# TU Clausthal

## Sharding-Based Scalability and Performance Analysis in NoSQL Databases: Case of MongoDB

## Ancel Kengmegne Sando, B.Sc., TU Clausthal

Nowadays, most companies are confronted with the problem of data growth. The storage and management by companies of enormous data volume, which continue every day exponentially to grow, is becoming a serious challenge. My research focuses on MongoDB, one of the most currently popular NoSQL databases, since it is designed to support very huge data volume of data, ensuring scalability, high performance and availability of transactions. Because of its ease to use and its efficiency, MongoDB gains a privileged place on the NoSQL database market. To ensure scalability, as well as high performance, MongoDB uses either **Scaling Up** (increase the capacity like memory, disks space, CPU ... of the self-server) or **Scaling Out** (also called sharding, the distribution of data set across multiple servers) strategy. The sharding is more suitable to ensure the scalability, high performance in a production system. Therefore, the principal objective of my work is to analyze the scalability of MongoDB, using a sharding environment–in more detail, how the division of the data set across multiple servers will influence the transaction performance when the data grows.

I begin my work with an overview of MongoDB: what it is, its functionality and how it supports the sharding. Here, all important elements for the understanding of MongoDB will be mentioned. To perform as easiest as possible performance and scalability tests, the best idea is to develop a tool that will support shard cluster administration tasks and will avoid the need of the mongo shell, as well as the MongoDB query language (JavaScript) to perform tasks. The ease of use will be ensured by a Graphical User Interface (GUI), allowing the user to perform tasks just by button click. The tool will help us next to perform performance tests, consisting of write, read and read-and-update of very large data set (Benchmark) in a shard cluster using successively one, two, three and four shards and the different test results (read, read-and-modify and write time) will be graphically represented. Through the obtained test results, we will analyze the scalability and performance of MongoDB and know whether MongoDB is well scaled or not. Then I will compare my results to the state-of-the-art, which means I will evaluate what the already done MongoDB's scalability analysis were, what different results were obtained, and what the position of my one result is, as well as the gain in knowledge compared to previous researches. At the end a conclusion of the work will be presented and ideas will be proposed about future works on that topic.

## Mittwoch, den 17.06.2015, 9 Uhr s.t.
## im Seminarraum 210, IfI, Am Regenbogen 15, D8