

Spatial and Graph Summit @

ANALYTICS AND DATA SUMMIT 2019

All Analytics. All Data. No Nonsense. March 12-14, 2019

Introduction to Graph Analytics and Oracle Cloud Service

Hans Viehmann Product Manager EMEA

Oracle

🥗 @SpatialHannes

Jean Ihm Product Manager US

Oracle

🔰 @JeanIhm

Analytics and Data Summit 2019

Spatial and Graph Sessions

- 25+ Spatial and Graph related sessions
 - See yellow track on agenda
 - Room 103 for most sessions
- Tuesday:
 - Morning: Graph technical sessions
 - Afternoon: Spatial technical sessions, Graph hands on lab
- Wednesday:
 - Morning: Spatial use cases
 - Afternoon: Graph use cases & Spatial sessions for developers
- Thursday:
 - Morning: Graph tech sessions & use cases (RDF & property graph)
 - Afternoon: Spatial analytics & big data focus





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

- Product Introduction
- ² Use Cases
- ³ Feature Overview





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 <u>due to the need to ask complex questions across</u> <u>complex data</u>, which is not always practical or even possible at scale using SQL queries.

Source: Gartner press release, 2/18/2019, www.gartner.com/en/newsroom/press-releases/2019-02-18-gartneridentifies-top-10-data-and-analytics-technolo





Oracle's Spatial and Graph Strategy

Oracle Database

Spatial and Graph Option

Enabling Spatial and Graph use cases on every platform

Oracle Big Data Spatial and Graph



Big Data Appliance Commodity Hadoop Spark

Cloud Services



Database Cloud Service Exadata Cloud Service

Copyright © 2019, Oracle and/or its affiliates. All rights reserved.

Two Graph Data Models

Property Graph Model

Social Network Analysis

• Path Analytics

- Social Network Analysis
- Entity analytics



- Financial
- Retail, Marketing
- Social Media
- Smart Manufacturing

Linked Data Knowledge Graphs

RDF Data Model

- Data federation
- Knowledge representation



- Life Sciences
- Health Care
- Publishing
- Finance

Use Case

ORACLE

Graph Model

Industry Domain

RDF Knowledge Graph Sessions: Technical overview, use cases on Knowledge Graphs and BIM for Engineering (Bechtel) – Thursday 8:45 – 10:40am

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
- Supported both on BDA and commodity hardware
 - Cloudera Distribution for Hadoop
- 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 and above (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







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 Applicance, 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



Paysafe:

Paysafe

- Providing online payment solutions
 - Real-time payments, e-Wallets
 - 1bn revenue/yr
 - 500000 payments/day
- Strong demand for fraud detection
 - Only feasible with graph data
 - In real-time, upon money movement
 - During account creation
 - In investigation, visualizing payment flows
- Storing payments in database
 - Refreshing graph using delta update



Graph use case sessions at AnD Summit '19

Wednesday

- Using Graph Analysis for Fraud Detection in Fintech at Paysafe S. Dalekova/Y. Ivanov, Paysafe – 1:00pm
- Building Consistent Crime Investigation Practices Using Big Data and Graph Technologies – D. Belchior/F. Ferreira, Rio Public Prosecutor's Office – 2:20pm
- $-\operatorname{Room} 103$



Feature Overview



Oracle Graph Analytics Architecture





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
 - JShell, Python, Javascript, ...
 - Apache Zeppelin integration
- Graphical UIs

- Property Graph Visualization component (forthcoming), Cytoscape, plug-in available
- Commercial Tools such as TomSawyer Perspectives









Example: Betweenness Centrality in Big Data Graph

Code 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



ORACLE

. . .

More on PGQL...

PGQL: A Query Language for Property Graphs – O. van Rest, Oracle – Wednesday, 3:25pm



Basic graph pattern matching

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



Basic graph pattern matching

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



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 (+)





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 (+)

+---+
| x.name | y.name | ancestor.name |
+---+
Peter	Retta	Paul
Peter	Dwight	Paul
Peter	Dwight	Retta
+---++

Result set





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

Zeppelin Notebook - Interpreter	Search	in your notebooks	Q	Connected	
Reachability					
Dur Green-Marl program will populate it. Then we will run some c value for it.	ode to query thi	s property and return	the graph node	s with a high	
<pre>// create a new property named 'count' count = graph.createVertexProperty(PropertyType.INTEGER, "cound")</pre>	nt");				
==> Vertex Property named 'count' of type integer belonging to	graph flight				
Now we are ready to run our Green-Marl program against the gra	iph:		Run thi (Sh	s paragraph ift+Enter)	
ORACLE Oracle Labs Data Studio				୍ Search ୍ ୧ ar	dmin@or
Copy of OOW/OOW 0	> C 🖬				4
blish Notebook 🔿	2 2 4 6 B		L	ayout 🗊 🛛 Default Templa	ate 🥒
Download Files			Build Graph		
	Graph Stats	aph == null) { .readGraphWithProperties('/va	r/shared/btc.json',	graphName) ▷ κ ³	1
land	lanari			25	53 ms @ 1
Bogdi SELECT COUNT(*) AS Vertices FROM COW, graph4 MATCH (v) Vertice		Vb9d > AS Edges > a ² im FROM COM_graph4 > A > A MATCH () -(e) >> () > A > A			 <i>≫</i>
					 Title
Y611068	Luges	251089			
364735	751089				Code
364735 Page 1 of 1 (1 of 1 items) K < [1] > ×	751089 Page 1 of	1 (1 of 1 items) K (1)	к		Code Resul Settin In IFra
364735 Page 1 of 1 (1 of 1 items) K < (1) > × Transactions per Bitcoin Address	751089 Page 1 of	1 (1 of 1 items) K (1)	Э Pagerank T	əp 6	 ✓ Code ✓ Resul Settin ✓ In IFra
384735 Page 1 of 1 (tof 1 items) × くて > × Transactions per Bitcoin Address 副 回 ③ 区 ビ ゆ ム 目 永 〇 田 田 凶 ② 本 占 * <i>争</i>	751089 Page 1 of	1 (1 of 1 items) K < 1 →	Pagerank T	op 6	 ✓ Code ✓ Resul Settir ✓ In IFra
364735 Page 1 of 1 (1 of 1 items) × くこう メ Transactions per Bitcoin Address 回回回回にはのム日本の回回にの 本 ふ* の	Page 1 of	1 (1 of 1 items) K < (] > ID 166682	Я Pagerank T value 0.0018036832049004	op 6	 ✓ Code ✓ Resul Settir ✓ In IFra
364735 Page 1 of 1 (1 of 1 items) × < 1 > × Transactions per Bitcoin Address 回 回 回 ビ ゆ ム 田 水 〇 田 地 凸 章	751089 Page 1 of	1 (1 of 1 items) K < 1 > ID 166682 287831	я Pagerank T value 0.0015036832049004 0.00159175341436224	op 6 724 94	 ✓ Code ✓ Resul Settir ✓ In IFra
364735 Page 1 of 1 (1 of 1 items) × < (1) × Transactions per Bitcoin Address	751089 Page 1 of	t (1 of 1 items) K < (1) > ID 166682 287831 247741	Pagerank T Value 0.0018036832049004 0.00159175341436224 0.0014478296328467	op 6 724 94 109	 ✓ Code ✓ Resul □ Settir ✓ In IFra
364735 Page 1 of 1 (1 of 1 items) K < (□) × Transactions per Bitcoin Address □	751089 Page 1 of	1 (1 of 1 items) × < 1 → ID 166682 287831 247741 150550	Pagerank T value 0.0018036832049004 0.00159175341436224 0.0014478296328467 0.0014050903831874	op 6 724 94 109 185	 ✓ Code ✓ Resul Settir ✓ In IFr
364735 Page 1 of 1 (1 of 1 items) K < 1 > × Transactions per Bitcoin Address © @ @ @ @ A II % ○ III № 0 [A III %] * * * * * * * * * * * * * *	751089 Page 1 of	1 (1 of 1 items) × < 1 → ID 166682 287831 247741 150550 181060	Pagerank T Value 0.0018036832049004 0.00159175341436224 0.0014478296328467 0.001406903831874;	op 6 724 94 209 :85 53	 ✓ Code ✓ Resul Settir ✓ In IFri

Demo



_	nation and the second second						
Datasource	es × +					-	
$\leftrightarrow \ \rightarrow \ G$	← → C 🛈 Not secure slc14rui:32000/graphstudio/?root=ds-datasources 🛛 🗢 🔍 🏠 🙆 🖓 🚱 😳 👘 🕲 📴 🍓) 🚳 🗄
						୍ Search ୧୦୦୦	jcs- <mark>dev-user</mark> 🔻
Ē₁ Data	Sources						Create 🔻
					Search	Sort by:	Ŧ
Туре	Name	Description	Owner	Status		Last Applied	
	Financial Demo Dataset	Financial Demo Dataset	ogcs-dev-user	Tue, 16 Oct 2018 20:18:	04 GMT		=
	slcai607	Sales Demo Dataset	ogcs-dev-use <mark>r</mark>	Fri, 12 Oct 2018 21:30:2	5 GMT		≡
			_				



Oracle Graph Analytics Architecture







Support for Graph Pattern Matching





Path Query (Parallel Recursive With)

PGQL:

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

Find the pairs of people who are connected to a common person through the "knows" relation

