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, 30. November 2016

Facharbeit: 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 problems is Linear Programs (LP) [1]. The term LP refers to optimizing a linear objective function under constraints that are expressed as a set of linear equalities or inequalities.

Linear Programs are utilized in the context of feed mills to determine compounds of feeds that satisfy various restrictions, that domain experts specify to optimize the efficiency of the livestock, and have minimal cost. Information about nutrient content of feeds is usually provided by warehouses like the Swiss Feed Database[1]. Relational database systems do not support optimization problems and the usual solution is extracting relevant data and calling external solvers. This is a time consuming and prone to errors process.

The goal of this project is to study algorithm Simplex [2], that is the prevalent algorithm for solving Linear Programs and design and implement an SQL operator in kernel of PostgreSQL that reads from relations representing a Linear Program and solves it using Simplex. As an example, assume the following relations:

[1] `http://feedbase.ch`

costs		
fid	feed	c
2	Hay	20
749	Barley	1
1104	Corn	4
234	Soy	2

diet			
nid	nutrient	l	u
88	Ca	9.3	10.2
141	P	20.4	21.2
37	Fe	12.8	13.1

fact_table		
nid	fid	g
88	2	3
88	749	1
141	1104	4
141	234	5
37	749	2
37	1104	1

that represent a LP of the following form:

$$\begin{aligned} \text{Minimize} \quad & c^T \cdot x \\ \text{Subject to} \quad & A \cdot x \geq b_l \\ & A \cdot x \leq b_u \\ & x \geq 0 \end{aligned} \tag{1}$$

Where,

$$A = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 0 & 4 & 5 \\ 0 & 2 & 1 & 0 \end{bmatrix}, \quad b_l = \begin{bmatrix} 9.3 \\ 20.4 \\ 12.8 \end{bmatrix}, \quad b_u = \begin{bmatrix} 10.2 \\ 21.2 \\ 13.1 \end{bmatrix},$$
$$c^T = \begin{bmatrix} 20 & 1 & 4 & 2 \end{bmatrix},$$

SQL syntax must be extended so that it can receive a query like the following

```
SELECT solution.fid, solution.feed, solution.v
FROM SIMPLEX (COEFFICIENTS fact_table(nid, fid, g) -- Matrix A
             LBOUNDS diet(nid, l) -- Matrix b_l
             UBOUNDS diet(nid,u)  -- Matrix b_u
             COSTS costs(fid, c)) AS solution; -- Matrix c^T
```

and return a relation representing the solution of the LP:

costs		
fid	feed	c
2	Hay	0.92
749	Barley	6.5
1104	Corn	0
234	Soy	4.08

Tasks

- Study and understand the linear optimization problem and the algorithm Simplex [1, 2].
- Study [3, 4] and the source code of PostgreSQL in order to understand the workflow of parsing, analyzing, optimizing, and executing a query.
- Extend PostgreSQL parser to accept operator `SIMPLEX` as described in the example above.
- Extend PostgreSQL analyzer and optimizer so that `SIMPLEX` can be integrated into a query `plan tree`.
- Develop a `plan node` that implements algorithm Simplex.
- Compare the implemented solution with existing external solvers (e.g., GLPK, R, SolveDB) using the dataset of the Swiss Feed Database or benchmark linear problems of `http://www.netlib.org/lp/data/`.
- Summarize you work in a report.

References

[1] D. G. Luenberger, Y. Ye, Linear and Nonlinear Programming, Springer International Publishing, 2016, Ch. Basic Properties of Linear Programs, pp. 11–31.

[2] D. G. Luenberger, Y. Ye, Linear and Nonlinear Programming, Springer International Publishing, 2016, Ch. The Simplex Method, pp. 33–82.

[3] The PostgreSQL Global Development Group, PostgreSQL 9.4.10 Documentation, Ch. 47. Overview of PostgreSQL Internals, pp. 1766–1773.
URL `https://www.postgresql.org/files/documentation/pdf/9.4/postgresql-9.4-A4.pdf`

[4] N. Conway, Introduction to hacking postgresql, pgCon (2007).
URL `http://www.neilconway.org/talks/hacking/ottawa/ottawa_slides.pdf`

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

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor