

ORACLE®

Technologies for Big Data and Graph Analytics at Oracle Labs

CERN OpenLab open day 2017

Davide Basilio Bartolini – <davide.bartolini@oracle.com>
Sr. Researcher, Oracle Labs Switzerland

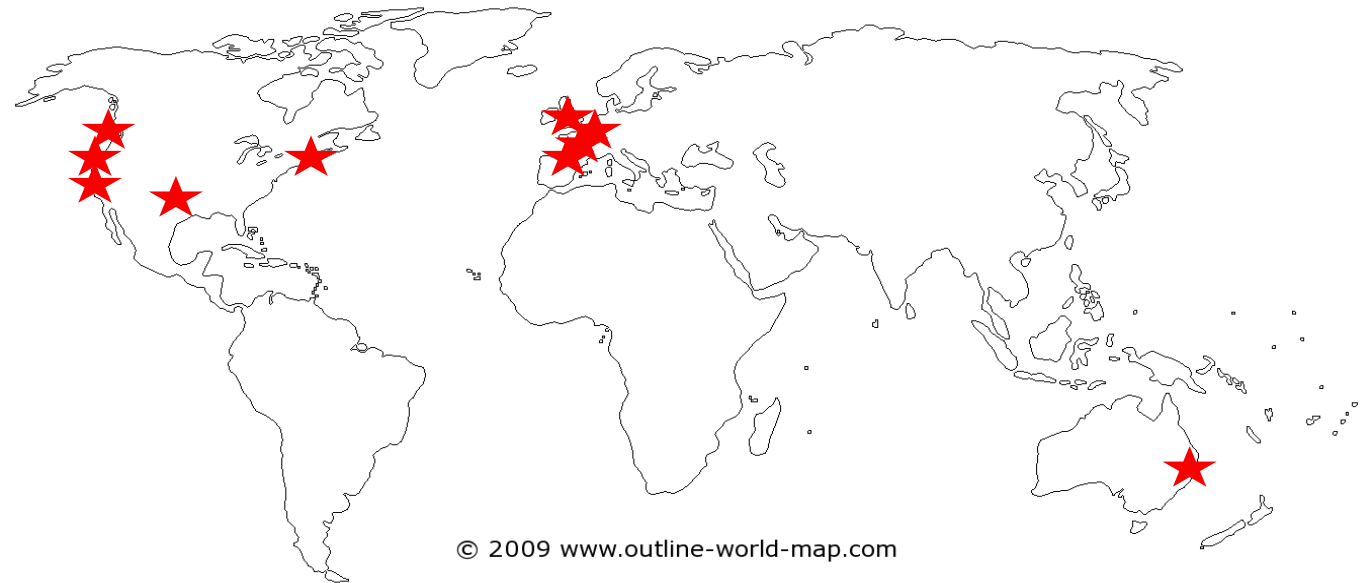
Geneva, September 21, 2017

Oracle Labs

Mission

“Identify, explore, and transfer new technologies that have the potential to substantially advance Oracle’s business.”

- With a global research team
 - 220+ PhD’s, MS, BS
- Wide range of research topics
 - Graph analytics
 - Machine learning and AI
 - More at labs.oracle.com

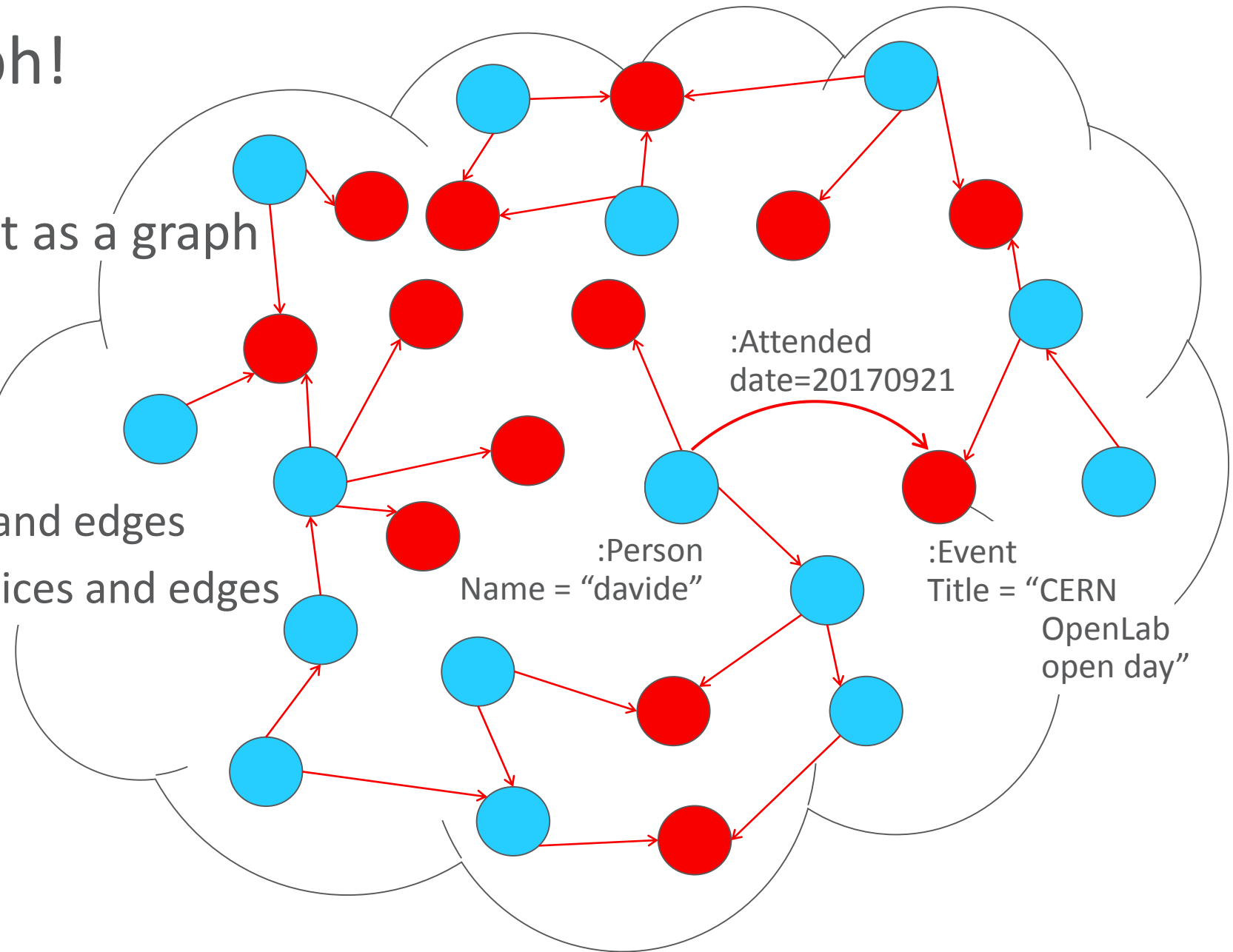


Safe Harbor Statement

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.

Your Data is a Graph!

- Represent your dataset as a graph
 - Entities are vertices
 - Relationships are edges
- Annotate your graph
 - Labels identify vertices and edges
 - Properties describe vertices and edges
- For the purpose of
 - Data modeling
 - Data analysis



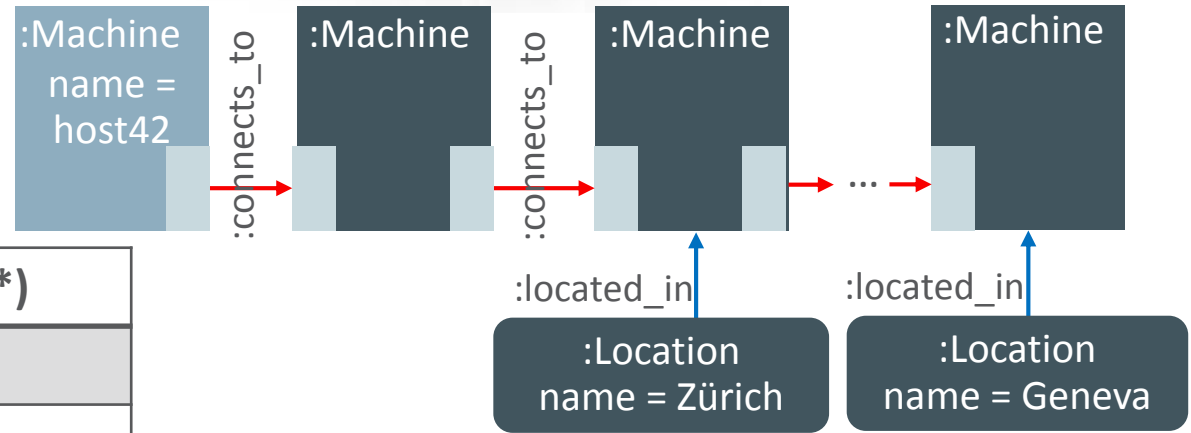
Example: Cloud Network Management

- Use a graph to represent the machines and their interconnections
 - Racks, machines, switches, ports, wires ...



