

Datei-basierte Analyse von großen Graphdaten in der Cloud

Praktikum / Masterarbeit

Wir suchen motivierte Informatik-StudentInnen für das folgende Praktikum oder Masterarbeit.

PROJEKT BESCHREIBUNG

Die Analyse von großen Graphstrukturen wird zunehmend wichtiger in Bereichen wie Sozialen Netzwerken und dem Internet der Dinge. Zum Beispiel kann das Problem, neue Kontakte in Facebook zu empfehlen durch eine Nachbarschaftssuche im "Freundschaftsgraphen" gelöst werden. Existierende Ansätze lösen solche Probleme durch eine Abbildung in ein relationales Modell oder benutzen spezialisierte Graphdatenbanken. Für Cloud-basierte Anwendungen stellt das Speichern von großen Mengen Graphdaten in Datenbanken oft ein zusätzliches Kostenproblem dar. In vielen Fällen speichern Kunden ihre Daten einfach in Dateien in einem Data Lake, zum Beispiel in Amazon S3.

Ziel dieses Projektes ist es ausgewählte Graphanalyseprobleme auf Datei-basierte Graphdaten in der Cloud anzuwenden. Die Hauptherausforderung ist hierbei, die Analyse "on-the-fly" beim (mehrfachen) Lesen der Dateien zu realisieren ohne den gesamten Graphen in den Hauptspeicher zu laden. Dazu sollen Datei-basiertere Algorithmen für ausgewählte Graphprobleme wie Nachbarschaftssuche, kürzeste Pfade oder einfaches Graph-Pattern Matching entworfen und prototypisch implementiert werden. Der Prototyp sollte in einer Standard-Programmiersprache entwickelt werden und im Hinblick auf Performance und Skalierbarkeit untersucht werden. Dabei sollten auch unterschiedliche Dateiformate und Partitionierungsschemas untersucht werden.

Das Projekt wird angeboten vom SAP Innovation Center Potsdam und dem Fachbereich Systemanalyse und Modellierung geleitet von Prof. Holger Giese am Hasso Plattner Institut.

KONTAKT

Dr. Christian Krause

SAP SE

christian.krause01@sap.com

Dr. Leen Lambers

Systemanalyse und Modellierung

Hasso Plattner Institut

leen.lambers@hpi.de

