

MorphStore – From an Idea to a Novel In-Memory Column Store Database System

[Abstract]

Dirk Habich
TU Dresden, Database Systems Group
Dresden, Germany
dirk.habich@tu-dresden.de

ABSTRACT

In-memory database systems pursue a main memory-centric architecture approach and assume that all relevant data (base data as well as intermediates) can be fully kept in the main memory of a computer or of a computer network. For OLAP workloads, in-memory column store systems are perfectly suited, because relational tables are organized by column rather than by row and based on that, queries only need to read relevant data columns. In these systems, lightweight integer compression algorithms already play an important role. Aside from reducing the amount of data, compressed data offers several advantages such as less time spent on load and store instructions and a better utilization of the cache hierarchy. Moreover, a direct processing of the compressed data is possible in many cases. However, existing systems only provide a very limited set of compression algorithms for base data. Furthermore, during query processing, these systems only keep the data compressed until an operator cannot process the compressed data directly, whereupon the data is decompressed, but not recompressed. Thus, the full optimization potential is not exploited.

To overcome that, several years ago we had the idea to make lightweight compression a first class citizen in query optimization and query processing, because the effort to access intermediate results is equivalent to the effort to access the base data. Thus, the optimization of intermediate results is interesting and has a high impact on the performance of the query execution. For this domain, we envisioned the continuous use of lightweight compression methods for base data as well as intermediate results and developed a novel compression-aware query processing concept. However, to minimize the overall query execution time, it is important to find a balance between the reduced transfer times and the increased computational effort during query optimization.

In my talk, I will give an overview of our research journey starting from an idea to the development of a novel in-memory column store database system called *MorphStore*.

re. As I will show, the unique features of *MorphStore* are: (i) support of a large variety of lightweight integer compression algorithms, (ii) a continuous handling of compression from base data through intermediate results, (iii) a cost-based decision for the best-suited compression algorithm, and (iv) morphing intermediates from one to another compression scheme to dynamically adapt the physical representation to the changing data characteristics at query run-time.