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",
       T3.T AS "s2.fname$T",T3.V AS "s2.fname$V",T3.VN AS "s2.fname$VN",T3.VT AS "s2.fname$VT"
FROM (/*Path[*/SELECT DISTINCT SVID, DVID FROM ( SELECT VID AS SVID, VID AS DVID FROM "GRAPH1VT$" UNION ALL SELECT SVID, DVID
      FROM (WITH RW (ROOT, SVID, DVID, LVL) AS ( SELECT ROOT, SVID, DVID, LVL FROM (SELECT SVID ROOT, SVID, DVID, 1 LVL
      FROM (SELECT T0.SVID AS SVID, T0.DVID AS DVID FROM "GRAPH1GT$" 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$" T1 WHERE (T1.EL = n'knows')) R, RW WHERE RW.DVID = R.SVID )
      CYCLE SVID SET cycle col TO 1 DEFAULT Ø SELECT ROOT SVID, DVID FROM RW ))/*]Path*/) T6,
     (/*Path[*/SELECT DISTINCT SVID, DVID FROM ( SELECT VID AS SVID, VID AS DVID FROM "GRAPH1VT$" UNION ALL SELECT SVID, DVID
      FROM (WITH RW (ROOT, SVID, DVID, LVL) AS ( SELECT ROOT, SVID, DVID, 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
WHERE T2.K=n'fname' AND T3.K=n'fname' AND T6.SVID=T2.VID AND T6.DVID=T7.DVID AND T7.SVID=T3.VID
ORDER BY T6.SVID ASC NULLS LAST, T7.SVID ASC NULLS LAST
```

Combining Graph Analytics and Machine Learning

Graph Analytics

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



Machine Learning

Machine learning session

 When Graphs Meet Machine Learning – S. Hong/R. Patra, Oracle – Thursday, 10:55am



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 vs. GraphX

- 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

Property Graph Visualization Tool (new)

- Lightweight visualization component
- Single-Page Web Application based on Oracle JET and D3.js
- Takes PGQL Query as input, renders result set visually
- Will support PGQL-to-PGX initially, but can work with anything that supports PGQL (including PGQL-to-SQL)





Graph Visualization – Commercial Tools



See Tom Sawyer vis demo: Partner lightning round – Tuesday 12:00pm (Auditorium), Finding Malicious Network Packets Using Anomaly Detection with Graph Analytics – Thursday, 12:00pm

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 interactive queries
 - Java
 - PGQL
 - Rest APIs

ORACLE

- Rich User Interface
 - Low code / zero code features
 - Notebook support and powerful data visualization features

Session: Graph Cloud Preview: How to Analyze Data Warehouse Data as a Graph – K. Schmid/J. Sharma, Oracle – Tuesday 11:15am





Summary Graph capabilities in Oracle Database and 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 Database or Big Data Platform
- Available both on-premise or in the Cloud



Graph sessions at AnD Summit '19

• Tuesday

- Graph Cloud Preview: How to Analyze Data Warehouse Data as a Graph K. Schimd/J.
 Sharma, Oracle–11:15am
- Hands On Lab: Introduction to Property Graphs in Oracle Databases K. Hare, JCC Consulting – 3:35pm room 202
- Wednesday
 - Using Graph Analysis for Fraud Detection in Fintech at Paysafe S. Dalekova/Y. Ivanov, Paysafe – 1:00pm
 - Building Consistent Crime Investigation Practices Using Big Data and Graph Technologies – D. Belchior/F. Ferreira, Rio Public Prosecutor's Office – 2:20pm
 - PGQL: A Query Language for Property Graphs O. van Rest, Oracle 3:25pm
 - Translating Natural Language to Graph Queries for Financial Crime Investigation M.
 Brantner, Oracle 4:30pm

Graph sessions at AnD Summit '19 (cont.)

Thursday

- Build Knowledge Graphs with Oracle RDF to Extract More Value from Your Data S.
 Das/M. Perry/M. Annamalai, Oracle 8:45am
- Semantic Middleware the Cornerstone of Your Next IT S. Gabler, Semantic Web Company – 9:50am
- Oracle Spatial and Graph RDF Semantic Model for BIM Classification & Scheduling T.
 McLane, Bechtel 10:15am
- When Graphs Meet Machine Learning S. Hong/R. Patra, Oracle 10:55am
- Finding Malicious Network Packets Using Anomaly Detection with Graph Analytics S.
 Hong, Oracle 12:00pm
- I know what you mean: leveraging graph for linking entities into knowledge base S.
 Hong, Oracle 3:40pm room 202

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 online tinyurl.com/oraclespatialcommunity









oraclespatialsig@gmail.com

Engage with the Spatial and Graph SIG



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

Members connect and exchange knowledge via online communities and at conferences and events Talk with us at the Summit! Look for badges with yellow ribbons

Birds of a Feather Lunch Wednesday 12-1pm Auditorium	Receptions Tues & Wed evenings
--	-----------------------------------

• Join us online

- (IOUG)
- tinyurl.com/oraclespatialcommunity

in

- Search for "Oracle Spatial and Graph Community"
- Contact us:

oraclespatialsig@gmail.com







ANALYTICS AND DATA SUMMIT 2019

All Analytics. All Data. No Nonsense. March 12-14, 2019



Analytics and Data Summit 2019