Looking for multi-hop connections

Count the number of Machines (indirectly) connected to host "host42", group by Location name



PGQL query

```
SELECT z.name, COUNT(*)
WHERE (x WITH name = 'host42') -/:connects_to*/-> (y: Machine),
      (y) <- [ :located_in ] - (z:Location)
GROUP BY z
```

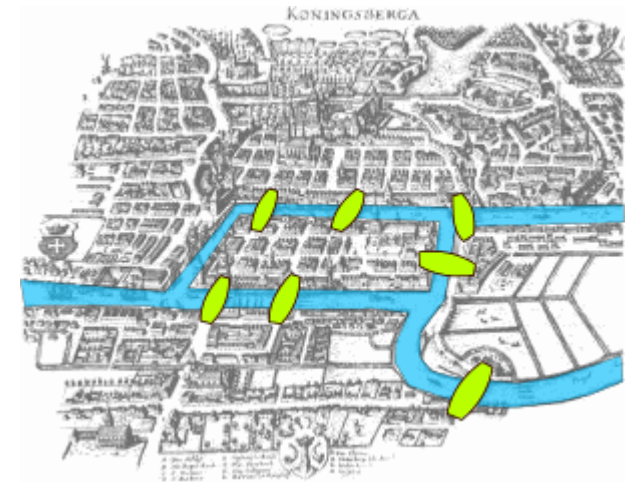
z.name	COUNT(*)
Zurich	1
Geneva	1

Easy to write a query and fast to get the answer thanks to the graph model



Beyond Queries: Graph Algorithms for Powerful Analytics

- Graphs have been studied in Maths for centuries
 - Since Euler’s “*Seven Bridges of Königsberg*”, 1736 [1]
- Classic problems on graphs [1,2]
 - Graph isomorphism
 - Traveling salesman’s problem
 - Max flow, min cut
 - ...
- More recent developments
 - Pagerank [3]
 - InfoMap [4]



Graph algorithms usually iteratively walk over the graph and update a vertex or edge property

