ORACLE

Build Recommendation Systems Using a Graph Database

Ryota Yamanaka and Melli Annamalai, Product Management, Oracle
Caroline Chan, System Development, CAGLA

July 2, 2020

AskTOM Office Hours: Graph Database and Analytics

- Welcome to our AskTOM Graph Office Hours series!
  We're back with new product updates, use cases, demos and technical tips
  https://asktom.oracle.com/pls/apex/asktom.search?oh=3084
- Sessions will be held about once a month
- **Subscribe** at the page above for updates on upcoming session topics & dates
  And submit feedback, questions, topic requests, and view past session recordings
- Note: **Spatial** now has a new Office Hours series for location analysis & mapping features in Oracle Database:
  https://asktom.oracle.com/pls/apex/asktom.search?oh=7761

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.

Agenda

1. Recap - Graph Analytics — Melli
2. Create a Graph from Tables — Melli
3. Compute Recommendations — Ryota
4. Build a Web Application — Caroline

Melli

Nashua, New Hampshire, USA
@AnnamalaiMelli

Ryota

Bangkok, Thailand
@ryotaymnk

Caroline

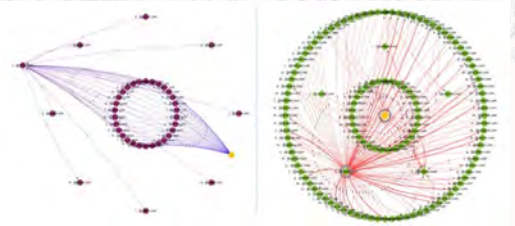
Toyota City, Japan,

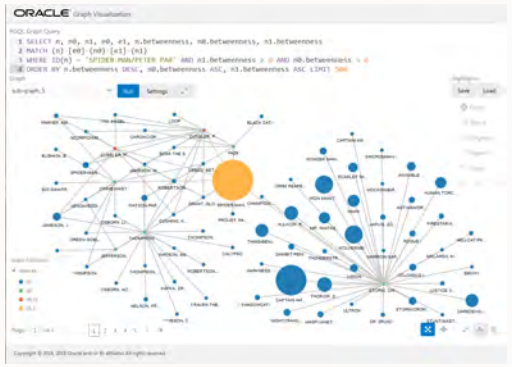
Recap - Graph Analytics

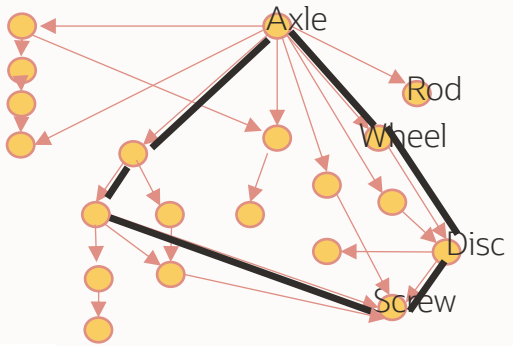
Graph Applications

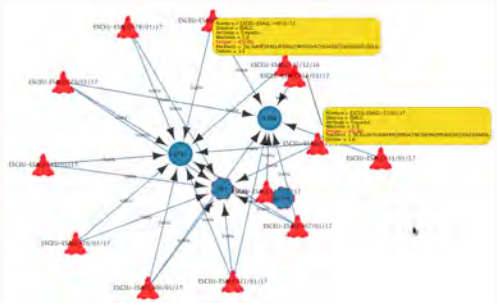
- Financial
- Law enforcement and security
- Manufacturing
- Public sector
- Pharma

and more









Graph Database and Analytics

Store, manage, query, and analyze graphs

- Enterprise capabilities: Built on Oracle infrastructure
- Manageability, fine-grained security, high availability, integration

Highly scalable

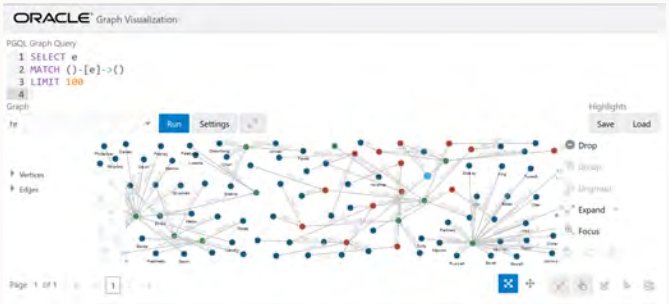
- In-memory query and analytics and in-database query
- 10s of billions of edges and vertices

