

File-Based Analysis of Large Graph Data in the Cloud

Internship / Master Thesis

We are looking for motivated computer science students for the following internship or Master thesis project.

PROJECT DESCRIPTION

Analysis methods for large graph structures are vital for many domains including social networks, the Internet of Things and supply chain management. For instance, recommending new friends on Facebook can be achieved by a neighborhood search in the "friendship graph." Existing approaches either map these problems to a relational model or use dedicated graph databases for the analysis. For cloud-based applications, storing large amount of graph data in a dedicated database can introduce an unacceptable cost overhead for storage and computation. In many cases, customers simply store their data as files in a data lake such as Amazon S3.

The goal of this project is to apply basic graph analysis problems to file-based data in the cloud. The main challenge is to perform the graph analysis "on-the-fly" while scanning the files without loading the entire data into main memory. The scope of this project includes designing and implementing file-based algorithms for graph analysis problems such as neighborhood traversals, shortest paths or simplified graph pattern matching. The prototype should be implemented in a standard programming language and evaluated in terms of performance and scalability and include a comparison of different file formats and partitioning schemes.

This is a joint project of the SAP Innovation Center Potsdam and the System Analysis and Modeling group headed by Prof. Holger Giese at the Hasso Plattner Institute.

CONTACT

Dr. Christian Krause

SAP SE

christian.krause01@sap.com

Dr. Leen Lambers

Systemanalyse und Modellierung

Hasso Plattner Institut

leen.lambers@hpi.de