=> Random accesses, memory-bound computation

[1] https://en.wikipedia.org/wiki/Graph_theory#History

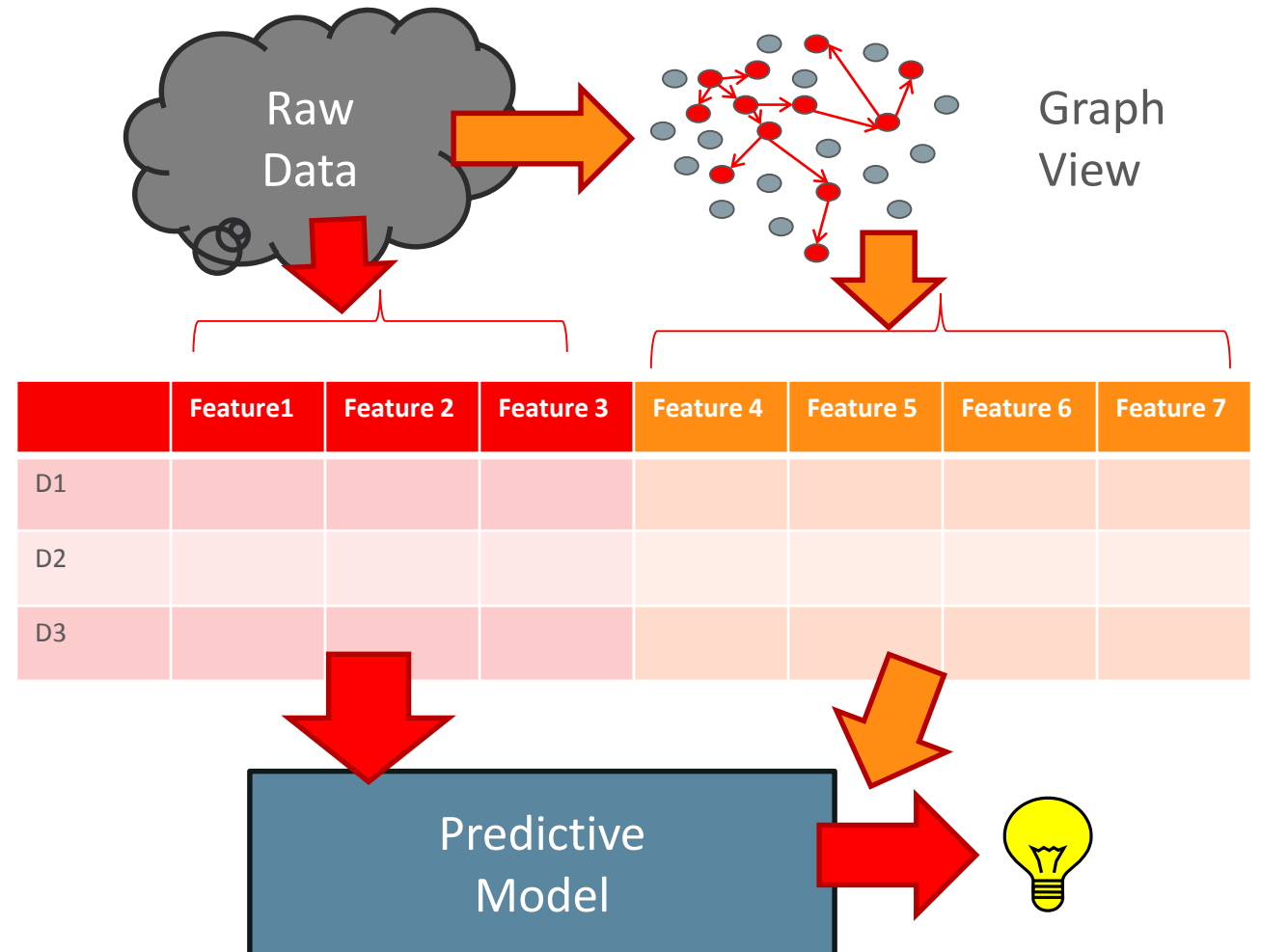
[2] https://docs.oracle.com/cd/E56133_01/latest/reference/algorithms/index.html

[3] <https://en.wikipedia.org/wiki/PageRank>

[4] <http://www.mapequation.org/>

Graph Analysis for Enhanced Machine Learning / AI

- Graph analysis can augment Machine Learning and AI apps
 - Typical ML techniques train models based on directly observed features: “properties”
 - Graph analysis can provide additional *strong* signals by analyzing relationships
 - ... so increasing the accuracy of the ML model



Oracle Labs

PGX – Parallel Graph Analytix

- A fast, parallel, in-memory graph analytics framework
- Provides a graph-specific query language (PGQL)

Match a graph pattern

```

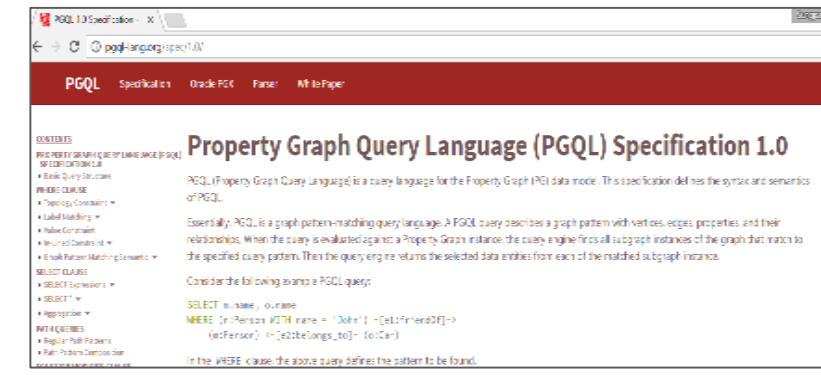
SELECT y.name, y.inDegree(), e, p
FROM snGraph
WHERE
  (x:Person WITH name = 'Paul') -[e:likes]-> (y),
  (z:Person WITH name = 'Amber') -/:likes*/-> (y),
  x.age > y.age
GROUP BY ... ORDER BY ... LIMIT ...
  