PGQL: Powerful SQL-like graph query language

Analytics Java API: 50+ pre-built graph analysis algorithms

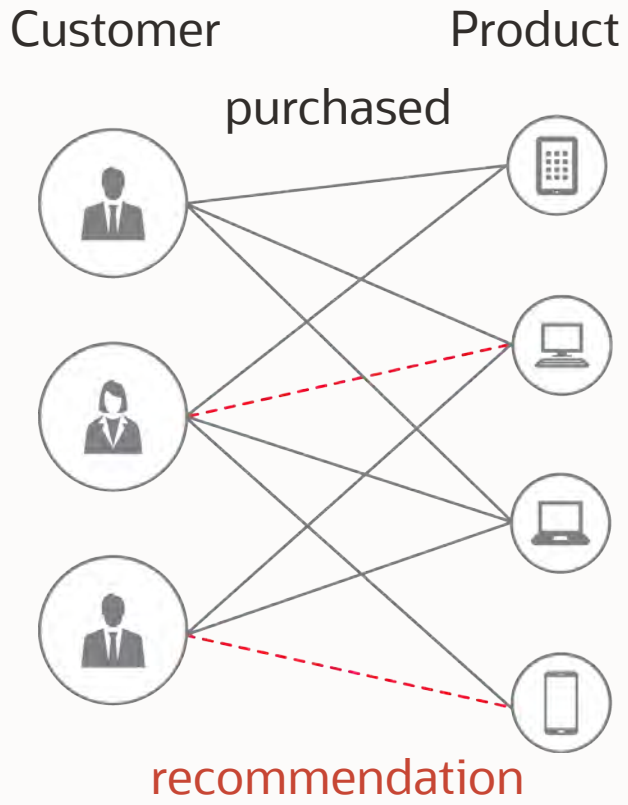
Visualization

- Light-weight web application, UI accessible from a browser



Graph Analytics and Recommendation

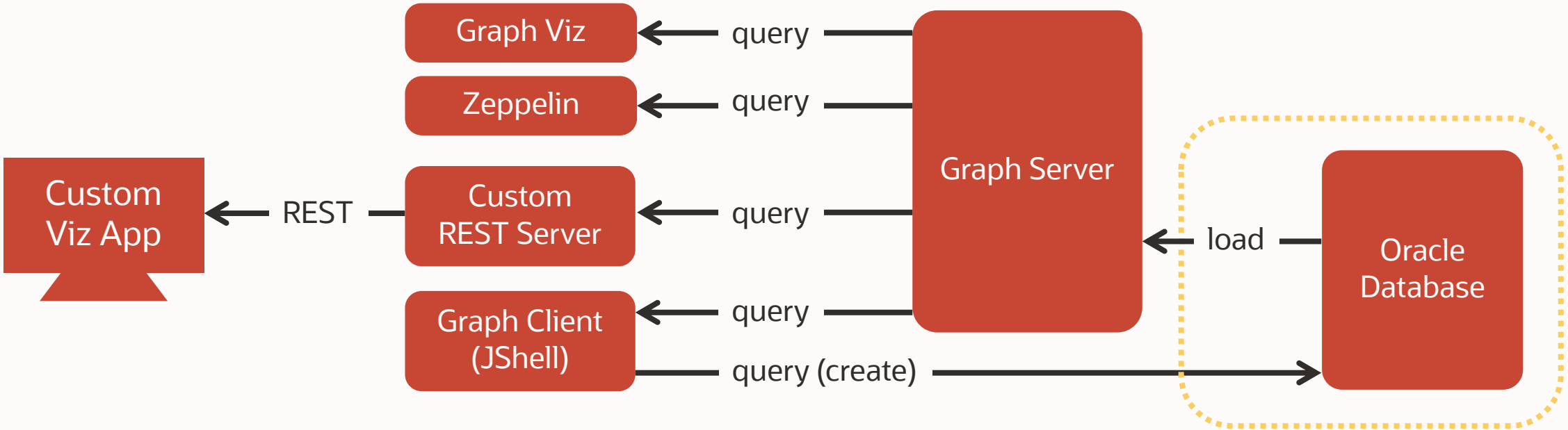
- Recommendation is a major A.I. problem:
  - Product recommendation in e-commerce
  - Optimized real-time adds in mobile apps
  - Personalized content in education
  - And more…
- Challenges
  - Precision in recommendation
  - Apply algorithms to existing data
  - High performance
  - Include connections between data in analysis
  - Integrate data from multiple sources



