Doctoral thesis

Towards Efficient Parallel Data Processing on Modern Hardware

RNDr. Zbyněk Falt

The thesis constitutes a set of advanced parallelization techniques and algorithms, which provide a logical path from low-level problems to the high-level ultimate goal.

The first notable contribution is the NUMA-aware scheduler for fine grained pipeline task parallelization. It is fully fledged task scheduler which speeds up computations in pipeline task environment by about tens of percent carefully using architecture cache hierarchy and task placement. The task scheduler has been implemented and thoroughly tested in the parallel pipeline task environment called Bobox. As a future work we have a plan to replace current TBB (Threading Build Blocks by Intel) scheduler by the presented one.

The second important contribution is a new language Bobolang for describing execution plans in the pipeline task environment. The language is then used either for describing execution plans by hand, but more importantly for writing execution plans by compilers of high-level languages, e.g. SPARQL compiler translates SPARQL query to the Bobolang code. The Bobolang code is compiled for a given system architecture in runtime before the query execution. Bobolang language offers a set of operators for describing generalized parallelization points, where the Bobolang compiler can make optimizations for the given system architecture. The Bobolang language and compiler were again implemented and thoroughly tested in the Bobox environment.

The next steps are novel parallel algorithms for data intensive computations. The author rewrote well-known sort and merge-join algorithms and created new algorithms, which are suitable for parallel execution in the pipeline task environment and offer high scalability and data skewness resistance. The new algorithms were used and tested in the SPARQL runtime built upon Bobox environment.

The final step is an implementation of a parallel runtime for SPARQL. Putting all previous contributions together results in one of the fastest SPARQL in-memory engine among all well-known SPARQL engines.

Our research cooperation with the author of the thesis has begun more than five years ago. We started our cooperation with author's master thesis "Scheduler and memory allocator for the Bobox system". The student graduated with an exceptional final examination score. He was really enthusiastic for using new technologies and paradigms in data oriented problems. Our cooperation has been prolonged as Zbyněk Falt started his PhD study under

my supervision. He has published several refereed papers, including papers on Core A conference, as well as some journal articles. He has proved he can make an independent research as well as working in a research team.

Overall, the presented thesis is a great contribution in the parallel computation research field and the scientific record of the student is remarkable as well. The author has demonstrated the ability for creative scientific work therefore I recommend granting the PhD. degree to RNDr. Zbyněk Falt.

Prague, July 7, 2014

RNDr. Jakub Yaghob, Ph.D.
Department of Software Engineering
Faculty of Mathematics and Physics
Charles University in Prague