LRlF3of-Hs
Integrated Cloud
Applications & Platform Services