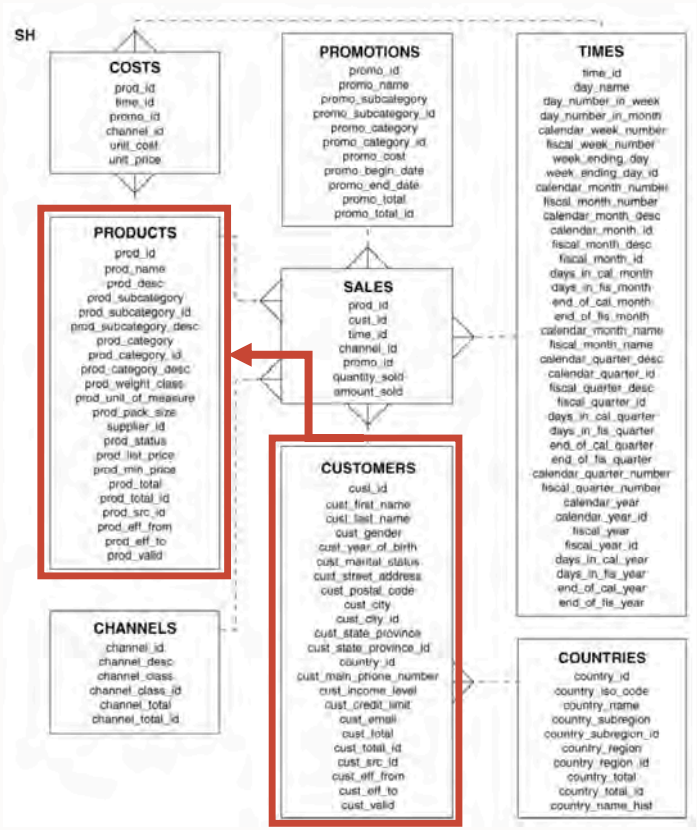
Creating a Graph Model

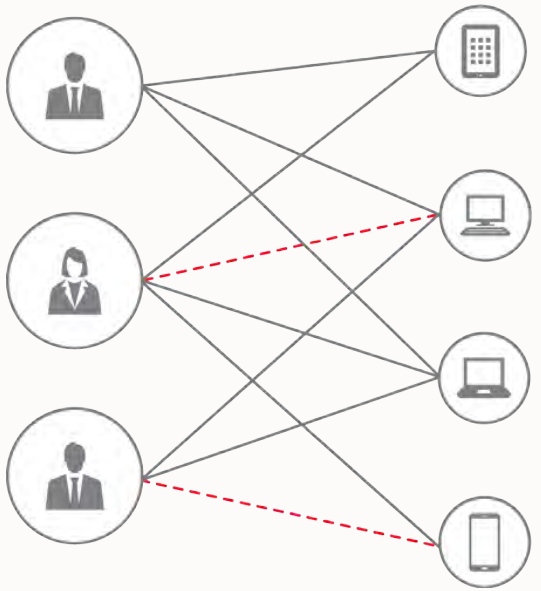
How to Create Graph from Relational

- Source data is often stored in relational database in table forma
- In Oracle Database, tables can be transformed into graphs using CREATE PROPERTY GRAPH query



How to Create Graph from Relational



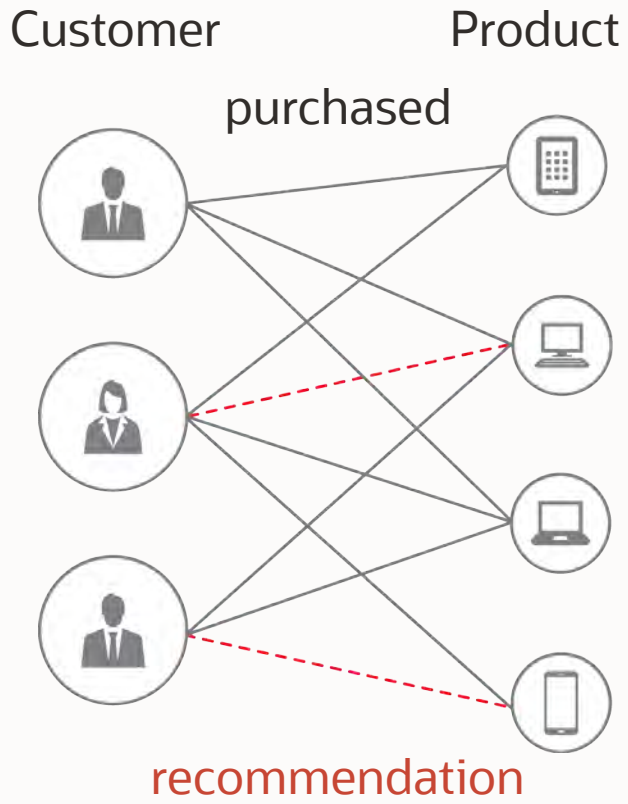


Step 1 - Define Graph Model According to Use Cases

Graph models are defined according to the use cases.

For example, when you need to analyze the product purchase activity of your customers and generate recommendation, your graph should contain the following information:

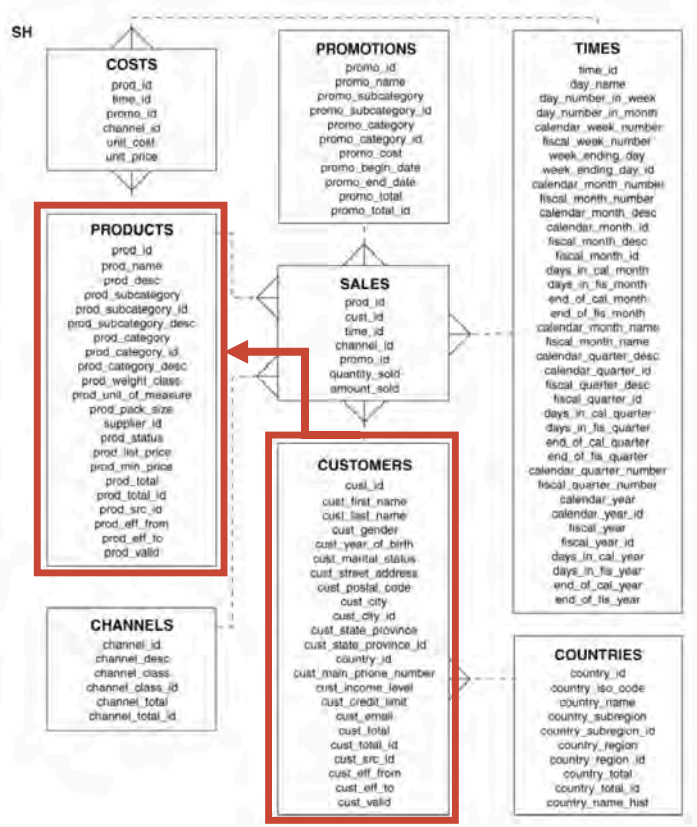
- **Customer** entities
- **Product** entities
- **Purchased** relationships



Step 2 - Understand the Source Data Model

Looking at the source data, typically in relational model, find from where we can obtain the necessary information = entities, relationships, and their attributes.

- **Customer** entity and its attributes (first name, last name, gender, ...) are from CUSTOMERS table
- **Product** entity and its attributes (name, category, size, ...) are from PRODUCTS table
- **Purchased** relationship and its attributes (quantity, amount, ...) are from SALES table. This table holds the n:m relationships between customers and products.



