



UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Youngseo Kim

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, May 14, 2024

MSc Thesis

Datenbanktechnologie

Topic: Implementing Inequality and Interval Overlap Joins

Joining relational tables using inequality join predicates is usually very slow, as they are executed as a nested-loop join, resulting in a quadratic run time. Khayyat et al. use an optimized sort-merge join by constructing a bit-array on the sorted tuples [1]. They implement a Bloom-filter algorithm which uses this bit-array index for the join filter.

An even more challenging join predicate involves checking whether two intervals overlap each other. Dignös et al. investigate a method that combines interval-overlap predicates with another join predicate based on equality [2]. They formulate these joins as unions of range joins with an additional equality condition. This approach does not work if the equality predicate is changed to one using inequality.

In summary, the goal of this thesis is to combine the two algorithms mentioned above to combine an interval-overlap predicate with an inequality one. In particular, the tasks are as follows:

Tasks:

1. Literature Review:

- Study the relevant literature [1], [2] to understand the algorithms for inequality and interval overlap-joins to get a deeper understanding of their computational complexity, implementation, and application.

2. Implement a brute-force (nested-loop) inequality and interval overlap join algo-

rithm as a baseline**3. Implement the approach in [1].**

- Implement the algorithm based sort-merge with Bloom-filters for the bit-array scan as described in the paper.

4. Combined the two approaches proposed in [1] and [2].

- Integrate the interval-overlap join approach described in [2] into the inequality join, modifying the algorithm as needed.

5. Evaluation:

- Experimentally evaluate the implementation of the combined approach with the baseline algorithm.
- Analyze the complexity of the algorithm in terms of run time and memory footprint.

6. Write up the thesis

- Describe the implementations, results, evaluations, and other findings in a thesis.
- Present and defend the thesis in a DBTG meeting.

Supervisor: Jamal Mohammed and Sven Helmer

Start date: 01.06.2024

End date: 30.11.2024

References

- [1] Z.Khayyat, W. Lucia, M.Singh, M.Ouzzani, P.Papotti, J.A Quianü-Ruiz, N.Tang P.Kalnis, Lightning Fast and Space Efficient Inequality Joins, VLDB 2015.
- [2] Anton Dignös, Michael H. Böhlen, Johann Gamper, Christian S. Jensen, Peter Moser, Leveraging range joins for the computation of overlap joins. VLDB Journal 2022

Department of Informatics, University of Zurich



Prof. Dr. Michael Böhlen