

Query Optimization for Business Intelligence

Nicole Bidoit and Ioana Manolescu

Database team, Université Paris Sud / OAK team, Inria Saclay

Yannick Cras and Bogdan Marinoiu

SAP France

The term *Business Intelligence* covers a set of technologies and tools seeking to facilitate the experience of exploring and exploiting a complex database, to non-expert users, and typically to managers. Such users need to be able to explore the available data, model the pertinent notions of their business model, navigate easily across complex concepts, ask business questions and have their answers computed fast, refine their questions by exposing more details etc. Formal data models and languages for querying complex data have been devised and their properties explored from the angles of database design, query optimization, efficient query evaluation etc. [1, 2]. However, the increasing data warehouse complexity, the multitude of data warehouses typically available in a large company, and the increasing need for ad-hoc warehouse cubes easy to specify by non-specialist, go beyond the expressive power and traditional tools established in data warehousing.

The OAK team brings together Inria and University of Paris Sud faculty working on large-scale management of complex data. The team has extensive background in data models for complex data management applications and a longstanding interest in algebraic techniques for query optimization. We have recently started an exchange with the SAP Research/Advanced Development team of SAP, working to devise Business Intelligence tools on top of structured data. The SAP team has devised an in-house, declarative data management language called DaSL (for *Data Specification Language*) for which an evaluation engine has also been built. The DaSL language bears some similarities on one hand to SQL, and on the other hand to Microsoft's Multidimensional Expressions (MDX); DaSL generalizes and extends both in many ways, and most importantly by being able to return as result of a DaSL expression, not only a table (as SQL does) or a cube (as MDX allows), but more generally a schema of interconnected relations and/or aggregated cubes.

The goal of the internship is to propose a formal model for the computations performed by DaSL queries, identify appropriate equivalence laws between language constructs, and exploit them to propose a framework for optimizing DaSL queries. Time permitting, a prototype implementation of the proposed techniques could also be developed.

We are currently seeking a Master intern with good background in databases and/or logic, to join us on this work. The position is paid by Inria. Our offices are located in the Paris Sud/Inria Saclay building in Gif-sur-Yvette, south of Paris. The position can start at any time, however interested applicants should be available for at least three consecutive months.

How to apply Send to Ioana Manolescu (Ioana.Manolescu@inria.fr) your CV, highlighting the courses taken in databases, and the name of one or two professors that can serve as references.

For more information, see:

- OAK team web site: <http://team.inria.fr/oak>
- Inria Saclay web site: <http://www.inria.fr/en/centre/saclay>
- LRI, CS lab of Université Paris Sud: http://www.lri.fr/index_en.php

References

- [1] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995.
- [2] Jeffrey D. Ullman. *Principles of Database and Knowledge-Base Systems, Volume II*. Computer Science Press, 1989.