Step 3 - Create Mapping

The mapping between relational models and graph models can be written in PGQL.

In CREATE PROPERTY GRAPH statement:

- VERTEX TABLES clause and EDGE TABLES clause list the source tables
- This statement creates a graph with materialized vertices and edges

```
CREATE PROPERTY GRAPH sh_purchase
  VERTEX TABLES (
    customers
      PROPERTIES (CUST_ID, CUST_FIRST_NAME),
    products
      PROPERTIES (PROD_ID, PROD_NAME)
  )
  EDGE TABLES (
    sales
      SOURCE customers
      DESTINATION products
      LABEL purchased
      PROPERTIES (quantity_sold)
  )
```

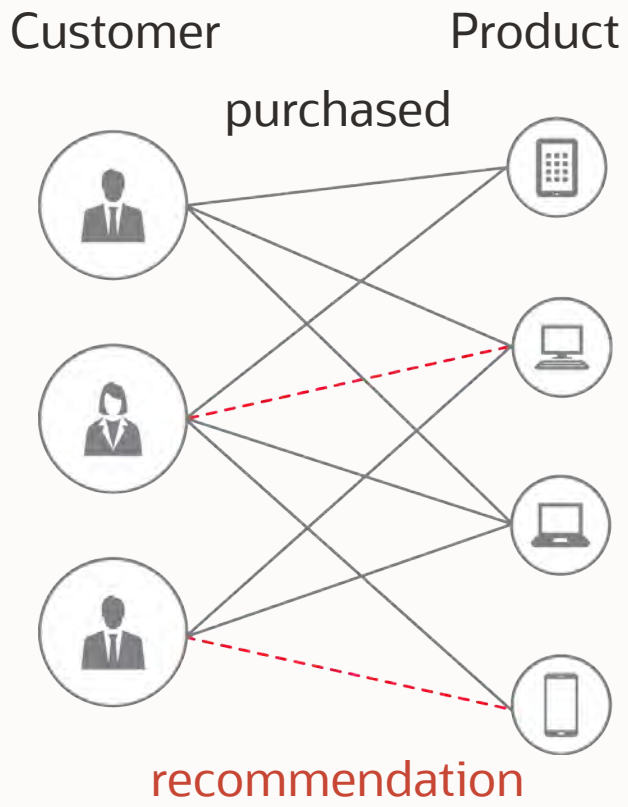
Specification (PGQL 1.3)
https://pgql-lang.org/spec/1.3/#creating-a-property-graph

Step 4 - Try and Improve the Model

Try your analysis use cases using graph queries and graph algorithms on the generated graph.

You might notice that you need to improve the mapping:

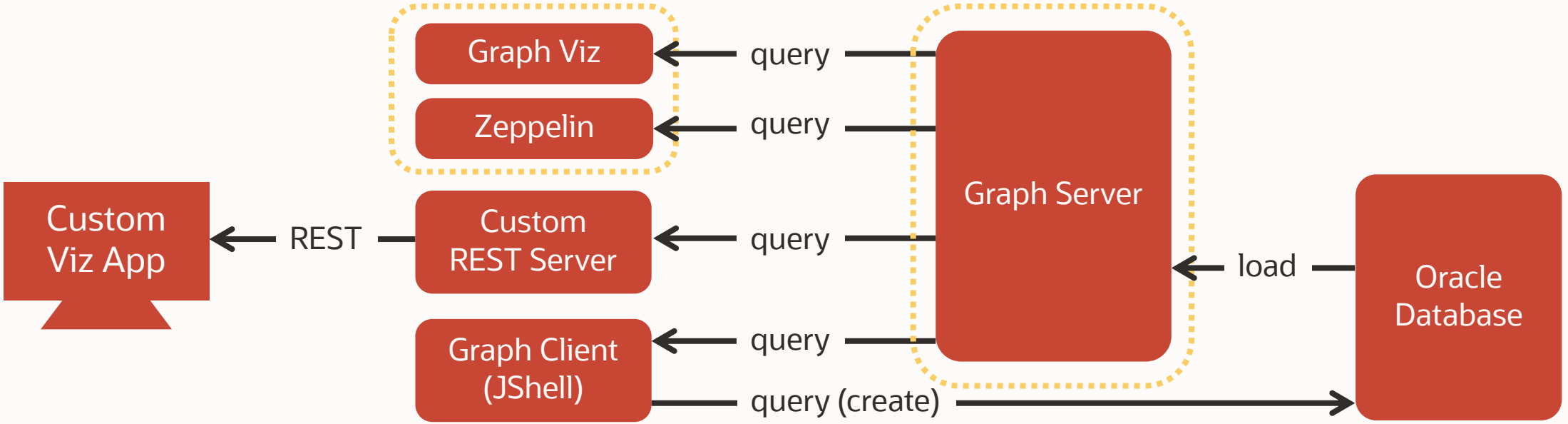
- **Different attributes are useful for different analyses.** For example, the graph should be filtered by time information or country information for generating better recommendation.
- **Vertices and edges are sometimes exchangeable.** For example, to add promotion information into each purchases, you might have to "vertexify" purchases as entities.
- **More information can be added.** More entities and relationships can be generated from different data sources and connected.



