



## School of Engineering

InIT Institut für angewandte  
Informationstechnologie

### Complex Query Processing with NoSQL

NoSQL-Datenbanksysteme erfreuen sich einer immer grösseren Beliebtheit, denn sie versprechen, grosse, semistrukturierte Datenmengen analysieren zu können. Doch es gibt grosse Unterschiede in den Anwendungsbereichen zwischen den verschiedenen Produkten, da sie für spezielle Bedürfnisse entwickelt wurden. In dieser Bachelorarbeit wird nach einem möglichst schnellen Weg gesucht, komplexe Abfragen auf solchen Datensätzen zu berechnen. Dabei werden drei unterschiedliche Datenbanksysteme unter die Lupe genommen. ZNS ist eine an einem Institut der ZHAW entwickelte NoSQL- und Main-Memory-Datenbank. Als bekannte NoSQL-Datenbank steigt MongoDB ins Rennen, welche oft für Webanwendungen verwendet wird. Spark, ein Main-Memory Datenanalyse Framework im Bereich Cluster Computing, ist das dritte untersuchte System.

Verarbeiten von grossen Datenmengen benötigt viel Rechenleistung - oft mehr als ein einzelner Server leisten kann. Für diese Bachelorarbeit werden deshalb die drei Systeme in einem Cluster installiert und in zwei verschiedenen Experimenten getestet. Im ersten Experiment wird die Skalierung auf verschiedenen Clustergrössen untersucht. Dabei zeigt sich, dass alle Systeme schneller werden, je grösser der Cluster ist. Spark kann dabei den kleinsten Verbesserungsfaktor vorweisen. ZNS ist auf allen Clustergrössen um mindestens den Faktor 6 schneller als MongoDB, dem nächsten Verfolger.

Im zweiten Experiment werden die Datenbanksysteme auf der Basis von realen Daten aus dem Bereich Medizin der U.S. Food and Drug Administration (FDA) getestet. Dabei wird der ganze Anwendungsablauf vom Datenmodellieren bis hin zum Stellen von Anfragen durchgeführt. ZNS überzeugt wiederum mit konstant schnellen Antwortzeiten. Aus dem zweiten Experiment wird die Wichtigkeit eines optimalen Datenmodells bei NoSQL-Datenbanken deutlich. Im Gegensatz zu relationalen Datenbanken kann hier auf kein Standard-Datenmodell zurückgegriffen werden. Um herauszufinden, welches die idealen Datenmodelle für die drei Datenbanksysteme sind, bedarf es weiterer Untersuchungen.



Diplomand  
Jonas Heitz

Dozent  
Kurt Stockinger

Bild klein 1.

Bild klein 2.