

The power of many: combining graph analytics with RDF queries in Rheem

Project Description:

Today there are more and more enterprises using knowledge graphs to facilitate and enhance the functionality of their services and products. For example, Google uses a knowledge graph for its search engine's infobox, while Walmart and Amazon use knowledge graphs in product search and advertising. Applications very often require sophisticated analytics on the knowledge graphs (represented in RDF) that combine declarative queries (in SPARQL) with generic graph mining algorithms. This goes beyond the capabilities of current RDF engines or graph systems. The goal of this project is to leverage [Rheem](#), a cross-platform data processing system, to offer such sophisticated analytics. For this, a query language interface on top of Rheem and an RDF engine as a platform inside Rheem need to be built.

Duties/Activities:

- Build a parser for SPARQL queries that support graph mining
- Translate SPARQL queries to Rheem jobs
- Add an RDF engine as a platform in Rheem

Required Skills:

- Good programming skills (mainly Java)
- Compilers background is a plus

Preferred Intern Academic Level:

- Junior or Senior students

Learning Opportunities:

- Knowledge graphs / RDF and SPARQL
- Cross-platform data processing/Polystores
- Big data technology

Expected Team Size: 1-2

Mentors

Name: Zoi Kaoudi

email: zkaoudi@hbku.edu.qa

Name: Jorge Quiané

email: jquianeruz@hbku.edu.qa