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

ORACLE

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

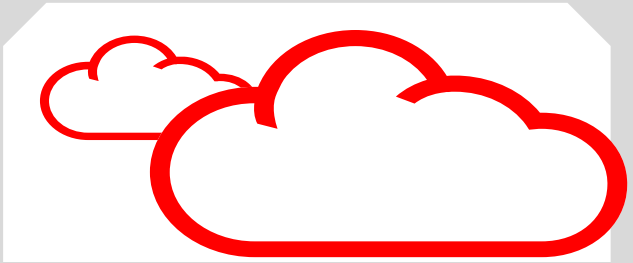
Oracle's Spatial and Graph Strategy

On Premise and Oracle Cloud

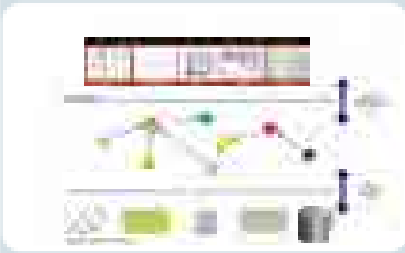
Oracle Database Spatial and Graph

Oracle Big Data Spatial and Graph

Spatial and Graph in Oracle Cloud

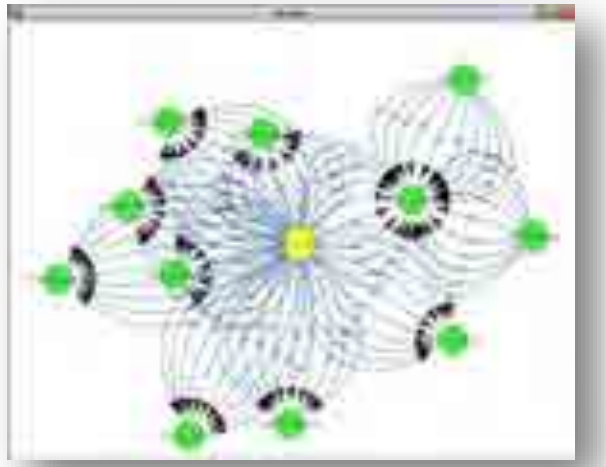


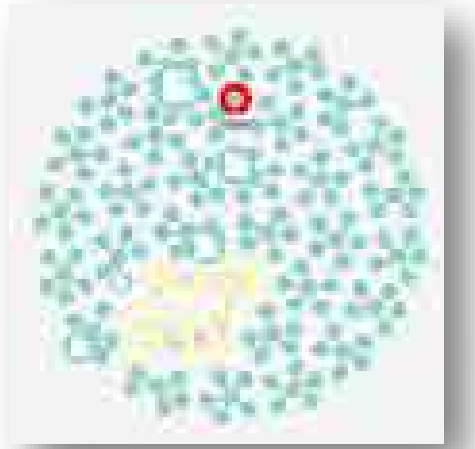
Two Graph Data Models

Use Case	Graph Model	Industry Domain
Social Network Analysis	**Property Graph Model** • Path Analytics • Social Network Analysis • Entity analytics	 ▪ Financial ▪ Retail, Marketing ▪ Social Media ▪ Smart Manufacturing
Linked Data Semantic Web	**RDF Data Model** • Data federation • Knowledge representation • Semantic Web	 ▪ Life Sciences ▪ Health Care ▪ Publishing ▪ Finance

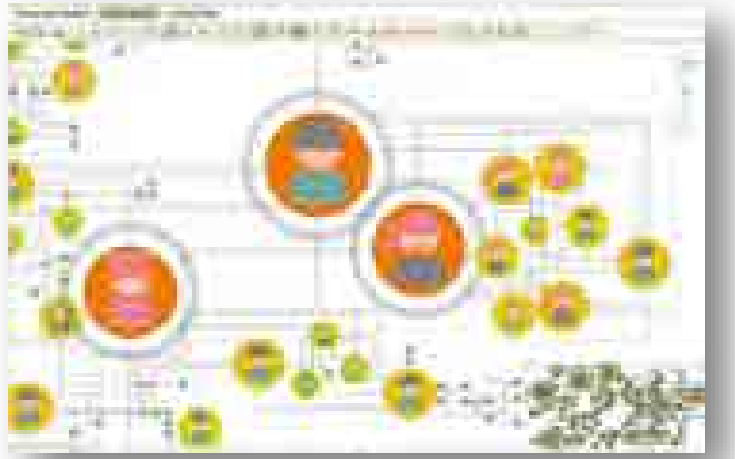
Graph Database Features:

- Scalability and Performance
- Graph analytics
- Graph Visualization
- Graph Query Language
- Standard interfaces
- Integration with Machine Learning tools





Courtesy Linkurious



Courtesy Tom Sawyer Perspectives

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)

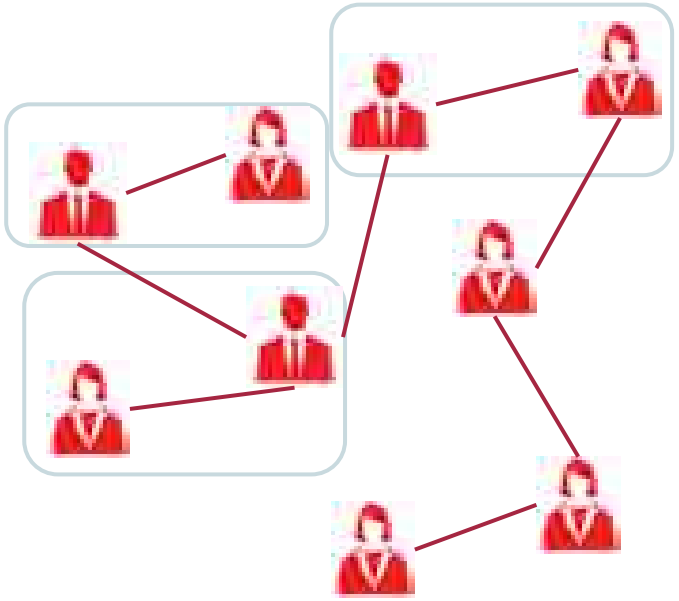
- Available with Oracle 12.2 / DBCS
- Using tables for graph persistence
- Graph views on relational data
- In-database graph analytics
  - Sparsification, shortest path, page rank, triangle counting, WCC, sub graphs
- SQL queries possible
- Included in Database Cloud Service

