Tuesday 27.09.2022, 14:00
Status-Talk by Johannes Pietrzyk
Supervisors: Wolfgang Lehner, Jeronimo Castrillon

Title:
"SharedVector – Using SIMD to optimize complex analytical database queries"
Abstract:
Modern database systems have to deal with an ever-growing amount of data and an increasing number of requests. Therefore, state-of-the-art database engines heavily utilize hardware capabilities.
This includes common paradigms like Multiple-Instruction-Multiple-Data (MIMD) provided by modern multi-core CPUs and data-parallel-processing (SIMD). Traditionally, MIMD is used to optimize both single query latency and workload throughput by leveraging intra- and inter-query parallelism, respectively. Conversely, SIMD is only considered to reduce single query latency by exploiting data-level parallelism. Despite individual strengths, all mentioned approaches face the memory-wall effect and thus are often memory-bound.
When multiple queries arrive in a system, they often execute the same (but potentially differently parameterized) operation on the exactly same data. This holds especially true for queries stemming from role-based software systems. Such queries involve, for instance, frequent filtering for roles played by a specific natural type that typically resides in a plays-relation. By exploring commonalities across incoming queries, work-sharing approaches can mitigate or even wholly omit redundant work and memory access.
This thesis investigates how to close the gap for SIMD-based throughput optimization through work-sharing and by leveraging a data-centric approach.
