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, May, 2019

Master Project(18 KP)
Datenbanktechnologie

Topic: Matrix Operations with Gathering in MonetDB

MEMON [1] is an extension of MonetDB [4] that supports matrix operations [5, 3]. Specifically, matrix operations can be applied to an area defined by selected columns and all rows of a relation. As an example consider Table 1, which illustrates relation r with attributes location (L), time (T), accuracy (A), humidity (H), and wind (W). The QR decomposition of MEMON can be applied to all cells in columns H and W.

r

L	T	A	H	W
CH	5am	1	1	3
US	5am	0	8	5
CH	7am	0.5	6	7
US	8am	1	2	3
US	9am	0	1	4
CH	10am	0.5	4	9

r

L	T	A	H	W
CH	5am	1	1	3
US	5am	0	8	5
CH	7am	0.5	6	7
US	8am	1	2	3
US	9am	0	1	4
CH	10am	0.5	4	9

Table 1: Applying gathering before QR decomposition to relation r

There exist many applications where matrix operations must be applied to different parts of a relation independently and where the number of parts is data-dependent and not known apriori. In our example the QR decomposition shall be applied to the light gray cells for CH and to the dark gray cells for the US.

In this project an extension of MEMON with gathering shall be designed, implemented and evaluated. The task is to perform the QR decomposition [2] over values of humidity and wind

gathered by location. To achieve this an integration of the grouping mechanism and matrix operations is needed. In order to perform QR decomposition with gathering over r, the tuples of r must be gathered into groups according to L, the groups must be sorted according to T, and QR must be performed for each group separately.

The work includes the following tasks:

1. Compile MonetDB

2. Implement new gathered BAT algebra operations needed for QR decomposition (gathered addition, gathered subtraction, gathered division, gathered sum, gathered sqrt, gathered multiplication) (about 5 weeks)

    - Operations additionally take and return BAT with OIDs and corresponding group_id values (GID BAT) for each input BAT and compute the result based on groups specified by GID BAT

    - Implementation must be done on all levels of the MonetDB routine: relational tree, statement tree, MAL plan, BAT algebra

3. Adapt the statement tree of implemented QR decomposition to the new operations with GID BATs (about 2 weeks)

4. Empirically evaluate implemented solution (4 weeks)

5. Write a report (approximately 20 pages)

Characteristic of work:

- Most tasks are practical and require programming skills

- Implementation of MonetDB and its extension are in C language

References

[1] Oksana Dolmatova, Nikolaus Augsten, and Michael H. Boehlen. A relational matrix algebra and its implementation in a column store. *Working paper*, 2019.

[2] W. Gander. Algorithms for the QR-Decomposition. Technical report, ETH Zurich, April 1980.

[3] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. John Hopkins University Press, 3rd edition, 1996.

[4] MonetDB. Online MonetDB reference. `https://www.monetdb.org/Home`, 2017.

[5] Quick R. R Matrix Algbera package overview. `http://www.statmethods.net/advstats/matrix.html`, 2017.

University of Zürich
Department of Informatics

Prof. Dr. Michael Böhlen