Use Cases

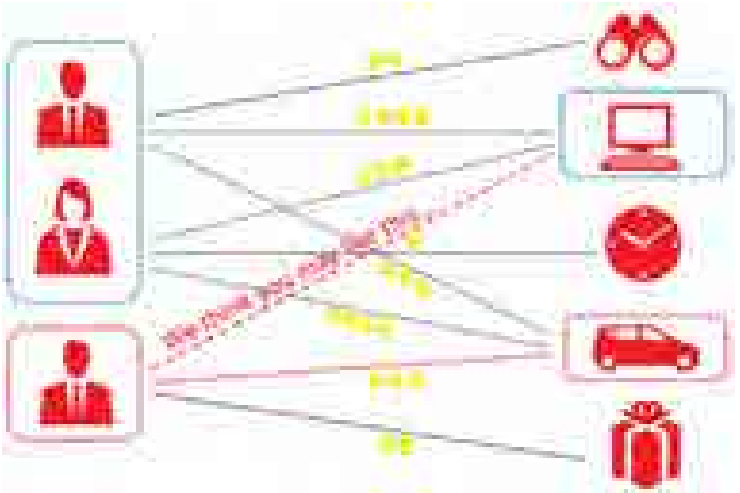
Graph Analysis for Business Insight

Identify Influencers

Discover Graph Patterns in Big Data

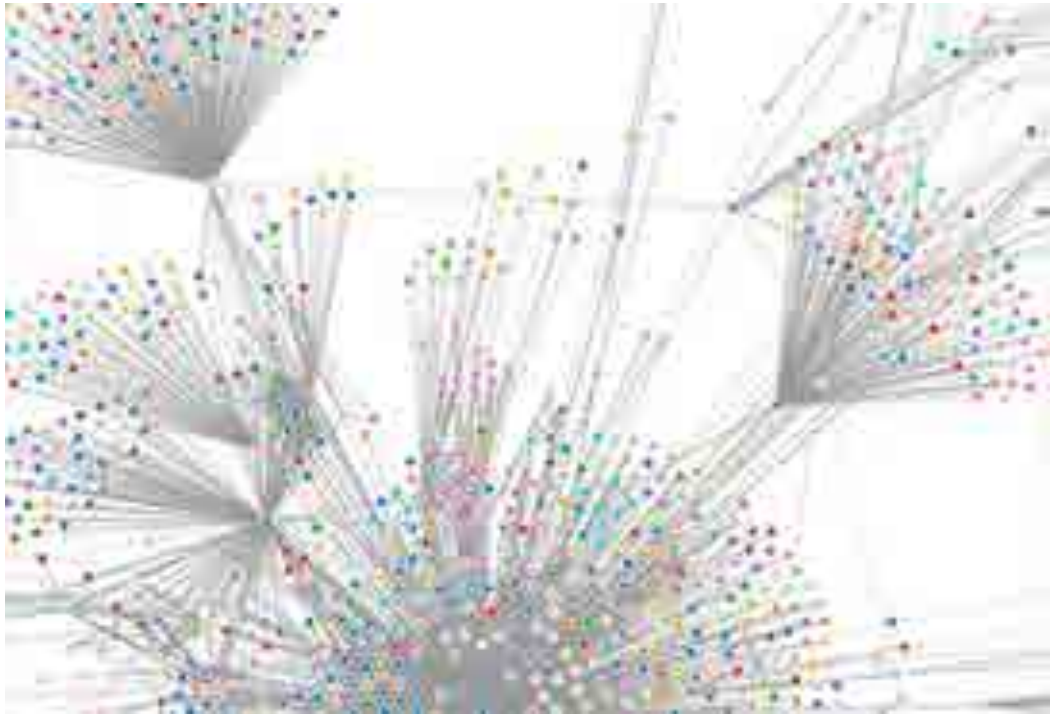


Generate Recommendations



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

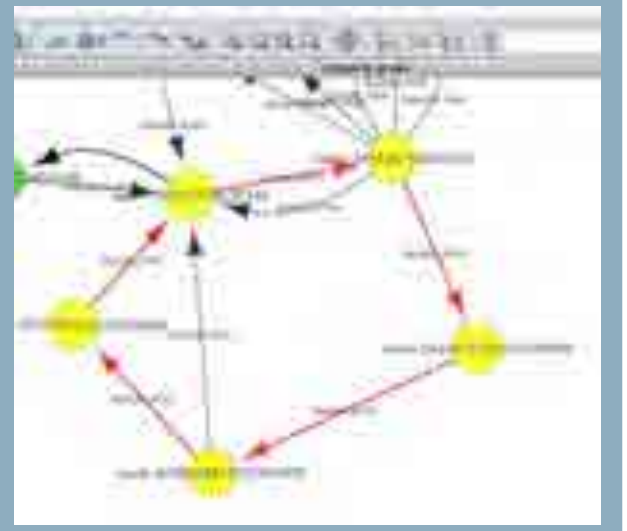
Applying Graph Analysis To Improve Customer Service

- 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.

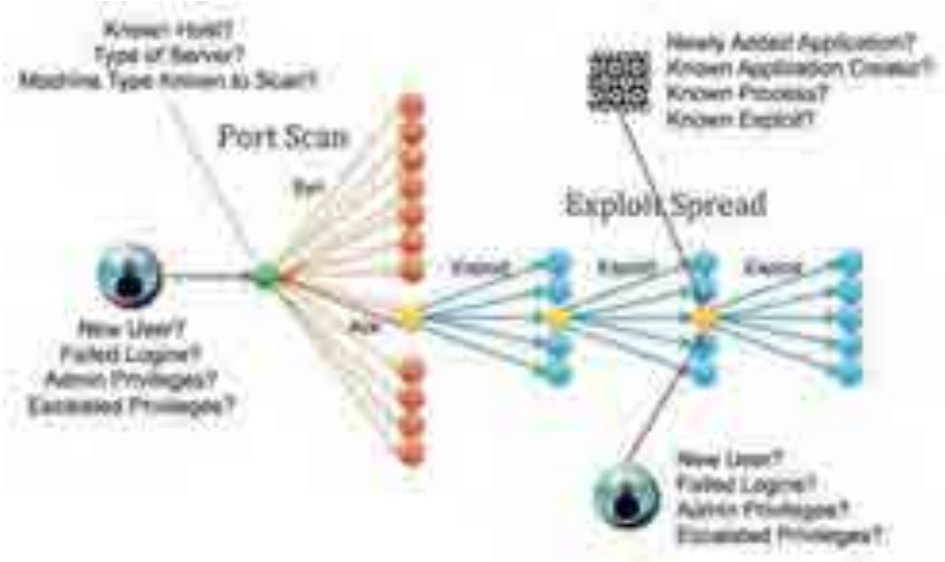
How graph analysis solves the problem

- Search for known individuals in web of content
- Analyze relationship with other criminals, travel history, addresses, employers
- Relationship between suspects and their financial affairs

Courtesy Tom Sawyer Perspectives

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 failture on an IoT system?
  - Reachability analysis: understand which routines, libraries, servers, routers are affected by a modification

