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, November, 2016

Datenbanktechnologie

Topic: Vector operations computation in MonetDB using LAPACK

The goal of this project is to delegate the computation of simple vector operation from MonetDB to LAPACK library.

MonetDB is a popular column-oriented system. Thus, the execution tree of a query is mapped to BAT (internal MonetDB data structure for attributes) algebra, which consists of a set of vector operations. For example, the expression "x+y" from the following query Q corresponds to vector addition (t is a table with two numeric attributes - x and y):

```
Q: "SELECT x + y from t;"
```

The work includes the following tasks:

1. Download and compile for Ubuntu the last version of MonetDB system.
2. Study the execution level of the system:
   - Study the MAL plan creation on the example of JOIN operation
   - Understand the MAL plan and the execution (used MAL instructions and functions) of the query Q.
3. Include LAPACK (LAPACKE interface) library into a compilation process of MonetDB.
4. Replace the internal MonetDB function of vector addition with an external function which calls LAPACK library instead. The data (input vectors) should be copied.
5. Evaluate the overhead of data copying: compare the run times of internal execution (the data is processed inside MonetDB) and partially external execution (vector operation is

delegated to LAPACK) of Q on the example of tables with different numbers of rows (1000, 10000, 1000000).

Optional part:

- Replace the internal MonetDB function of vector addition with an external function which calls LAPACK library instead. The data must be passed to LAPACK functions without copying.
- Compare the efficiency of internal execution of Q and external (without copying) execution on the example of tables with different numbers of rows (1000, 10000, 1000000).

Supervisor: Oksana Dolmatova
Start date:
End date:

University of Zürich
Department of Informatics

Prof. Dr. Michael Böhlen