```

Edge -[..]->

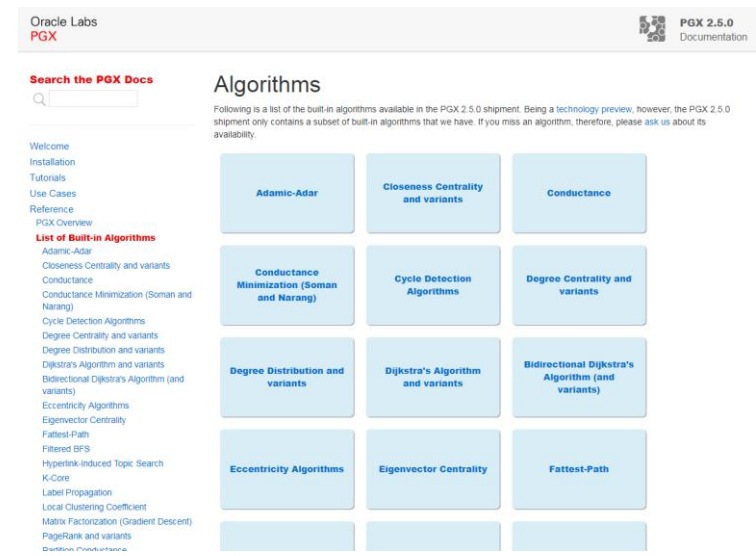
Vertex (..)

Path -/:*/->

Return a "result set" (i.e., a table)

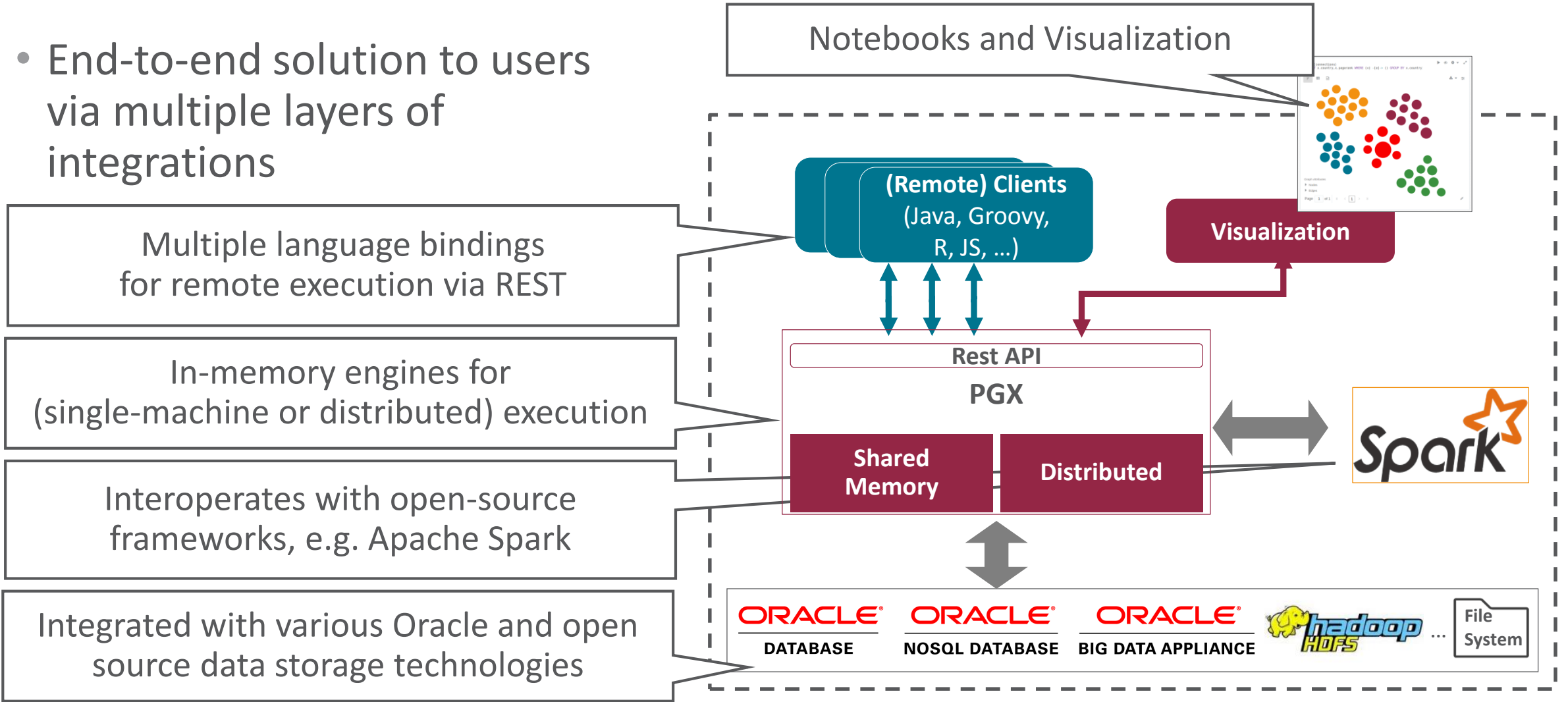


- Offers 35+ built-in, native graph analytics algorithms
- Is both already embedded in Oracle products and an active research project at Oracle Labs