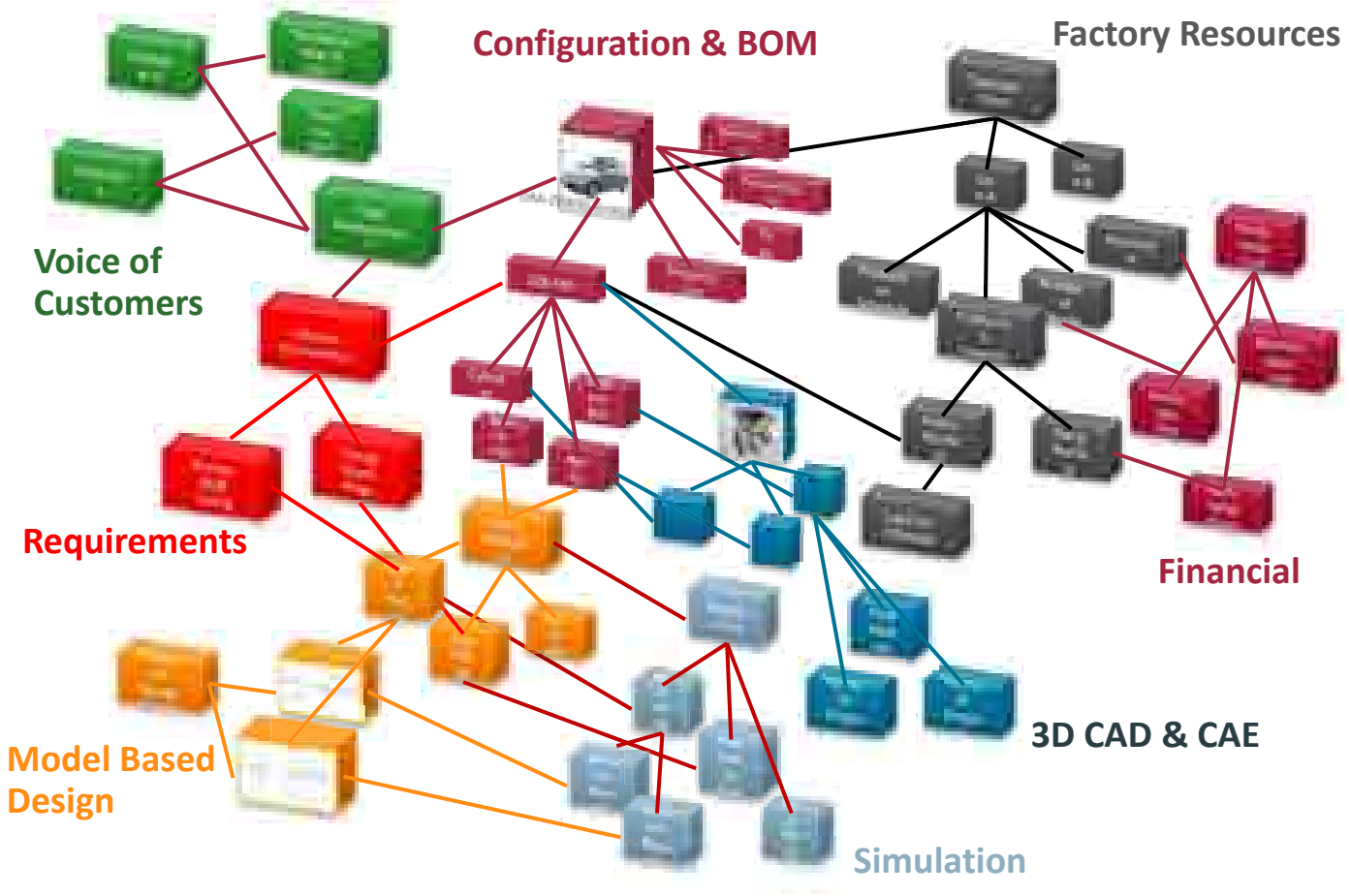


Automotive Manufacturing

Support high variance, short innovation cycles of complex autos

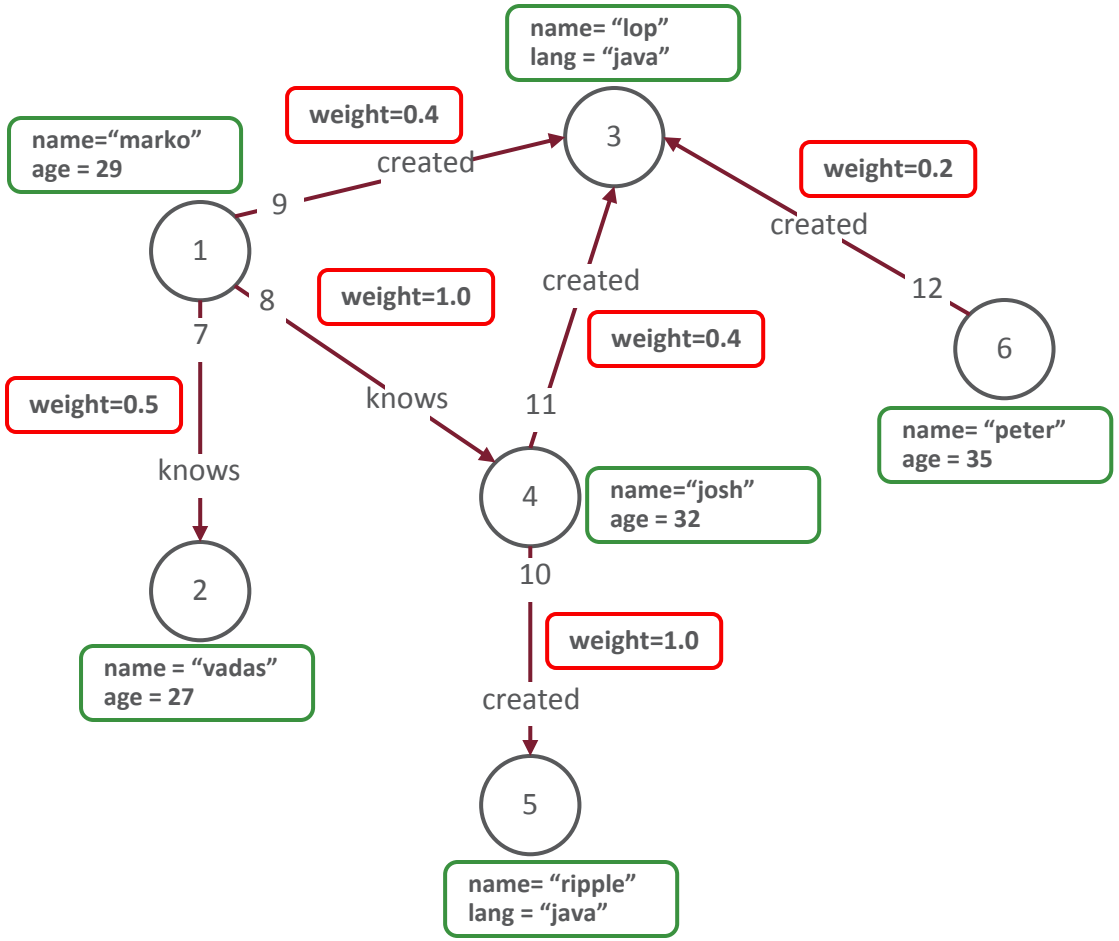
Graph View of Enterprise Data

- 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

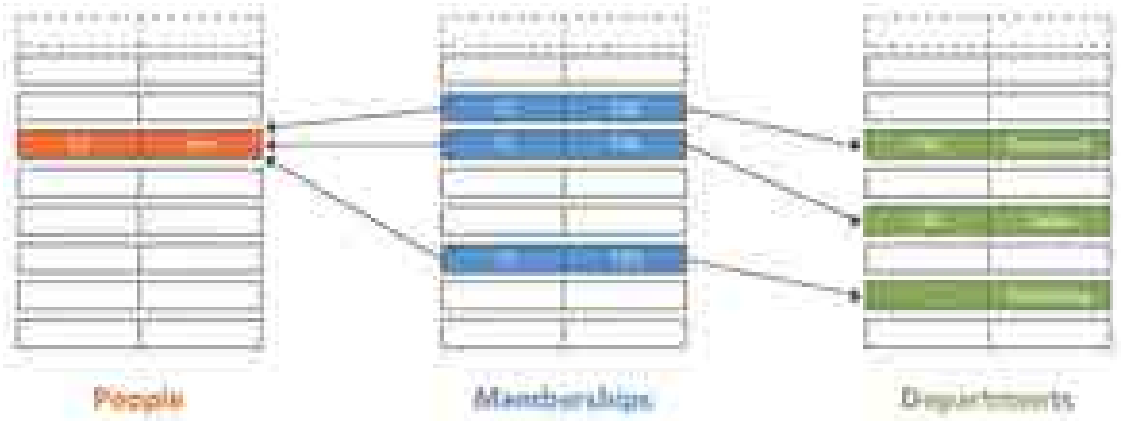


https://github.com/tinkerpop/blueprints/wiki/Property-Graph-Model

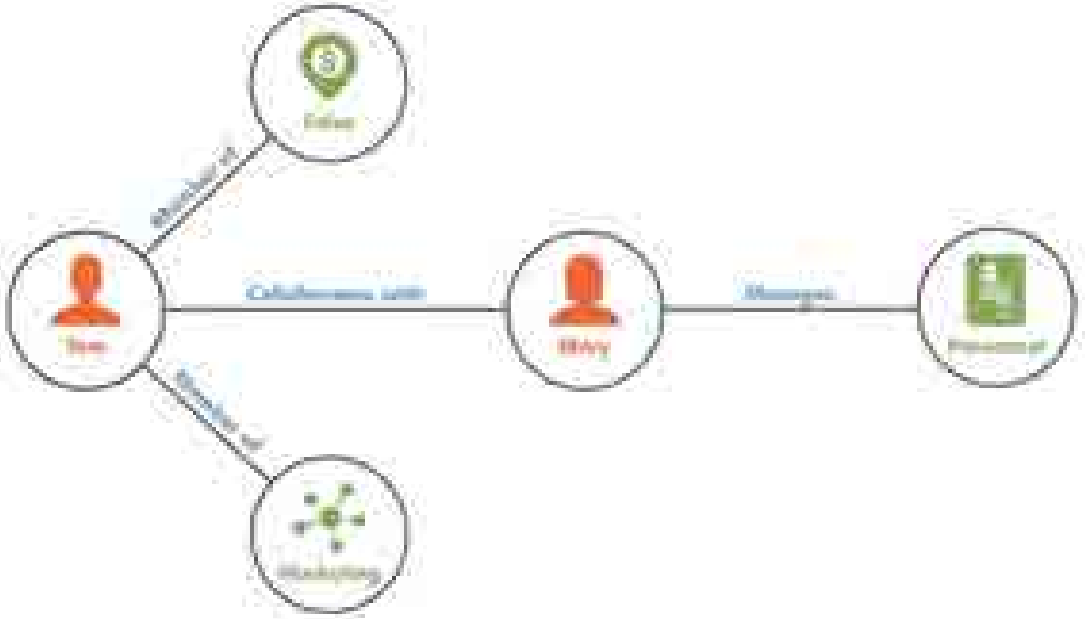
- A set of vertices (or nodes)
  - each vertex has a unique identifier.
  - each vertex has a set of in/out edges.
  - each vertex has a collection of **key-value** properties.
- A set of edges (or links)
  - each edge has a unique identifier.
  - each edge has a head/tail vertex.
  - each edge has a label denoting type of relationship between two vertices.
  - each edge has a collection of **key-value** properties.

Relational Model vs. Graph Model

- Relational Model

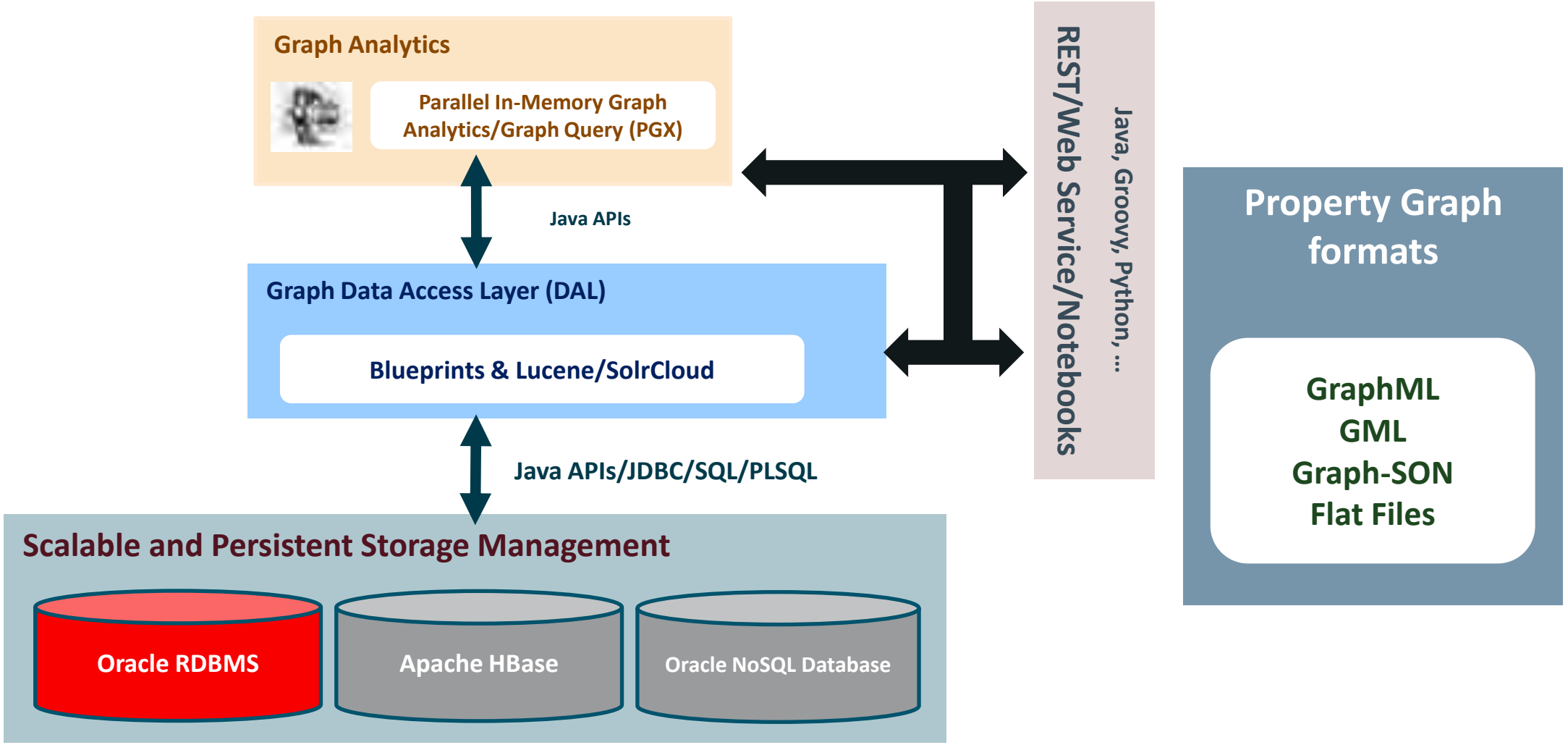


- Graph Model

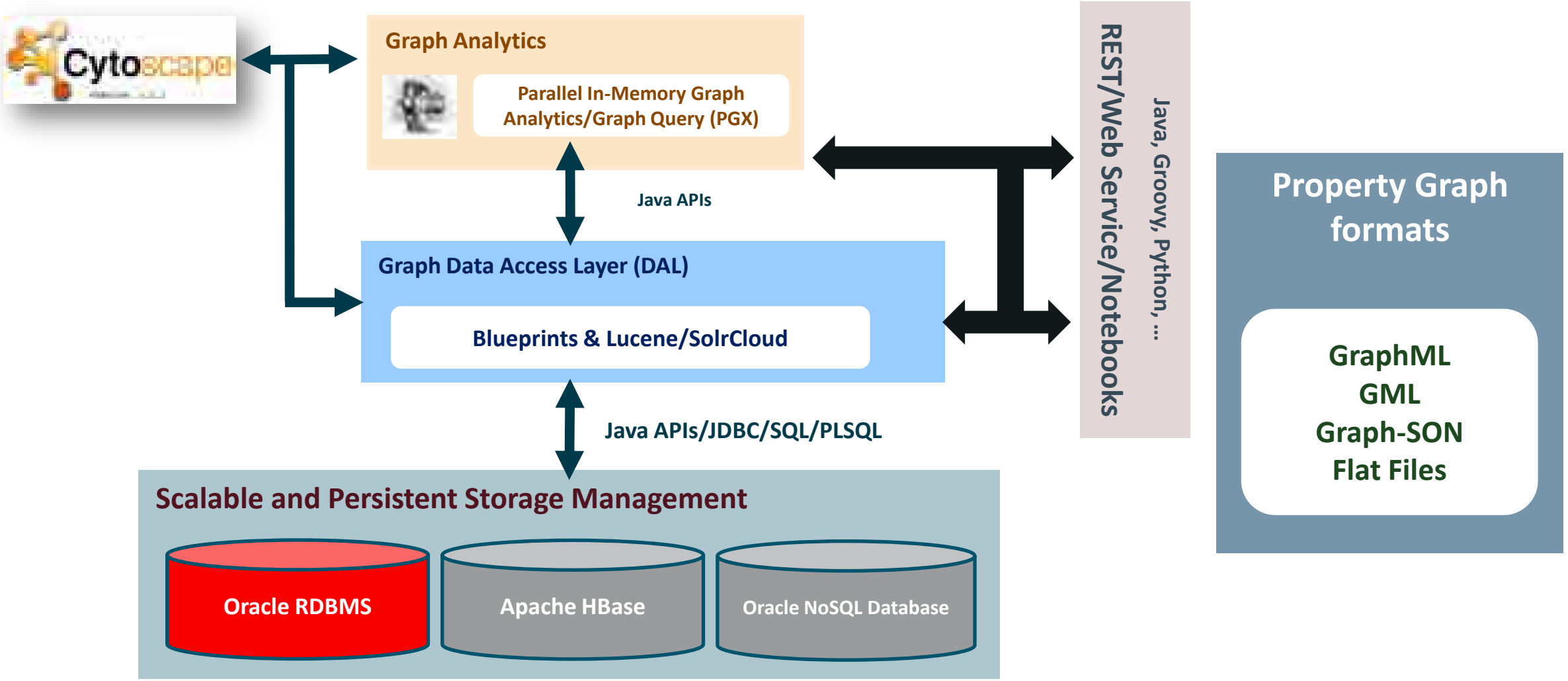


Courtesy: Tom Sawyer 2016

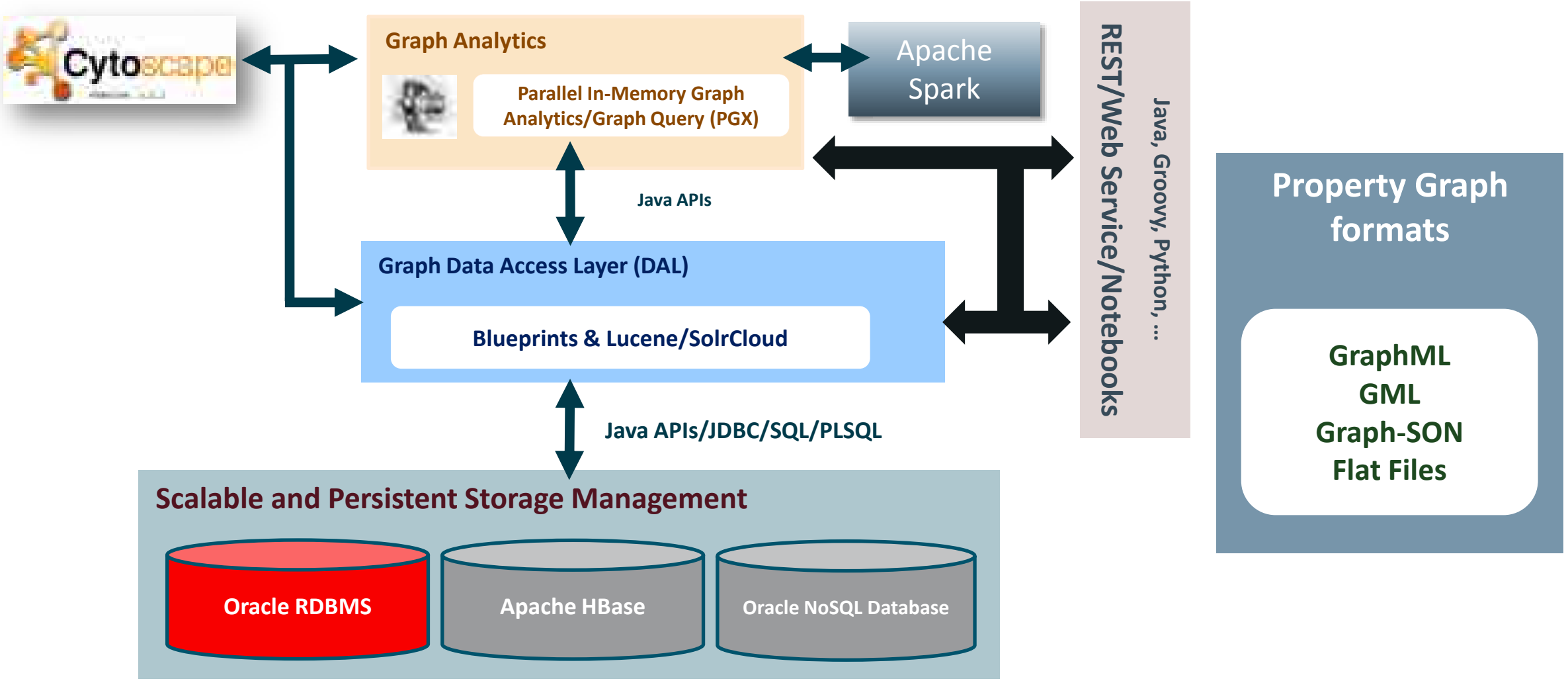
Architecture of Property Graph Support



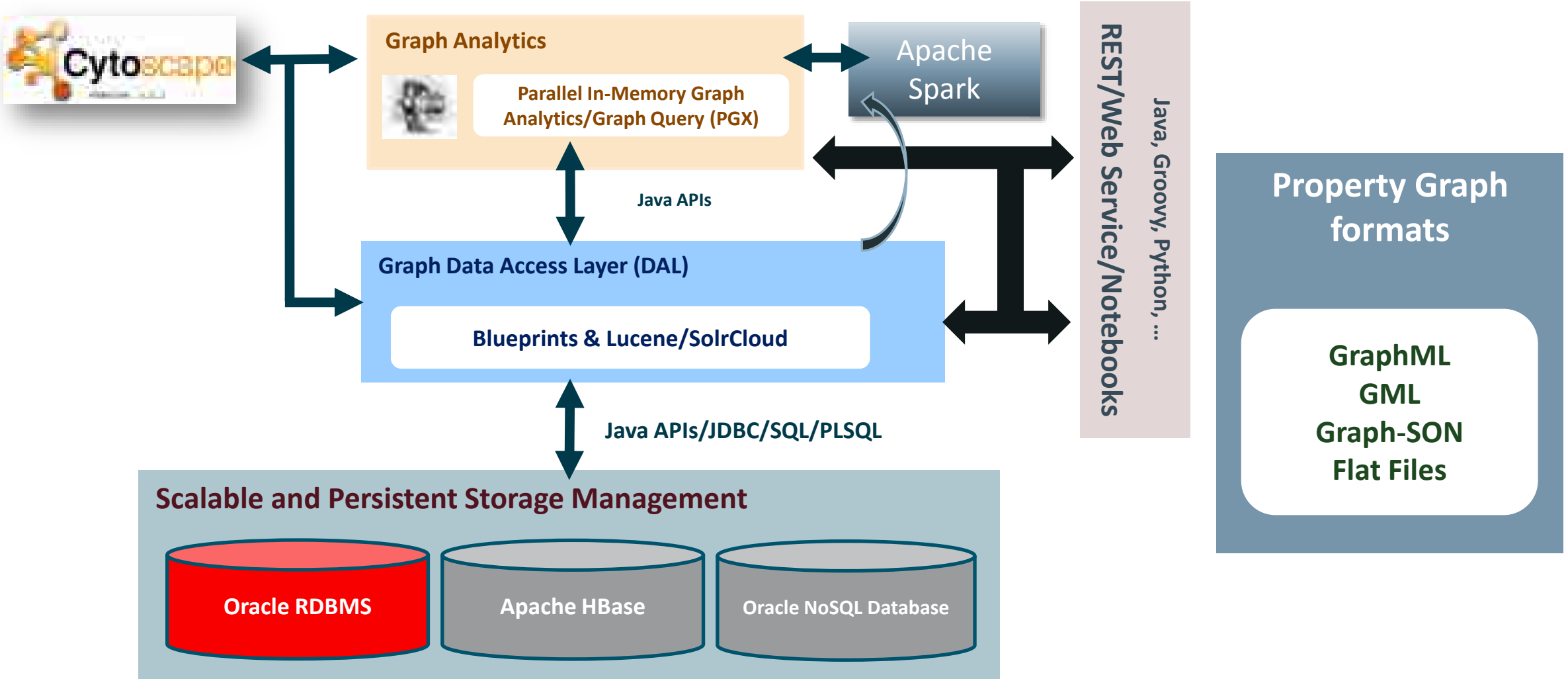
Architecture of Property Graph Support



Architecture of Property Graph Support



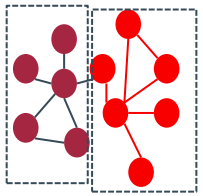
Architecture of Property Graph Support



Computational Analytics: Built-in Package

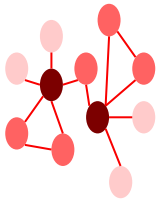
Rich set of built-in parallel graph algorithms

Detecting Components and Communities



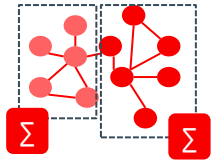
