

Intern with Oracle Labs PGX!

The Machine Learning team within Parallel Graph AnalytiX (PGX) group at Oracle Labs has open internship positions available on Graph Machine Learning development topics.

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.

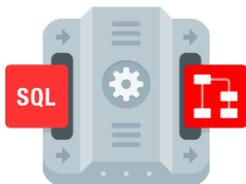
Oracle's application suites, platforms, and infrastructure leverage both the latest technologies and emerging ones – including artificial intelligence, machine learning, blockchain, and Internet of Things – in ways that create business differentiation and advantage for customers. Continued technological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

Relationships in the data are becoming a key feature to enable knowledge discovery from large datasets. Graphs are a powerful abstraction to support this analysis, thanks to their explicit representation of relationships as edges. 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, but can have tremendous value once uncovered.



PGX [1] is a toolkit for graph analysis that supports running (i) algorithms such as PageRank on graphs, (ii) performing SQL-like pattern-matching on graphs using the results of algorithmic analysis, and (iii) graph machine learning techniques like DeepWalk or Graph Neural Networks. Algorithms are parallelized for extreme performance. The PGX toolkit includes both a single-node in-memory engine, and a distributed engine for extremely large graphs. Graphs can be loaded from a variety of sources including flat files, SQL and NoSQL databases and Apache Spark and Hadoop; incremental updates are supported.

PGX is both already available as an option in Oracle products and an active research project at Oracle Labs, with a world-class team of researchers further advancing the capabilities of the toolkit.

Internship Details

At Oracle Labs PGX group, we are actively working on challenging machine learning problems with a focus on graph-represented data. We are developing a state-of-the-art Machine Learning library primarily to support various Graph-based ML techniques. These specific graph-based algorithms mainly deal with vector representations of (i) vertices in a graph, or (ii) sub-graphs, or (iii) even a complete graph. The use-cases of these approaches are vast and spans across multiple domains starting from Finance, Bio-med, to Cybersecurity. While implementing these functionalities, our focus is on the following objectives.

1. **Scalability:** Scalable implementations of multiple ML algorithms (graph-based and classical ones) that leverage parallel accelerators for optimal performance..
2. **Efficient Memory Consumption:** Graph-based ML algorithms are extremely memory-greedy so it is crucial to employ efficient memory management to handle large-scale graphs on a single machine.

The goal of this project is to design and implement novel graph learning algorithms (or optimize existing algorithms) that scale-up on large-scale graphs by accounting for the above-mentioned challenges.

Required Skills

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

- Basic understanding of machine learning and deep learning algorithms
- Experience with Java/Python/C++ programming
- Experience with ML platforms like Tensorflow or PyTorch
- Excellent problem-solving and analytical skills
- Experience in parallel programming (multicore CPU or GPU) is a plus
- Familiarity with graph-analytic algorithms is a plus
- A high grade in a machine learning course is required.
- Suitable for regular internship or M.Sc. Thesis

For more information, contact Rhicheck Patra (rhicheck.patra@oracle.com).

References

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

“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

Related Topics

Our group at Oracle Labs further offers various internship topics in the following areas:

- Extending a Distributed Graph Engine (Oracle Labs PGX)
- Autonomous Graph Support in the Oracle Database
- Extending Oracle Lab's Autonomous Middleware Platform
- Extending a Web-Based Data Science Platform
- Graph-Powered CyberSecurity Intelligence
- Machine Learning for Compliance
- Machine Learning and Data Analysis Techniques for Domain Global Graphs

If you are interested in more than one of these areas, it is sufficient to apply once. In our interview process, we are going to take all your areas of interest into account.