

# Master-/Diplomarbeit zu vergeben!

## Cloud-basierte Speicherung und Abfrage von RDF im MapReduce-Framework

Seit einigen Jahren werden immer mehr Daten als RDF (Resource Description Framework) [1], einem graph-basierten Datenmodell, veröffentlicht. So sind mittlerweile Milliarden von Fakten als „Linked Open Data“ frei verfügbar [2]. Die Verarbeitung dieser riesigen Mengen an Daten stellt neue Herausforderungen an die technische Infrastruktur und erfordert den Einsatz von verteilten, parallelen Systemen um Skalierbarkeit zu gewährleisten. Zur Speicherung von RDF-Daten wurde am Institut AIFB CumulusRDF [3] entwickelt, ein RDF-Store, der auf Apache Cassandra aufsetzt, einer populären cloud-basierten Datenbank.

Im Rahmen dieser Master-/Diplomarbeit soll der Einsatz von MapReduce, einem populären Framework zur Verarbeitung großer Datenmengen in Clustern, für RDF-Daten evaluiert werden. Ziel ist es, die Ausführung von graph-basierten Algorithmen und Analysen zu ermöglichen. Dabei sollen bestehende Open-Source Frameworks wie Apache Hadoop [4] und DataStax Brisk [5] zum Einsatz kommen und CumulusRDF nutzen.

Wünschenswert sind Kenntnisse in und Interesse für Java-Programmierung und Datenbanken.



Bei weiterem Interesse:

Günter Ladwig, Institut AIFB

[guenter.ladwig@kit.edu](mailto:guenter.ladwig@kit.edu)

[http://www.aifb.kit.edu/web/Günter\\_Ladwig](http://www.aifb.kit.edu/web/Günter_Ladwig)

0721 / 608 - 47946

Weiterführende Links:

[1] [http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework)

[2] <http://linkeddata.org/>

[3] <http://code.google.com/p/cumulusrdf/>

[4] <http://hadoop.apache.org/>

[5] <http://www.datastax.com/products/brisk>