

Intern with Domain Global Graphs!

The Domain Global Graphs team at Oracle Labs, which focuses on machine learning and data/graph analytics for enterprise domain data, has open internship positions available.

Oracle

Oracle, a global provider of enterprise cloud computing, is empowering businesses of all sizes on their journey of digital transformation. Oracle Cloud provides leading-edge capabilities in software as a service, platform as a service, infrastructure as a service, and data as a service.

Oracle's application suites, platforms, and infrastructure leverage both the latest technologies and emerging ones – including artificial intelligence, machine learning, blockchain, and Internet of Things – in ways that create business differentiation and advantage for customers. Continued technological advances are always on the horizon.

Oracle Labs

Oracle Labs is the advanced research and development arm of Oracle. We focus on the development of technologies that keep Oracle at the forefront of the computer industry. Oracle Labs researchers look for novel approaches and methodologies, often taking on projects with high risk or uncertainty, or that are difficult to tackle within a product-development organization. Oracle Labs research is focused on real-world outcomes: our researchers aim to develop technologies that will someday play a significant role in the evolution of technology and society. For example, chip multithreading and the Java programming language grew out of work done in Oracle Labs.

Parallel Graph AnalytiX (PGX)

Graph analysis lets you reveal latent information that is encoded, not as fields in your data, but as direct and indirect relationships between elements of your data – information that is not obvious to the naked eye. PGX [1] is a toolkit for graph analysis that supports *running algorithms* such as PageRank on graphs, and performing SQL-like *pattern-matching* on graphs (PGQL – property graph query language), using the results of algorithmic analysis, and *graph Machine Learning (ML)*. PGX includes both a single-node in-memory engine, and a distributed engine for extremely large graphs. Graphs can be loaded from various sources including flat files, SQL, NoSQL databases, Apache Spark and Hadoop; incremental updates are supported. PGX is already available as an option in Oracle products and an active research project at Oracle Labs.



Oracle Labs Data Studio

Oracle Labs Data Studio is a web-based notebook platform for data scientists. By combining live code collaboration in multiple programming languages with graph analytics and rich, interactive visualizations, it accelerates the process of exploring and gaining insights from your data. Data can be imported from various sources and analyzed with interpreter environments for many programming languages (Python, R, Shell, Spark, etc.). For graph data, Data Studio comes packaged with PGX and PGQL, adding an interactive visual layer that supports filtering graphs, highlighting elements, visualizing geographical data, and expanding/contracting the view. Data Studio components form a re-usable base for enterprise software products tailored to specific industries. Example use cases include financial crime detection and compliance [2], machine learning for health sciences, and market segmentation for retail.

Domain Global Graphs

For enterprise use cases of PGX and Data Studio, organizations often adopt a graph data model to integrate various data sources into one global view from their enterprise domain (e.g., financial [2], retail or healthcare data), so that they can run graph analytics and conduct investigations on them. The Domain Global Graph team helps with the integration of PGX and Data Studio into solutions that support the investigation of domain global graphs, and helps with further research on how to produce additional insight to facilitate investigation, e.g., by using Machine Learning techniques.

Internship Details

As part of your internship, you will be working with the Domain Global Graphs team – for example by researching, designing and implementing new features, enhancing existing functionality, measuring and improving performance, fixing bugs, etc. While specific internship topics may change, here is a selection of current focus items:

- Development and enhancement of machine learning and data analysis algorithms for Global Graph applications. Examples use cases:
 - Recognizing graph entities and their relationship from text and tables, employing and improving techniques such as Natural Language Processing (NLP), Named Entity Recognition (NER), Relation Extraction, and Co-Reference Resolution.
 - Generating a narrative text from a domain global graph, adapting them to the graph structure.
 - Improving Entity Resolution (ER) techniques, e.g., by considering entities' fields combinations and value thresholds, or by considering the graph structure and properties of the resolved entities, with explainability as well.
 - Improving global graph investigation, e.g., regarding financial patterns by detecting similar or relevant subgraphs, or by automating the processing and classification of cases under investigation.
 - Improving name and address parsing, standardization and matching, across different languages as well.



Image Caption 1. Our beautiful offices at Prime Tower, Zurich, CH.

- Optimization of data pipeline for integrating data sources into graph model and keeping the data synchronized.
- Ensuring data permission and lineage control for large-scale data integration.
- Customized visualization for interactive data analysis and explainability.
- Optimization and validation of data analysis pipeline and its deployment with modern software architecture.

Qualifications

The successful candidate is expected to complete the internship using a wide and diverse set of skills.

Required Skills

- Thorough understanding of CS fundamentals including data structures, algorithms and complexity analysis
- Experience with object-oriented programming
- Experience with Java and Python programming
- Good problem-solving and analytical skills

Preferred Skills

- Basic understanding of parallel, concurrent, and/or distributed programming
- Basic understanding of machine learning and deep learning algorithms
- Experience with Linux (e.g., bash scripts)
- Familiarity with graph-analytic algorithms
- A high GPA in the master studies
- A high GPA in a machine learning course

Desired Skills

- Experience with Big Data (e.g., Hadoop, Spark).
- Experience with SQL, Kubernetes, Docker, Gradle, Jenkins.
- Experience with text search systems (e.g., Elasticsearch).

References

[1] Parallel Graph AnalytiX (PGX), <http://www.oracle.com/technetwork/oracle-labs/parallel-graph-analytix/overview/index.html>

[2] Financial Crime and Compliance Studio, <https://www.oracle.com/industries/financial-services/analytics/crime-compliance-studio/>

Internship Facts

- Suitable for an internship or a master thesis of three to six months

“Oracle Labs is a great place to perform meaningful research which can directly impact the industry. Projects are well supervised, and the people are always there to help you in order to end up with a successful project.”

Mikael Morales

Former EPFL student,
6-month intern

For more information about the internship, contact Iraklis Psaroudakis (iraklis.psaroudakis@oracle.com).