Internship or MSc Thesis at Oracle Labs

The PGX team at Oracle Labs has an opening for an internship to explore and apply software language engineering techniques in the context of graph analytics.

Internship Facts

- Advisors at Oracle Labs: Tomas Faltin (local advisor)
- Suitable for an internship of four to six months
- Can be combined with a MSc thesis
- Location: Oracle Labs Prague, Czech Republic
- Gross salary: 65.000 CZK per month

Software Language Engineering for Graph Analytics

Graphs are a powerful abstraction to enable knowledge discovery from relationships in large datasets, thanks to their explicit representation of relationships as edges. Graph analysis reveals latent information that is encoded, not as fields in the data, but as direct and indirect relationships between elements of the data – information that is not obvious to the naked eye, but can have tremendous value once uncovered.

A variety of tools such as Neo4j, Apache Spark, and Apache TinkerPop support graph analysis. With this variety of tools comes a variety of domain-specific languages for graph analysis such as Cypher, GraphQL, and Gremlin. PGX is a toolkit for graph analysis that supports both *running algorithms* such as PageRank on graphs, and performing SQL-like *pattern-matching* on graphs, using the results of algorithmic analysis. PGX provides domain-specific languages to express graph algorithms (Green-Marl) and graph queries (PGQL).

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. The PGX team has an opening for an internship to explore and apply software language engineering techniques to execute a variety of graph analysis languages with multiple graph analysis tools. This includes, but is not limited to, the following topics:

Source-to-source compilation: Graph analysis languages are often tied to a single graph analysis tool. Compiling one graph analysis language into other graph analysis languages enables execution with other tools as well. We want to explore the compilation of a variety of graph query languages to Green-Marl and PGQL to enable execution in PGX, and to SQL and PL/SQL to enable execution in the Oracle Database.

Intermediate representation: Source-to-source compilation typically does not scale well with increasing numbers of languages and execution tools. We want to explore an intermediate representation as a common compilation target for different graph analysis languages, which can then be mapped to different execution tools.

Qualifications

Minimum qualifications

- BSc degree in Computer Science or related technical field, or equivalent practical experience.
- Good knowledge and understanding of Software Language Engineering techniques such as compilation and intermediate representations.
- Software Development experience through hands on coding in a general purpose programming language.

Preferred qualifications

- Software Development experience in Java.
- Distinguished problem-solving skills.
- Working proficiency in verbal and written English.
- Software Language Development experience through hands on coding in a language workbench such as Spoofax, MPS, xText.
- Some familiarity with data processing in general and graph processing in particular.

About Oracle and Oracle Labs

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 invests heavily in research and development: US$6.2 billion in FY 2017. 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 is the only organization at Oracle that is devoted exclusively to research.

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.

For more information about internships and MSc thesis topics in the PGX project, contact Tomas Faltin (tomas.faltin@oracle.com).