IDEA GROUP PUBLISHING
01 E. Chocolate Avenue, Hershey PA 17033-1117, USA
el: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB7516

Chapter 14

Efficient Query Processing with Structural Join Indexing in an Object Relational Data Warehousing Environment*

Vivekanand Gopalkrishnan and Qing Li
City University of Hong Kong

Kamalakar Karlapalem
University of Science and Technology, Hong Kong

In an Object Relational Data Warehousing (ORDW) environment, the semantics of data and queries can be explicitly captured, represented, and utilized based on is-a and class composition hierarchies, thereby resulting in more efficient OLAP query processing. In this chapter, we show the efficacy in building semantic-rich hybrid data indexes incorporating Structural Join Index Hierarchy (SJIH) on the ORDW views. Given a set of queries, we use a hill-climbing heuristic algorithm to select (near) optimal SJIHs, thereby embedding query semantics into the indexing framework. Finally, by a cost model, we analyze the effectiveness of our approach vis-a-vis the pointer chasing approach.

INTRODUCTION

Data warehouse (DW) equips users with more effective decision support tools by integrating enterprise-wide corporate data into a single repository from which business end-users can run reports and perform ad hoc data analysis (Chaudhuri and Dayal, 1997). As DWs contain enormous amount of data, often from different sources, we need highly efficient Indexing structures (Sarawagi, 1997; Gupta, Harinarayanan, Rajaraman and Ullman, 1997), materialized (stored) Views (Roussopoulos, 1997), and query processing techniques (Gopalkrishnan and

Previously Published in *Challenges of Information Technology Management in the 21st Century*, edited by Mehdi Khosrow-Pour, Copyright © 2000, Idea Group Publishing.
This chapter appears in the book, Data Warehousing and Web Engineering by Shirley Becker.

Karlapalem, 1999) to efficiently answer *on-line analytical processing (OLAP)* queries. Materialized Views represent integrated data based on complex aggregate queries, and should be available consistently and instantaneously. Maintaining the integrity of these Indexes and Views (Mohania and Kambayashi, 1999) imposes a challenging problem when the source data changes frequently, when the size of the DW keeps growing, and/or when the user queries become more and more complex. An extensible framework that can accommodate *dynamic warehousing* (Dayal, 1999) of changing data gracefully, and have adaptive handles for processing OLAP queries efficiently is needed.

In Gopalkrishnan, Li and Karlapalem (1998), we showed that besides establishing a semantically richer framework for multi-dimension hierarchies, the Object Relational View (ORV) model provides excellent support for complex object retrieval. In Gopalkrishnan et al. (1999), we presented the Object Relational Data Warehousing (ORDW) approach to address some of the issues discussed in Gopalkrishnan et al. (1998) on data warehousing. More specifically, we devised a translation mechanism from the star/snowflake schema to an object oriented (O-O) representation. In particular, we have also studied some query processing strategies utilizing vertical partitioning (Karlapalem and Li, 1998) and SJIH indexing techniques (Fung, Karlapalem and Li, 1998) for complex queries on complex objects.

In this chapter, we show the efficiency of building semantic-rich hybrid data indexes incorporating Structural Join Index Hierarchy (SJIH) on the ORDW views. Given a set of queries, we run a hill-climbing heuristic algorithm to select (near) optimal SJIHs, thereby embedding query semantics into the indexing framework. Finally, by a cost model, we analyze the effectiveness of our approach vis-a-vis some other existing techniques.

To put our research in perspective, we review some related work and briefly outline our previous work in the field of ORDW and SJIH on OODBs in the next section. We further motive our study by presenting on the ORDW schema some sample queries whose patterns are classified based on DW operations and by OO concepts. Obtaining an optimal indexing scheme to process this set of queries is the main focus of the following section, where we employ a hill-climbing heuristic algorithm to select a (near) optimal SJIHs. This algorithm is *profiling driven*, and can also be further extended to incorporate other semantics. In the section entitled, SJIH Evaluation, we compare results of retrieval costs using the SJIH versus other retrieval methods. Finally we conclude in the last section, and briefly state our future work.