Tarjan's, Kosaraju's, Weakly Connected Components, Label Propagation (w/ variants), Soman and Narang's Spacification

Ranking and Walking



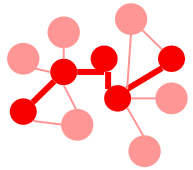
Pagerank, Personalized Pagerank, Betwenness Centrality (w/ variants), Closeness Centrality, Degree Centrality, Eigenvector Centrality, HITS, Random walking and sampling (w/ variants)

Evaluating Community Structures



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

Path-Finding

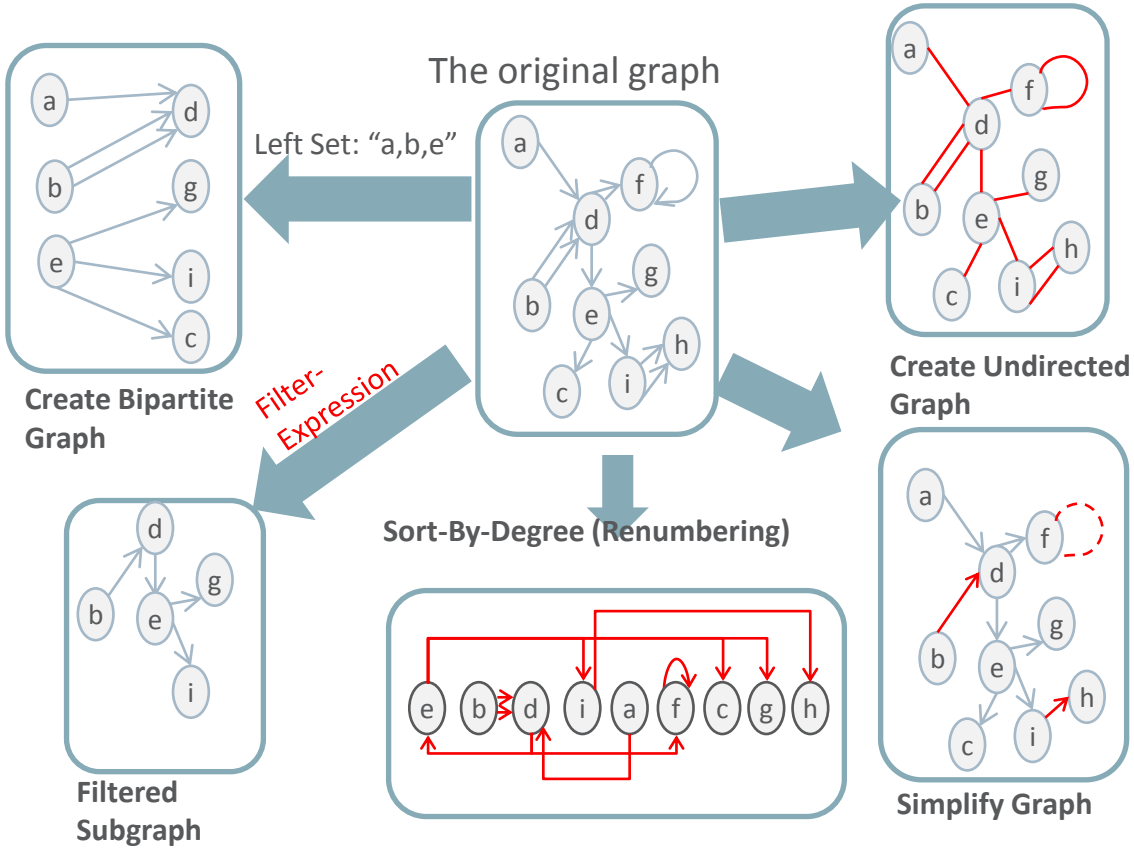


Hop-Distance (BFS) Dijkstra's, Bi-directional Dijkstra's Bellman-Ford's

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

Other Classics Vertex Cover Minimum Spanning-Tree (Prim's)

