

Query Optimization at the Crossroads

Surajit Chaudhuri (Moderator)

Microsoft Research, Redmond

surajitc@microsoft.com

Theme

The ability to optimize queries is considered one of the key enabling technologies for relational databases and is an active area in research and development. There continues to be work in discovering new transformations for SQL. The problem of building extensible query optimizers continues to draw a lot of attention, not only in academic research but also in commercial development.

Despite such strides, it appears that we have no good solutions to some central problems in query optimization that has plagued the field since its early days. The problem of estimating selectivities and calculating accurate costs continues to be hard and becomes a nightmare as we consider Abstract Data Types or distributed and heterogeneous queries. Compared to developing transformations, we have made much less progress in understanding the nature of the search space, interactions among transformations, and the quality of plans generated by an optimizer. In particular, the problem of how to integrate the transformations in a cost-based fashion remains a daunting task. For example, optimizing nested SQL queries efficiently is still an open problem.

Unlike traditional OLTP applications, decision-support systems will provide the real test for optimizers. Emergence of decision support systems has also triggered significant growth in middleware to serve these applications. Many believe that it is the middleware layer, packaged applications and front-end tools that will drive the optimization of queries by using the appropriate optimizer hints. Such a trend will change the landscape of optimization and may lead to diminishing importance of generic SQL optimizers.

We would like to hear from the panelists their views on the state of the art and the future of query optimization in databases:

- Have we made any significant progress on the fundamental problems in query optimization (e.g., exploring search space, selectivity estimation, cost models)?
- Will the problem of optimization only get more ad-hoc as databases embrace new data types?

- How will the optimization task be shared between the middleware and the servers? (e.g., between ROLAP and the back end vendors)?
- How can we objectively measure the quality of query optimizers, if possible orthogonally to specific execution algorithms?

Panelists

- Sanju Bansal (Microstrategy)
- Rick Cole (Redbrick)
- Goetz Graefe (Microsoft)
- Yannis Ioannidis (University of Wisconsin, Madison)
- Hamid Pirahesh (IBM Almaden Research Center)
- Mike Stonebraker (Informix)

Permission to make digital/hard copy of part or all this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD '97 AZ,USA
© 1997 ACM 0-89791-911-4/97/0005...\$3.50