Compute Recommendations

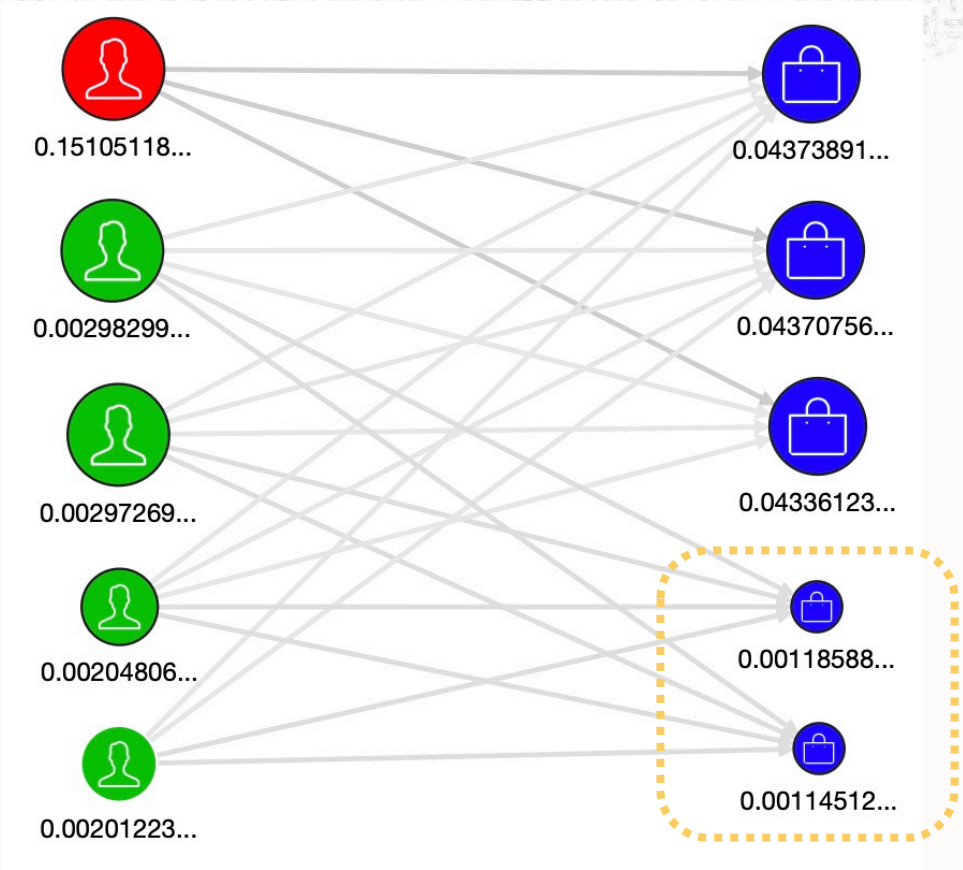
Compute Recommendations

- Graph Server has the **built-in algorithms** which are often used for recommendation systems
- Users can run the algorithms with Zeppelin, and visualize the results with Garph Viz



Algorithm 1 - Personalized Pagerank

- Idea
  - Customers (or products) which have more paths from the target customer can get higher ranks
  - Customers with higher ranks are similar to the target customer in their purchase histories
  - Products with higher ranks should be recommended
- Algorithm
  - Give the initial rank to the starting node only
  - Each node gets ranks from incoming edges, and distribute own ranks to outgoing edges
  - Interate until the ranks are settled



https://blogs.oracle.com/bigdataspatialgraph/intuitive-explanation-of-personalized-page-rank-and-its-application-in-recommendation

Algorithm 2 - Collaborative Filtering

- Idea
  - The customers sharing their purchase activities have the same features (= tastes)
  - The products also hold common features (e.g. bat and grove share "baseball" feature)
  - According to the features between a customer and a product, we can predict their affinity
- Algorithm
  - Run matrix factorization to discover important features and the feature vectors for each customer and product
  - Multiply the vectors to get the predicted score (= the possibility of purchase)

-	Star Wars	Dog Day Afternoon	Blue Velvet	The Toxic Avenger
Alice	1	1	5	4
Bob	1	-	-	4
Jack	5	1	3	-

-	Feature 1	Feature 2	Feature 3	Feature 4
Star Wars	0.25	0.8	0.01	0.053
Blue Velvet	0.96	0.325	0.13	0.46
Dog Day Afternoon	0.21	0.1	0.87	0.83
The Toxic Avenger	0.64	0.23	0.1	0.74

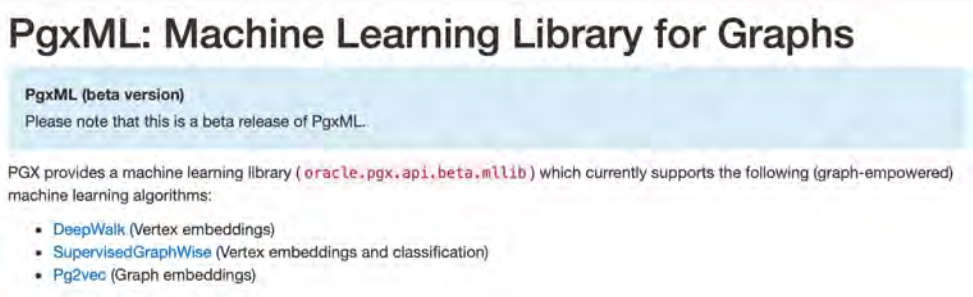
-	Feature 1	Feature 2	Feature 3	Feature 4
Alice	0.25	0.8	0.01	0.053
Bob	0.96	0.325	0.13	0.46
Jack	0.21	0.1	0.87	0.83

https://github.com/oracle/pgx-samples/blob/master/movie-recommendation/README.md

Algorithm 3 - DeepWalk

- Idea
  - Similar customers (or products) have similar connection patterns in the graph
  - Products purchased by similar customers are recommended
  - Similar products are suggested (when the customer is interested in particular products)
- Algorithm
  - Generate random walks for each node
  - Using the random walks as input sequences, run word2vec algorithm to create the vector representation of each node
  - Calculate the distances between the nodes

