

Doctoral thesis

Distributed Graph Query Engine Improvements for Big Data Graphs

Mgr. Tomáš Faltín

The focus of this dissertation is to improve key parts of a distributed in-memory graph query engine. The first part is to optimize the distributed DFS algorithm for asynchronous pattern matching by combining it with BFS, thus achieving an overall performance improvement of the engine. DFS has less spatial complexity, while BFS has much better parallelization, locality of access, and load-balancing capabilities. The second part is a completely novel extension of the pattern matching algorithm to RPQ with support for variable length patterns based on regular expressions. This approach allows us to manage memory usage much better while maintaining performance. And the third key part is the improvement of the planner, which significantly affects the performance of the whole query system. Our planner now collects information at runtime (in lightweight manner) and uses it to select a proper computation plan given the current computational environment. This dissertation is based on previously published papers by extending them with additional details and experiments.

I have worked with the author throughout his studies. I supervised his bachelor's thesis, master's thesis and finally his PhD thesis. The original plan when starting the PhD was to build on the master's thesis and extend our in-house parallel computing system Bobox with distributed computing. However, it soon became clear that we did not have sufficient resources for such a development. Moreover, we lacked really big data to perform sufficiently high-quality experiments. At this point, we received a helping hand from Oracle, which has been working very well with MFF UK on several projects and is an important partner of ours. Our faculty provided theoretical and practical knowledge, and Oracle was able to supply the necessary resources.

The slightly longer duration of the PhD is due to the desire to publish at the highest quality conferences (CORE A and better) and also to ensure a sufficiently high-quality implementation followed by a high-quality experimental evaluation of the results of the work. This has resulted in two publications at CORE A ranked conferences and one workshop at a CORE A* ranked conference. During his PhD studies, Tomas had a 6-month internship in Zurich at Oracle Labs. He is currently part of the WHISPER team working on operating systems at Inria in Paris as a researcher since March and will move to a postdoc position after his successful defense.

Overall, the resulting work is an excellent theoretical algorithmic work supported by good implementation and measurement. The publication results achieved are also at an excellent level. The author has demonstrated the ability for creative scientific work therefore I recommend granting the PhD. degree to Mgr. Tomáš Faltín.

Prague, Nov 28, 2023

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