



**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, 12. Juli 2016

BSc Vertiefungsarbeit: Linear Optimization in Relational Databases

The concept of optimization refers to solving complex decision problems, involving the selection of values for a number of interrelated variables. The goal of an optimization problem is to maximize (or minimize) an objective function, that quantifies the quality of the decision, subject to a set of constraints that limits the selection of decision variable values. A special case of optimization is Linear Programming, that refers to optimization problems with linear objective functions and constraints that are expressed as a set of linear equalities or inequalities. Linear Programming problems are called Linear Programs (LP).

The Swiss Feed Database¹ contains data about nutrient content of animal feeds. It is used by stock-farmers to find out dietary quality of animal feeds and so determine a good diet for their livestock. The Diet Problem targets on determining the most economical diet that satisfies some basic minimum nutritional requirements. So, a LP highly connected to the Swiss Feed Database is the Diet Problem. A popular algorithm for solving Linear Programming problems is Simplex [1]. The idea of Simplex is to proceed from one feasible solution to another one that results in an improved value of the objective function, until an optimum is reached.

The goal of this project is to study the implementation of Simplex in the context of relational databases as a set of user defined functions of PostgreSQL. Since, Simplex represents an LP in a matrix format and makes heavy use of matrix operations (e.g., multiplication, Gaussian elimination) the implementation shall extend MADlib, which is a free, open source library of in-database analytic methods. MADlib provides a C++ abstraction layer that permits efficient linear algebra operations (using standard libraries like LAPACK) in UDF and the machinery to iteratively call such functions to complete the implementation of data analytics algorithms.

¹<http://feedbase.ch>



Tasks

- Study and understand the linear optimization problem and the simplex algorithm [1].
- Read and understand [2].
- Implement Simplex as a user defined function of PostgreSQL.
- Summarize your work in a short report.

References

- [1] D. G. Luenberger, Y. Ye, The Simplex Method, in: Linear and Nonlinear Programming, Springer, 2016.
- [2] J. M. Hellerstein, C. Ré, F. Schoppmann, D. Z. Wang, E. Fratkin, A. Gorajek, K. S. Ng, C. Welton, X. Feng, K. Li, A. Kumar, The madlib analytics library or MAD skills, the SQL, PVLDB 5 (12) (2012) 1700–1711.

Supervisor: Georgios Garmpis (ggarmpis@ifi.uzh.ch)

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor