Diploma/Master Thesis Topics in the Field of Tuning and Self-Tuning in Databases

Offered and supervised by Eike Schallehn:

1. Overview of Tuning and Self-Tuning Storage Approaches for Column-oriented DBMS

Column-oriented DBMS like MonetDB, LucidDB, C-Store, etc. are current solutions suitable especially for analytical databases (e.g. Data Warehouses), where tables are stored in terms of columns rather than as tuples in classical relational DBMS. This includes current research on tuning the storage structures or implementing self-tuning storage solutions, e.g. by Database Cracking. The goal of the thesis would be to provide an overview of the foundations of column-oriented storage and existing approaches for optimizing storage structures for a specific application.

2. Concepts for Multi-Tenancy Data Management and Tuning

Hosted applications serving several customers/organization based on one hardware platform introduce a problem for managing and tuning data within one system for the requirements of several application instances. The goal of the thesis is to investigate existing concepts for the customization of the underlying data management and propose concepts how requirements of different customers regarding the performance of data management can be satisfied.

3. Foundation of Control Theory for Database Self-Tuning

Self-tuning in DBMS is roughly based on concepts borrowed from Control Theory, which applies concepts of engineering and mathematics to describe the behaviour of self-controlled systems. While Control Theory is a well established field describing complex aspects of these systems in a detailed way, database self-tuning still lacks a strong theoretical foundation. The goal of the thesis would be to provide an overview of relevant fields from Control Theory and relate these to current approaches in database self-tuning.

4. Access-balanced Index Structures

Index structures for database systems are typically organized dynamically according to the data set they represent to provide fast access to all data items they manage. This is good for equally distributed accesses on all data items, but causes a huge overhead if certain subsets of data are accessed less frequently than others. Accordingly, indexing data depending on how often they are accessed can provide improved resource consumption and response time. The goal of the thesis would be to development concepts for adjusting existing data structures, e.g.
- Grid-Files
- B+-Trees
- Hashing
- kdB-Trees
- Multi-dimensional clustering (MDC)
- ...

(each of these is a topic for a separate thesis) to provide access-balanced internal organization. Furthermore, a prototypical implementation and evaluation of the concepts would be part of the work on the thesis.

Other topics: I want to encourage students to propose their own topics in the field of database tuning and self-tuning and related areas!