Zürcher Hochschule
für Angewandte Wissenschaften

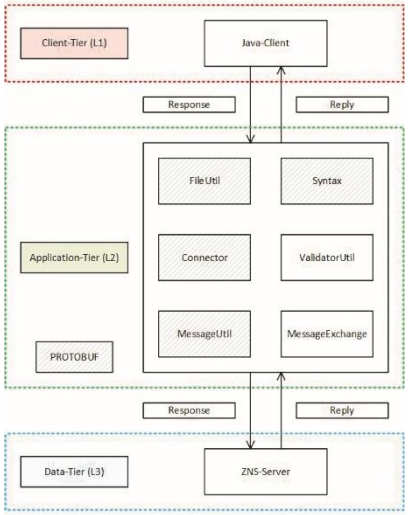
School of
Engineering

# Entwicklung und Performance-Evaluierung eines SQL-light-Clients für NoSQL-Technologien

Diplomierende
Güney Gülsoy
Dominik Rieder

Dozierende
Thomas Oskar Weinmann
Kurt Stockinger


In der Industrie werden NoSQL-Systeme vermehrt eingesetzt, da sie die Verarbeitung grosser Datenmengen erlauben, schemaflexibel sind und eine bessere horizontale Skalierbarkeit versprechen. Durch die Analyse von Rohdaten k&#127;önnen Unternehmen z.B. Kundenbed&#127;ürfnisse identifizieren oder auch Profile erstellen. Doch wie so oft werden auch in dieser Branche verschiedenste Produkte mit vielen Anwendungsm&#127;öglichkeiten angeboten.
In dieser Bachelorarbeit wird einerseits ein Java-Client f&#127;ür ZurichNoSQL (ZNS) programmiert und andererseits multidimensionale Queries auf verschiedenen Datensätzen ausgef&#127;ührt, um so eine Performance-Evaluierung zwischen zwei verschiedenen Systemen durchf&#127;ühren zu k&#127;önnen. Für ZNS wird eine SQL-ähnliche Syntax definiert, die anhand des Java-Clients in eine ZNS-konforme Anweisung umgewandelt wird. Die zu untersuchenden Datenbanksysteme sind MongoDB, eine dokumentbasierte NoSQL-Datenbank, und ZNS, eine an der ZHAW entwickelte Main-Memory-Datenbank. Der ZNS-Server wird mit protobuf von Google &#127;über den entwickelten Java-Client angesprochen, wobei Insert-, Select- und Update-Anweisungen von der API unterstützt werden.
Die beiden Systeme wurden zun&#127;ächst auf zwei verschiedenen Testumgebungen nach der vertikalen Skalierung untersucht. Daraufhin wurden die Experimente in einem Cluster von zwei und vier Knoten wiederholt, um klare Schl&#127;üsse bez&#127;üglich der Performance zu ziehen. Die Experimente beziehen sich dabei auf Inserts, Selects und Updates. So beinhalten bei den Experimenten die generierten Datens&#127;ätze eine Dichte von 1% und besitzen zwischen $10^6$ und $10^8$ Einträge. Die Resultate bestehen aus einem Vergleich zwischen MongoDB und dem Java-Client sowie aus dem entwickelten Java- und dem bestehenden C++-Client.
Die Ergebnisse dieser Experimente haben gezeigt, dass der Java-Client ungeachtet der eingesetzten Queries mit konstant schnellen Antwortzeiten &#127;überzeugt. Der Vergleich mit MongoDB hat ergeben, dass der Vorteil vom Java-Client bei steigenden Dimensionen und grösseren Datenmengen bez&#127;üglich der Verarbeitungszeiten zunimmt. Der Importvergleich mit dem C++-Client zeigte, dass der Java-Client durch weitere Datenverarbeitungen an Zeit einbüsst. Durch begrenzte Hardwareleistungen der Testumgebungen bedarf es weiterer Untersuchung, um die Performance bei realen Datens&#127;ätzen evaluieren zu k&#127;önnen.




Die Architektur des zu entwickelnden Java-Clients für ZurichNoSQL (ZNS). Ersichtlich sind die drei Schichten, wobei die Entwicklung der API im Application-Tier stattfindet.