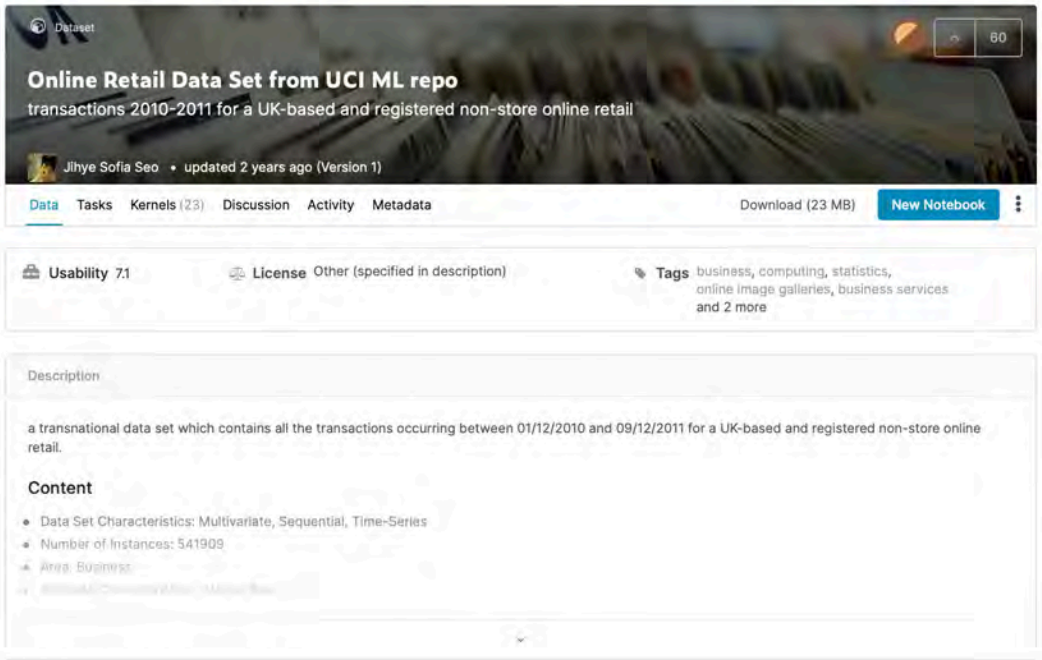
Graph Embeddings — The Summary
https://towardsdatascience.com/graph-embeddings-the-summary-cc6075aba007



https://docs.oracle.com/cd/E56133_01/latest/prog-guides/mllib/deepwalk.html

Demo - Sample Dataset

- Online Retail dataset (Kaggle)
  - 4,339 customers
  - 3,919 products
  - 396,370 purchases
- Data preparation
  - Removed duplicated purchases (266,795 distinct purchases)
  - Added reverse edges (533,590 edges in total)
- Script
  - https://github.com/ryotayamanaka/oracle-pg/tree/master/graphs/retail