PGX Architecture Overview

- End-to-end solution to users via multiple layers of integrations



PGX Provides High Performance Graph Analytics

- Graphalytics (VLDB2016)
 - A benchmark framework for graph analysis
 - Composed of multiple algorithms and graphs
 - Joint work among academia and industry
 - Industry: IBM, Intel, SAP, Huawei
 - Academia: Led by TU Delft

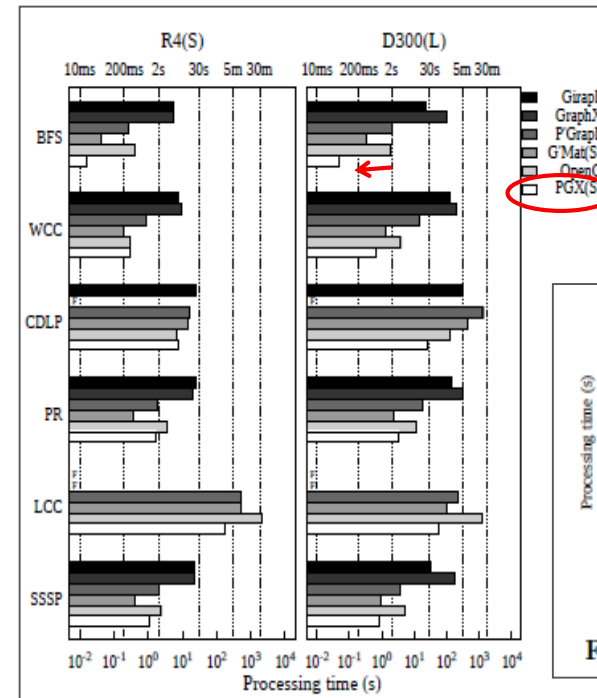
Don't do this at home

- PGX is available as
 - OTN, non-commercial evaluation package
 - Within Oracle BDSG and OSG
 - Within BDCS and DBaaS in the Oracle Cloud
- [spoiler] Tune in to OOW2017 for more [/spoiler]

LDBC Graphalytics: A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms

Alexandru Iosup[△] Tim Hegeman[△] Wing Lung Ngai[△] Stijn Heldens[△] Arnau Prat Pérez[□]
 Thomas Manhardt[◇] Hassan Chafi[◇] Mihai Capotă[◇] Narayanan Sundaram[◇]
 Michael Anderson[◇] Ilie Gabriel Tănase[▽] Yinglong Xia[◇] Lifeng Nai[◇] Peter Boncz[▽]

◇ Oracle Labs ◇ Intel Labs ▽ IBM Research ◇ Huawei Research America
[△] Delft University of Technology □ UPC Barcelona ◇ Georgia Tech ▽ CWI Amsterdam



PGX shows top-tier performance and scalability

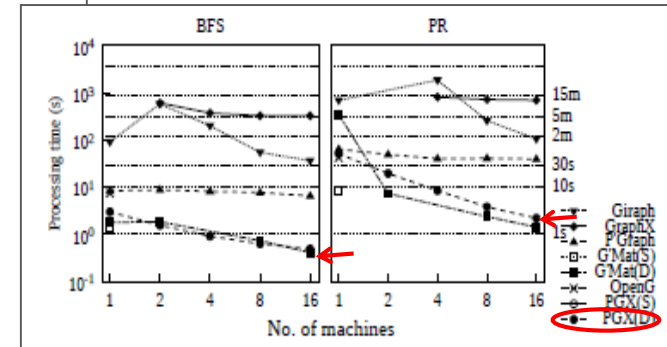



Figure 8: Strong scalability: T_{proc} vs. #machines.



Don't miss the **demo** in the **exhibition area**



Zenodo Keyword Auto-Suggest with Oracle PGX

CERN openlab open day

Davide Bartolini, Manuel Martín Márquez, Luis Rodríguez

21/09/2017

1



PGX on the OTN



labs.oracle.com

Davide Basilio Bartolini — <davide.bartolini@oracle.com>

