RAM: array database management through relational mapping

van Ballegooij, A.R.

Citation for published version (APA):
Database technology has not penetrated scientific computing in the same way it has the business world. Yet scientific instruments and computer simulations are creating vast volumes of data to be organized, managed, and analyzed: These are the primary tasks of a database management system. The lack of acceptance of (relational) database technology in science can be attributed to a number of issues: the lack of performance offered by existing database management systems; the mismatch between scientific paradigms and the relational data model; and the unclear benefit of the investments required to switch from existing application frameworks, which at present suffice, to a database driven environment.

Trends in the evolution of database technology are addressing the challenges posed by very large scientific data sets [1]. Yet the interface hurdle imposed by the mismatch between scientific paradigms and the relational model, generally known as the impedance mismatch, remains. We seek a solution for this problem by introducing array data structures in a database environment: Support for multi-dimensional arrays as a primary data type has been argued to be the essential ingredient required for database technology to be embraced by the scientific community [2].

The research objective of this thesis is the realization of an extensible array database architecture using relational mapping and existing relational database technology. Previous efforts toward array-oriented database systems were based on the development of complete array DBMS from the ground up. We opt for an alternative approach based on relational mapping: the translation of operations over non-relational data to relational queries over a relational representation of that data. This approach has been used in object-relational database solutions where object-oriented database functionality is realized by mapping operations to a relational DBMS [3]. Following the success of the object-relational approach, the emergence of XML databases and the XQuery language [4] has lead to various XML-relational mapping schemes [5, 6, 7].

The overall research objective is addressed through the following three goals. The first goal is the specification of an efficient array-mapping scheme: We present an array-oriented data model and show how this data model can be implemented in a
relational environment.

The second goal is to explore the benefit of query optimization at the array level in addition to relational query optimization of translated array queries. We explore the suitability of traditional relational optimization techniques to be applied in the array domain.

The third goal is to show that translation of array operations directly into primitive relational operations allows for more efficient queries than high-level relational query languages would. We explore the specifics of translation to several back-ends and discuss the merits of generating “smart” physical relational query plans directly rather than relying on the relational system to optimize naively generated query plans.

Research is conducted in the context of a prototype relational array mapping system (called the RAM system) [8, 9]. This system is used in a case study and several smaller scale experiments to validate the effectiveness of both the relational mapping scheme and the different optimization techniques developed [10, 11, 12].
Bibliography