https://www.kaggle.com/jihyeseo/online-retail-data-set-from-uci-ml-repo

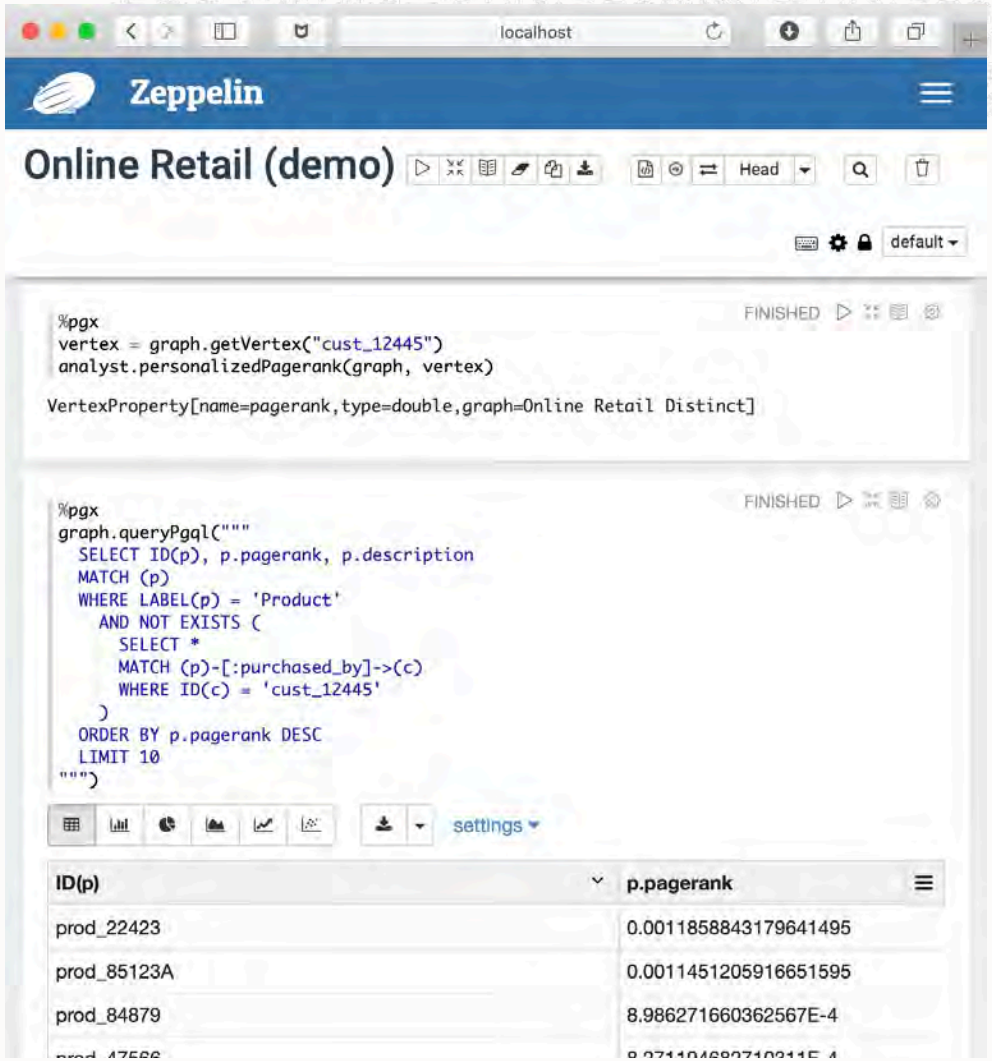
How to Run - Random Walk

- Run algorithm

```
vertex = graph.getVertex("cust_12445")
analyst.personalizedPagerank(graph, vertex)
```

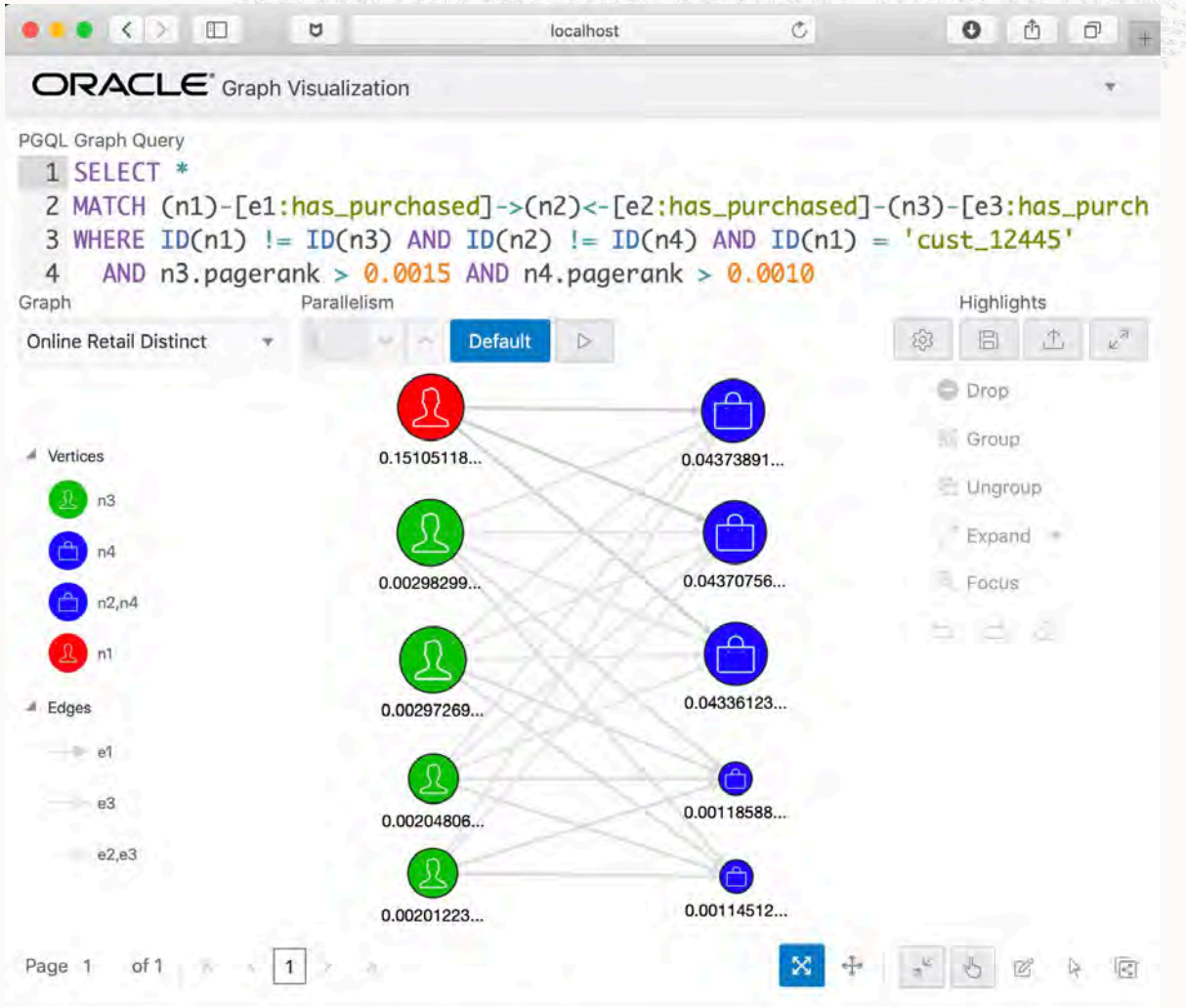
- Retrieve the result (using PGQL query)

```
graph.queryPgql("""
  SELECT ID(p), p.pagerank, p.description
  MATCH (p)
  WHERE LABEL(p) = 'Product'
    AND NOT EXISTS (
      SELECT *
      MATCH (p)-[:purchased_by]->(c)
      WHERE ID(c) = 'cust_12445'
    )
  ORDER BY p.pagerank DESC
  LIMIT 10
""")
```



Graph Visualization

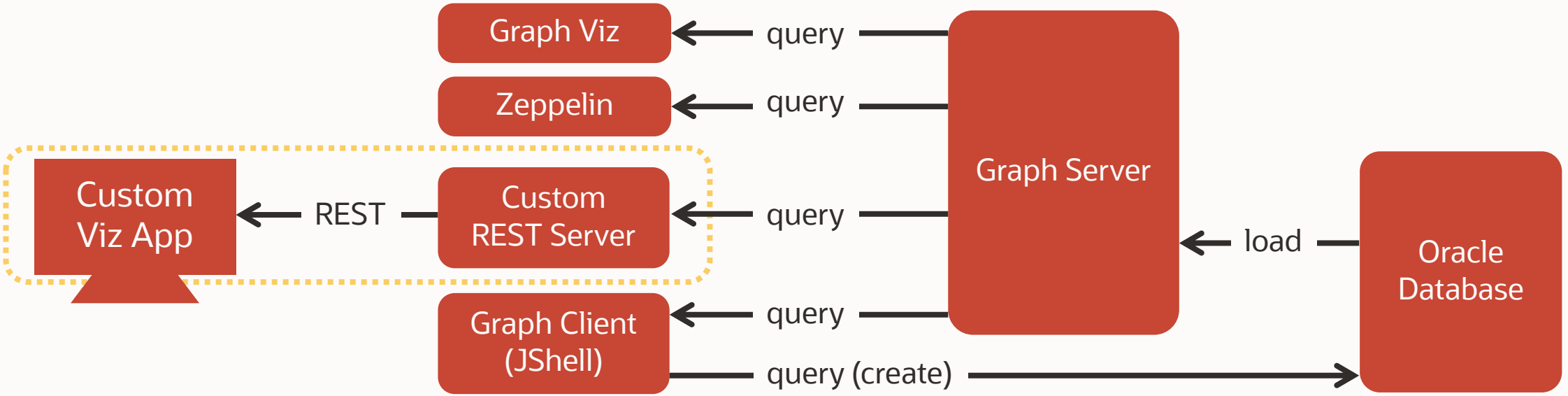
- For the users who write PGQL,
  - Built-in graph visuzlization tool (GraphViz)
  - Interactive visualization against queries
  - Highlights setting can be saved (e.g. the color and size of nodes, layout, ...)
- For the users who do not write PGQL and/or customized visualizations,
  - Custom apps can be designed using Graph Client Java API



Build a Web Application

Implementing Custom App

- Custom applications can be implemented with **Graph Client Java API**.
- In this demo, the REST Server and Viz App are implemented in Java and JavaScript, respectively

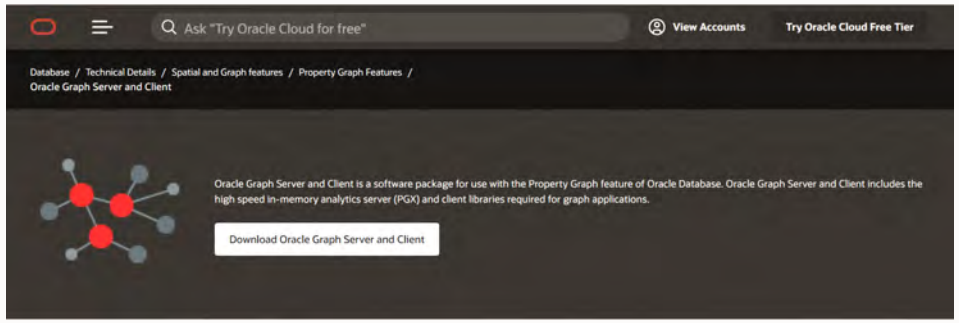


(CAGLA)

Wrap up

Helpful Links

Search for "Oracle Graph Server and Client" to download from oracle.com



- Graphs at Oracle
  https://www.oracle.com/goto/graph
- Oracle Property Graph
  http://www.oracle.com/goto/propertygraph
- Blog: Examples, Tips and Tricks
  http://bit.ly/OracleGraphBlog
- AskTOM Series: https://asktom.oracle.com/pls/apex/asktom.search?office=3084
- Social Media
  - Twitter: @OracleBigData, @SpatialHannes, @Jeanlhm, @ryotaymnk
  - LinkedIn: Oracle Spatial and Graph Group
  - YouTube: youtube.com/c/OracleSpatialandGraph

AskTOM Office Hours: Graph Database and Analytics

- Welcome to our AskTOM Graph Office Hours series!
  We're back with new product updates, use cases, demos and technical tips
  https://asktom.oracle.com/pls/apex/asktom.search?oh=3084
- Sessions will be held about once a month
- **Subscribe** at the page above for updates on upcoming session topics & dates
  And submit feedback, questions, topic requests, and view past session recordings
- Note: **Spatial** now has a new Office Hours series for location analysis & mapping features in Oracle Database:
  https://asktom.oracle.com/pls/apex/asktom.search?oh=7761

ORACLE