University of Zurich[UZH]

**Department of Informatics**

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

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

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

Zürich, April 19, 2021

**Master's Thesis (30 KP)**
**Datenbanktechnologie**

**Topic: Transitioning between data and schema values in MonetDB**

Selected relational matrix operations turn data values in an argument relation into attribute names in the result relation. The relational model does not foresee such operations at either the logical or physical level. Since the relational model does not support these operations, relational systems have been built without support for this functionality. This is an impediment for the implementation of linear algebra operations in relational system.

One significant issue when transitioning between schema and data values is that the number of attributes in a result relation depends on the number of input tuples. For example, the transpose operation constructs the schema of the result based on the data values of one of its columns [1], which means that the number of attributes in the result is equal to the number of tuples in the input relation. This is a problem since the information about number and names of attributes is essential for MonetDB when building a query tree. When mapping data to schema values this information only becomes available during query execution when the actual values of input relations are accessed. This prevents MonetDB to seamlessly support relational matrix operations that turn data into schema values.

The goal of this Master's thesis is to design, implement, and evaluate a solution to cleanly transition between data and schema values in a relation, and to illustrate and evaluate the solution in terms of the transpose operation.

The work includes the following tasks:

1. Implement the transpose operation together with representative queries that include the transpose operation (e.g., $\sigma(tra(r))$):

   (a) Identify and illustrate with carefully chosen examples the parts of MonetDB that are affected by the missing information (i.e., the number of attributes in the result relation and their names).
   (b) Suggest approaches to handle the missing information (e.g., placeholders for the attribute names, an extension of query tree nodes, preprocessing of queries, etc) and investigate their advantages and disadvantages.
   (c) Choose an approach to turn data into schema values and implement your approach in MonetDB.
   (d) Empirically evaluate the properties of the proposed solution.
2. Develop optimization rules that allow to omit relational matrix operations that require to turn data into schema values.
3. Write a thesis (approximately 50 pages).
4. Present your thesis in a DBTG meeting.

**References**

[1] Oksana Dolmatova, Nikolaus Augsten, and Michael H. Böhlen. A relational matrix algebra and its implementation in a column store. In David Maier, Rachel Pottinger, AnHai Doan, Wang-Chiew Tan, Abdussalam Alawini, and Hung Q. Ngo, editors, *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*, pages 2573–2587. ACM, 2020.

University of Zürich
Department of Informatics

Prof. Dr. Michael Böhlen