

Workload-Aware Aggregation Strategies for In-Memory Databases

Background and Introduction

In typical mixed OLTP and OLAP workloads, aggregations and joins are the most performance-intensive operations on the database level. Although in-memory column store databases on multi-core systems are ideally suited for calculating aggregations based on large data sets on the fly, there may be situations that favor the materialization of aggregates.

Motivation

The goal of this project is to develop a general cost model for the aggregation operation in column stores. Factors that influence the aggregation performance, for example, are the underlying data structures, the update frequency, the horizontal partitioning of historical data and the utilization of CPU cores. The developed cost function will be evaluated with experiments and detailed measurements. Furthermore, workload-aware caching strategies inside the aggregate operator are explored.

Requirements

- Advanced C++ programming skills,
- Experience in in-memory database development (e.g. HYRISE, SAP TREX, MonetDB, etc.) required
- Excellent English language skills
- Required Courses: TuK II (required), In-Memory Data Structures (optional)

Organisation

A group of three to five (3-5) students may participate in the project. In the first phase (October and November 2011), participants will work on initial measurements to define the cost model and associated parameters. The main steps in implementation are to be executed from December 2011 till March 2012 while the project finishes with a submission of a scientific project report. The project will be carried out at the Hasso Plattner Institute in Potsdam.

Contacts

- Dr. Alexander Zeier,
office-epic@hpi.uni-potsdam.de, +49 (331) 97 99 2 - 560, V-2.12,
- Jens Krüger,
jens.krueger@hpi.uni-potsdam.de, +49 (331) 97 99 2 - 573, V-2.05
- Stephan Müller,
stephan.mueller@hpi.uni-potsdam.de, +49 (331) 97 99 2 - 526, V-2.02