… and parallel graph mutation operations



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)

```
sum([v.query() \
    .direction(blueprints.Direction.OUT).count() \
    for v in OPGIterator(v0.query() \
    .direction(blueprints.Direction.OUT) \
    .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

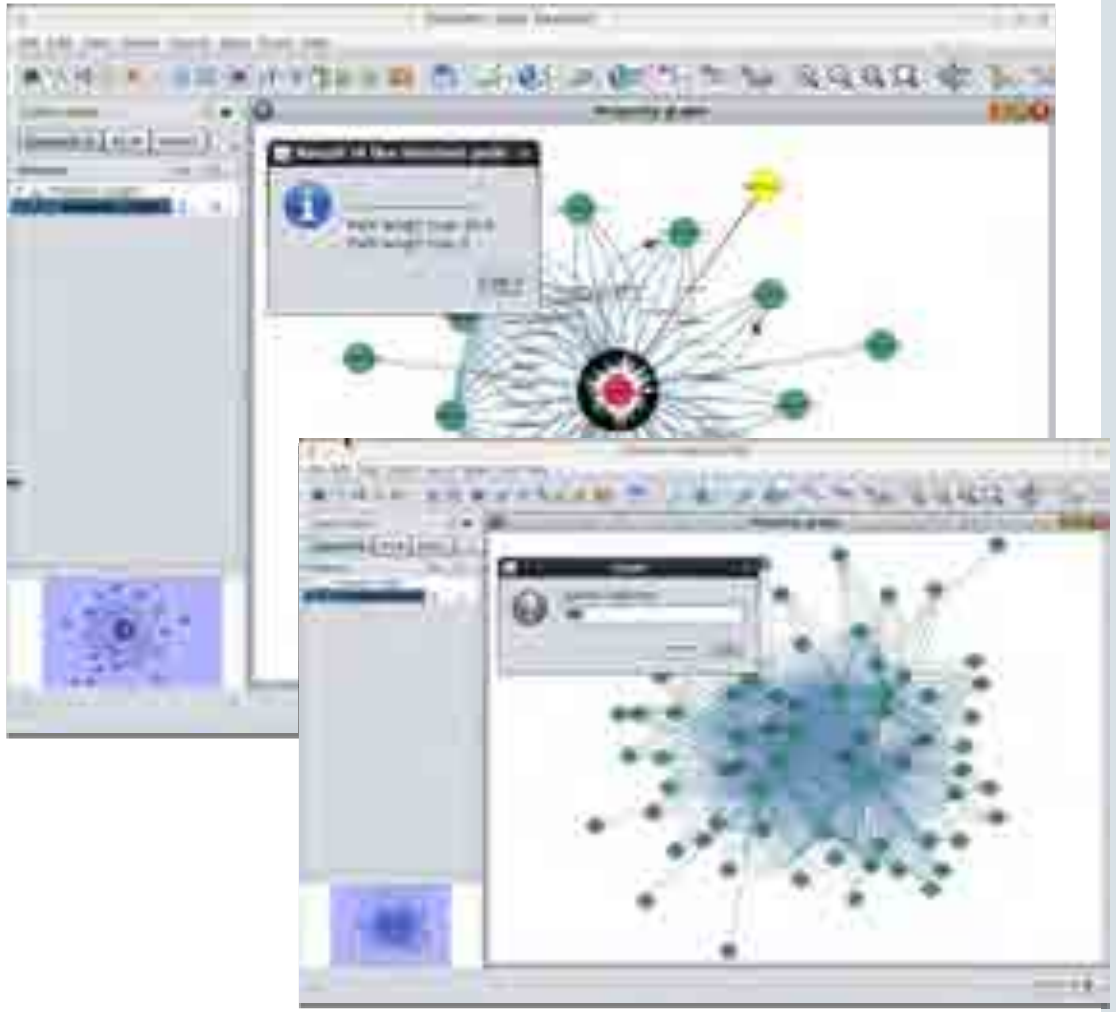
- Automatic update based on a subset of property keys
- Avoid linear scan to access an element by key/value

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

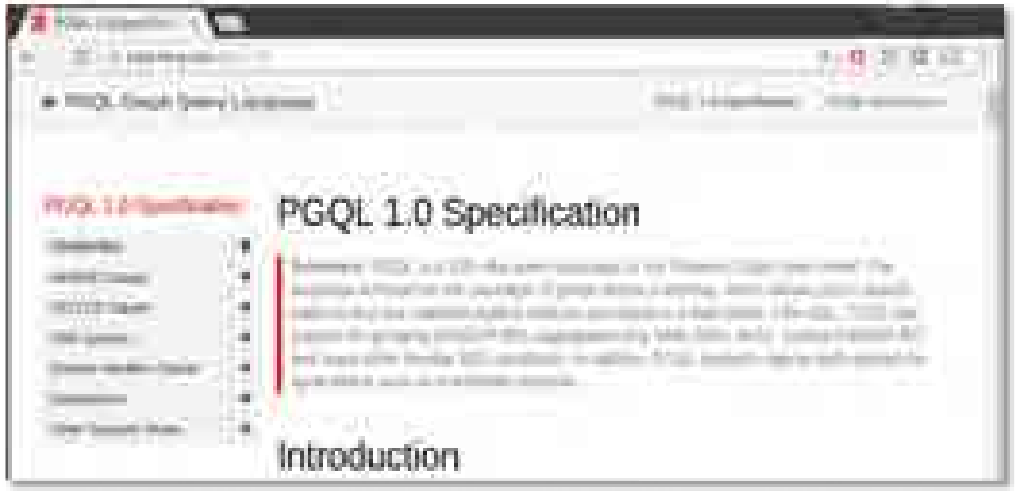
TomSawyer, Cambridge Intelligence, Linkurios, Vis.js,...

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
  - …

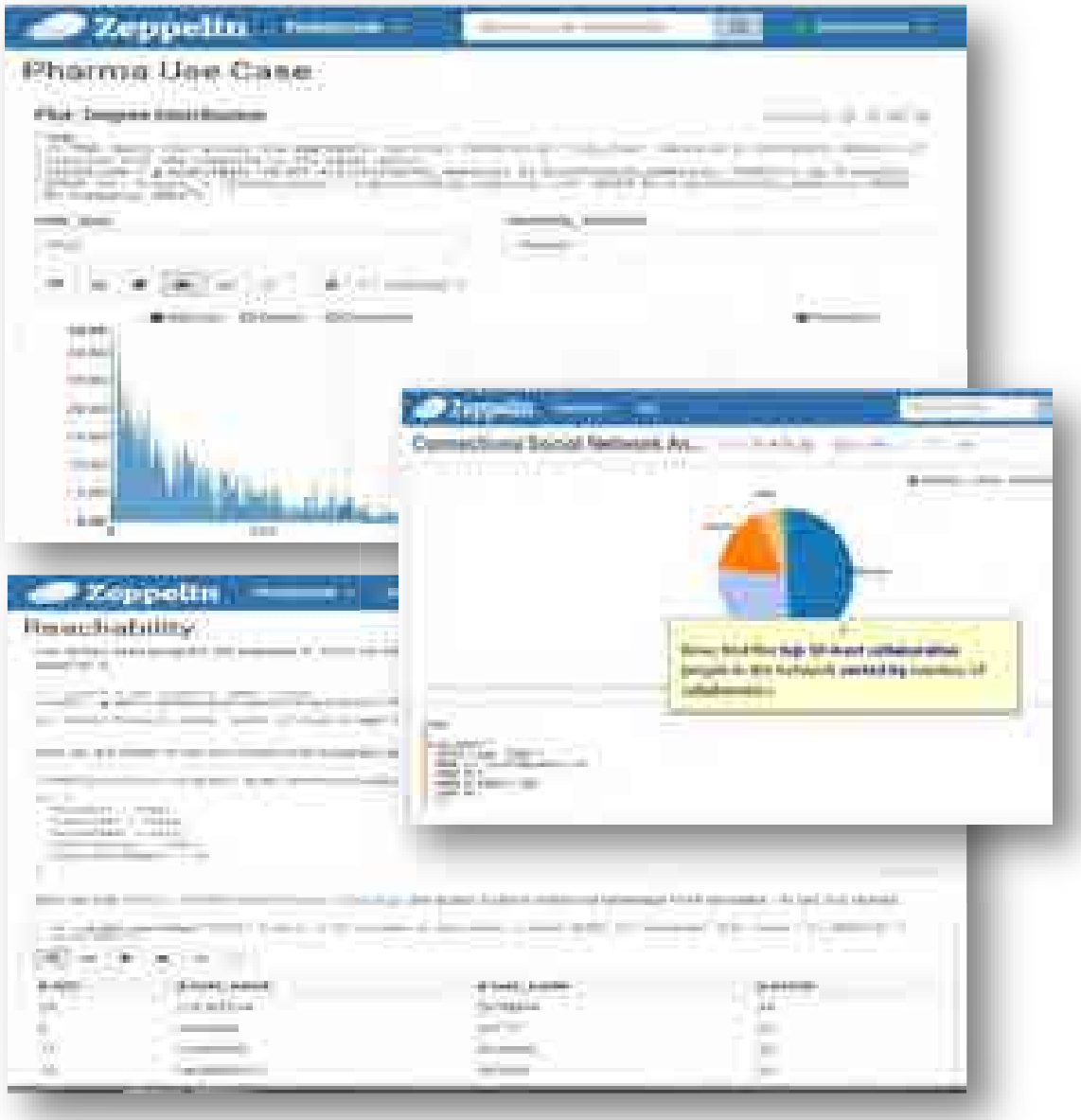
- Proposed for standardization by Oracle
  - Specification available on-line
  - Open-sourced front-end (i.e. parser)

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

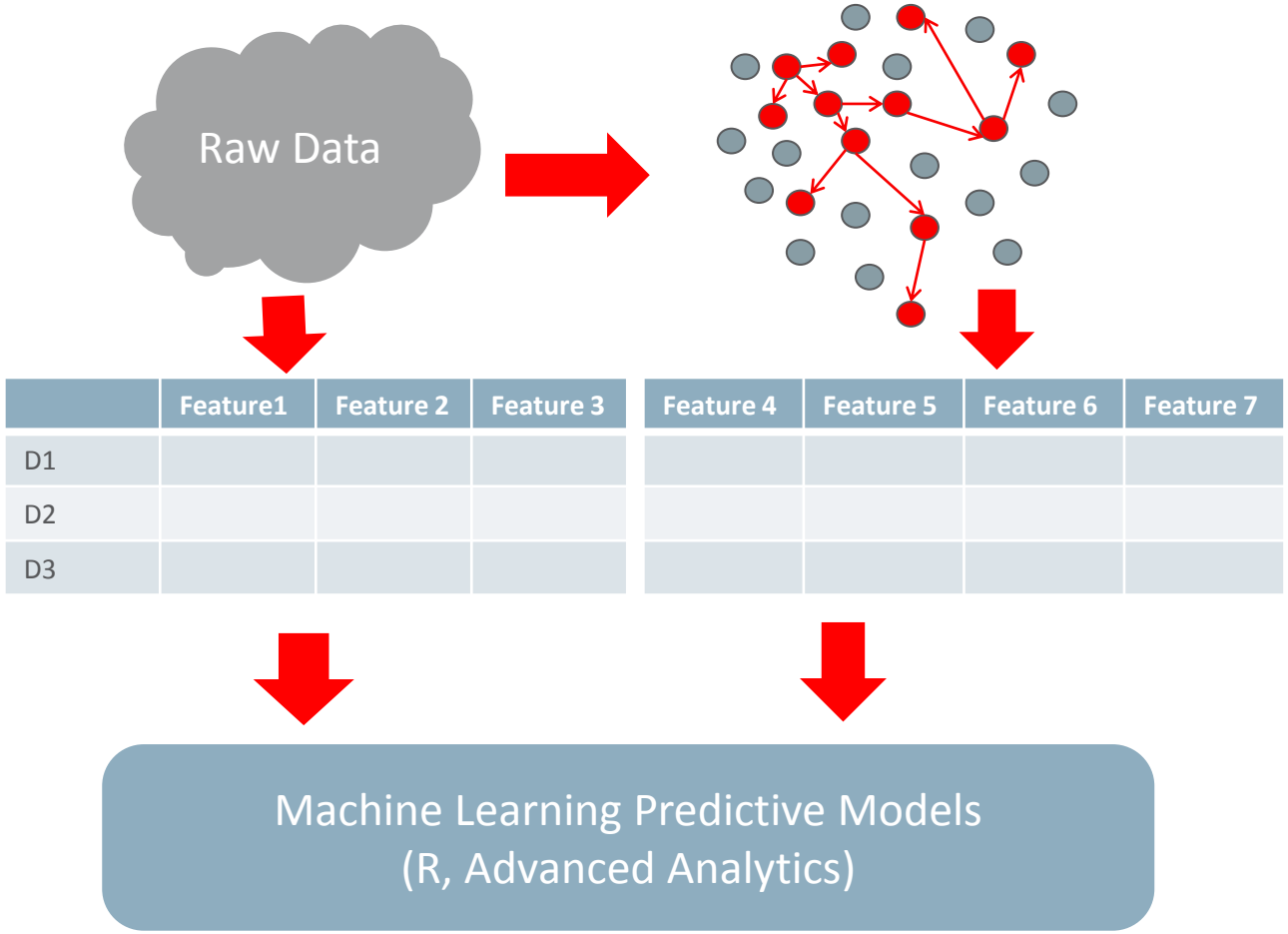
- Access through APIs
  - Implementation of Apache Tinkerpop Blueprints APIs
  - Based on Java, REST plus SolR Cloud/Lucene support for text search
  - SQL/PLSQL for property graph functions in Oracle Database

- 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

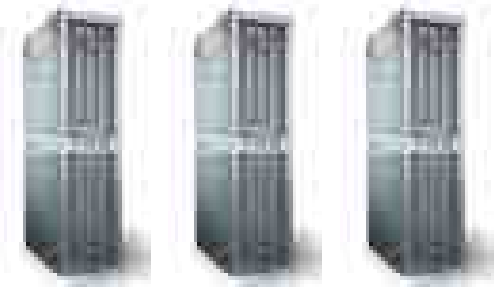
Enhancing ML and Data Analytics with Graphs

- 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



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

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

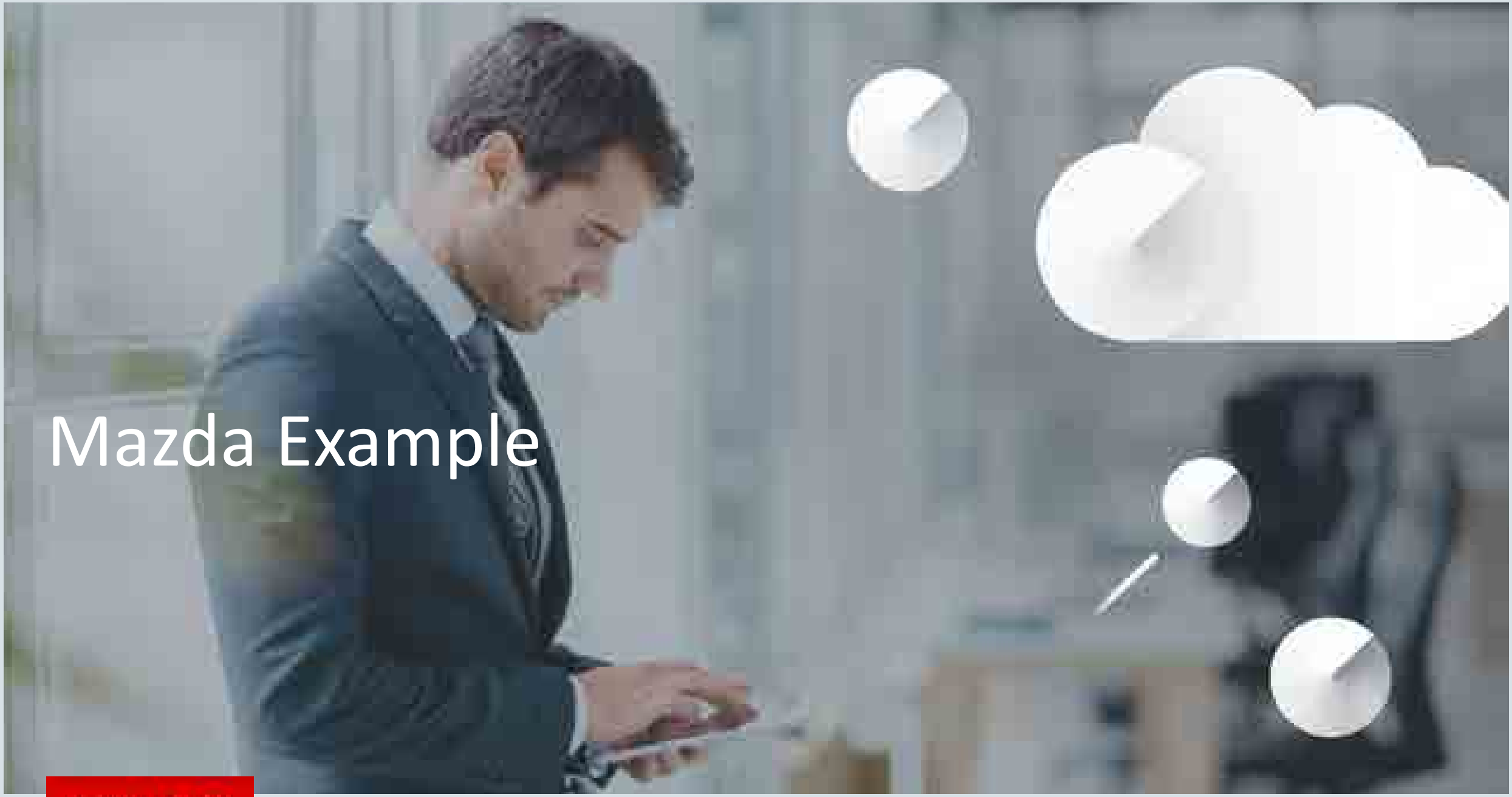
Graph Studio

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

Company name	Mazda Motor Corporation
Founded	January 30, 1920
Headquarters	Hiroshima / Japan
Revenue	$30 Billion (FYE Mar 2017)
Retail Volume	1.5 million units (same FY as above)
Number of employees	48,749 (consolidated) (same FY as above)
R&D center	5 locations (Hiroshima, Yokohama, US, Germany, China)
Production Site	3 factories in Japan Hiroshima Plant (Head Office, Ujina), Hofu Plant (Nishinoura, Nakanoseki), 7 factories overseas China, Thailand, Mexico, Vietnam, Malaysia, Russia

Mazda Plant

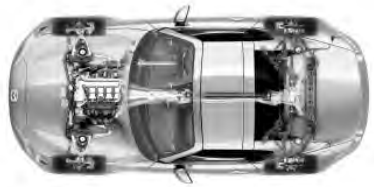
Mazda Plant

Mazda's Problem

Imagine

Auto Manufacturer

Vehicle



Vehicle

Parts (constructed by small parts)

Mazda's Problem
Data Structure

Relational ?

Many Business Domain

- Finance
- Sale / Marketing
- Production
- Bill Of Materials
- …

Graph ?

Which Data Structure is better for each Data ?

Mazda’s PoC

	May 8	15	22	29	June 5	12	19	26	July 3	10	17	24	31	Aug 7	21	28	Sep 4	11	18	25	Oct 2	9	16	23	30	Nov 6	13	20	27	Dec 4	11	18
Ite0																																
Ite1																																
Ite2																																
Ite3																																
Ite4																																
Ite5																																
Ite6																																
Ite7																																

Mazda's PoC (4th Stage)

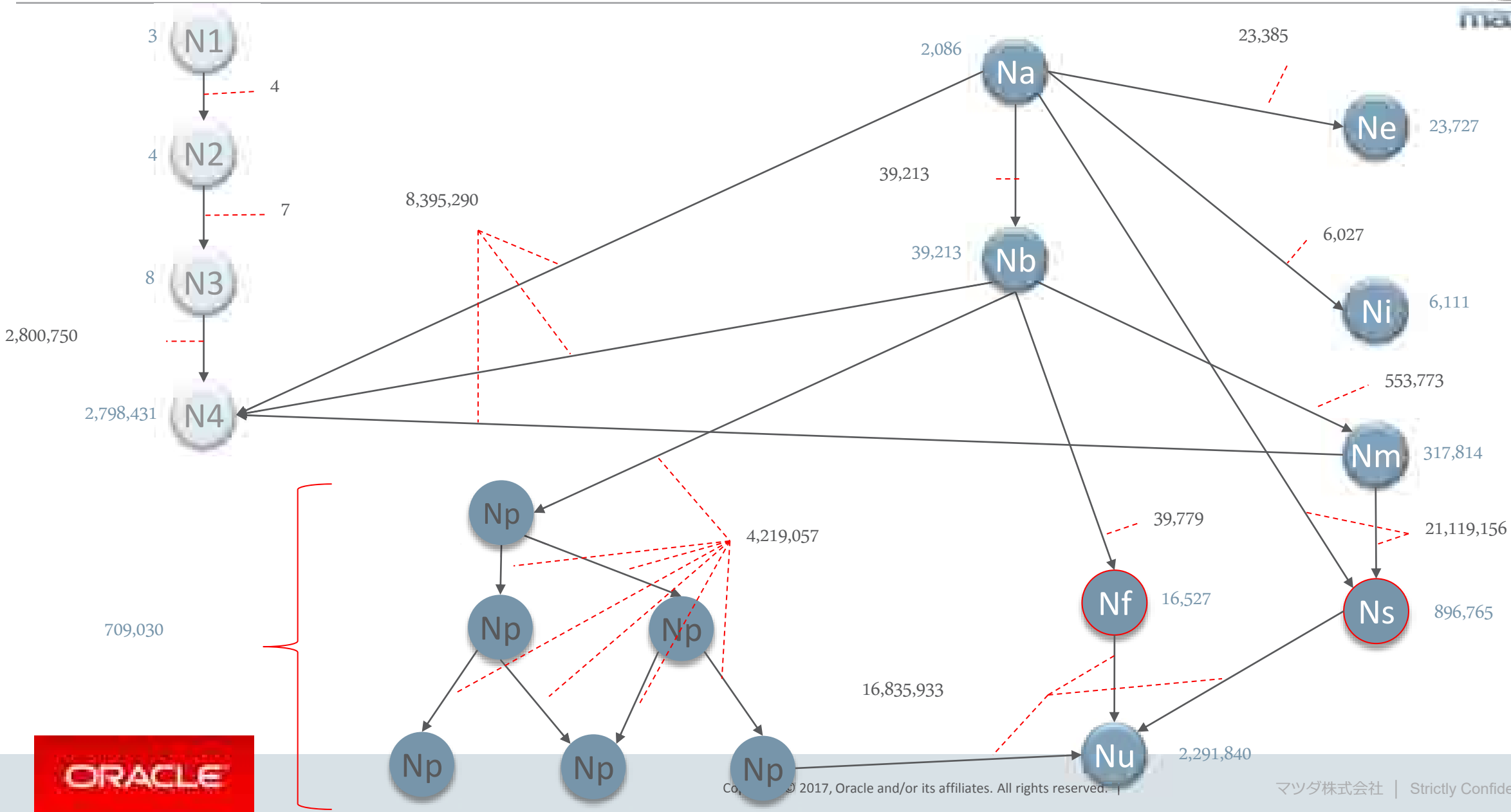
Total number of Edges : 53,993,161

Total number of Nodes : 7,099,473

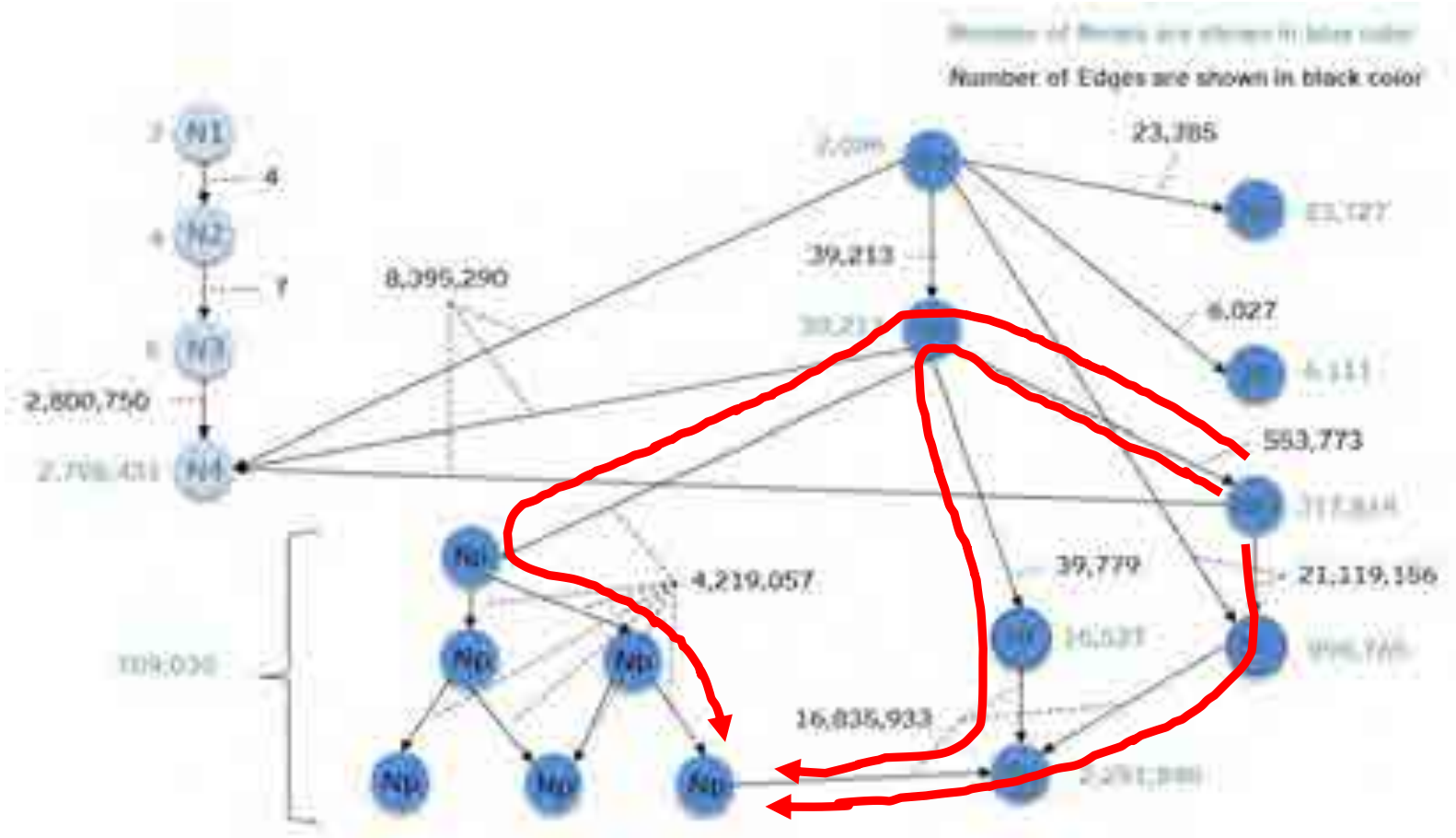
Mazda's PoC (4th Stage)

Number of Nodes are shown in blue color

Number of Edges are shown in black color



Performance (PGQL Query)



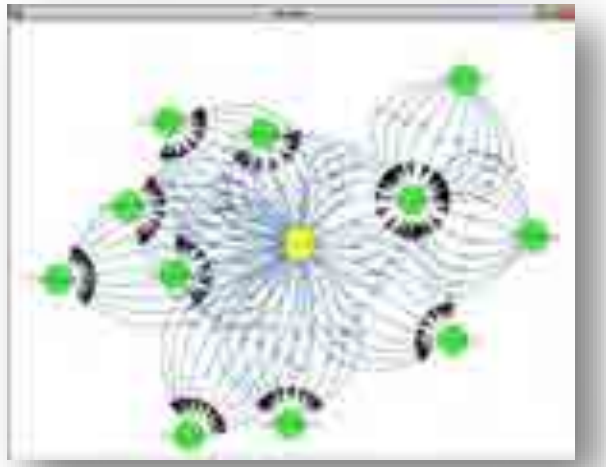
Nm	Num	Query time (ms)
aaaaaaaa	62	43
bbbbbbbb	66	51
cccccccc	78	46

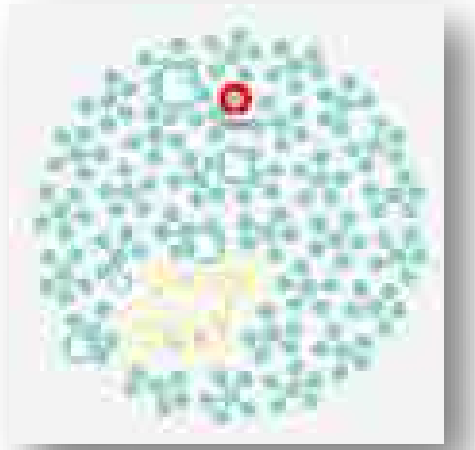
Summary (Current Result)

- Performance is Good !
- Issues: Refinement of complex PGQL queries
- Next Step: On going collaboration with Oracle Team
  - Oracle Japan, US Development, Oracle Labs

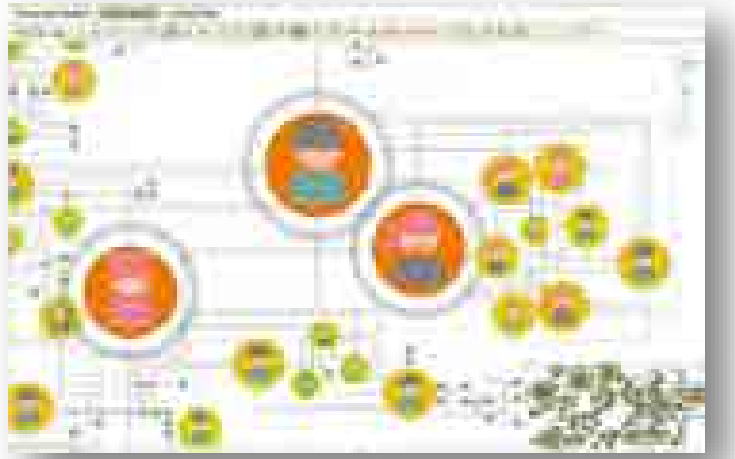
Overview:
Complete Graph Solution

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





Courtesy Linkurious



Courtesy Tom Sawyer Perspectives

Spatial and Graph at OOW 2017

tinyurl.com/SpatialGraphOOW17

Spatial and Graph Sessions

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

Spatial and Graph Demos

Date/Time	Title	Location
Monday - Wednesday	Oracle's Spatial Technologies for Database, Big Data, and the Cloud	Moscone West Exhibit Hall 1st floor Oracle Cloud Platform > Analytics & Big Data, pod SOA 131
Monday - Wednesday	Oracle's Graph Database and Analytics for Database, Big Data, and the Cloud	Moscone West Exhibit Hall 1st floor Oracle Cloud Platform > Analytics & Big Data, pod SOA 127

www.AnalyticsandDataSummit.org

Call for speakers is now open with rolling acceptances.


ORACLE