PREFACE TO THE DATA WAREHOUSE SPECIAL ISSUE

Patrick O'Neil and Kenneth A. Ross

It is difficult to assign a date to the earliest Data Warehousing systems. The MODEL 204 DBMS, implemented to run on the IBM S370, was used in the early 1970's to perform high-performance queries on data extracted from operational systems and stored in a read-only environment [M20487]. Systems such as this were known as [*Executive or Marketing*] *Information Systems* [*EIS*, *MIS, IS*], or sometimes as *Decision Support Systems* [*DSS*]. MODEL 204 schemas at that time were single tables with object-relational-like nested information substituting for relational joins, queried by a procedural language. The concept of a Star Schema, a denormalized Fact table and ancillary Dimension tables to be joined in SQL, seems not to have arisen until about 1991. Teradata incorporated specialized Star-Schema optimization algorithms in 1991 and the original Wal-Mart data warehousing effort which began in 1990 devised a Star Schema (really a Snowflake Schema) for their database. Since then, Wal-Mart has invested over $4 billion supporting this technology [WEST01]. An excellent detailed explanation of practical Data Warehousing Star Schema design is provided by Ralph Kimball [KIMB96].

SQL access is implicit in the Star Schema concept, but a different approach to dimensional data arose in the *OLAP* Model. Products in the original MOLAP, or *Multi-Dimensional OLAP*, category stored data in a *dimensional cube*, and provided data access operations with aggregation semantics that owed nothing to SQL. The OLAP data model was later hosted on relational databases with the same semantics, providing the ROLAP product category. OLAP analysts are empowered to examine aggregated data on any subset of dimensions and explore up and down dimensional hierarchies using appealingly simple commands. An early MOLAP product called *Express*, developed in the 1980's by MDS/IRI, and ultimately acquired by ORACLE, became a model for later OLAP systems. A white paper written for Arbor Essbase in 1993 by E.F. Codd et al. [CCS93] coined the OLAP term, an acronym for *Online Analytical Processing*, and rigorously defined the OLAP Model for the first time.

While Star Schema Data Warehouses originally gained high performance from well-designed database indexes, both MOLAP and ROLAP took the approach of aggregating data grouped by dimensionl hierarchies to speed up queries by orders of magnitude. When this came to the attention of the database research community, it became clear that MOLAP/ROLAP efficiency, leaving specialized semantics apart, could be traced to an approach of precomputing query results, or in database terms *materializing views*, and a tremendous outpouring of papers on materialized views followed. In the last several years, most database vendors have been working to incorporate materialized views in their systems to speed up their SQL queries.

The current special issue on Data Warehousing provides a collection of papers on recent advances in Data Warehouse research. It contains three papers selected from many high-quality manuscripts submitted to this special issue.

The first paper by A. Gupta, I.S. Mumick, J. Rao, K.A. Ross, "Adapting materialized views after redefinitions: Techniques and a performance study," considers techniques for adapting existing materialized views to answer new queries. The authors take a different slant on the well-known View Maintenance problem -- how to modify materialized views to adapte to changed data -- by assuming the data remains stable and the queries change. New queries can often be answered much more quickly by leveraging former results than by evaluating the queries from scratch, especially in a common case where users are exploring data by making small dynamic modifications to pose a sequence of queries. This paper significantly extends a previous article from SIGMOD [GMR95], to which there are already numerous citations (see the DBLP URL, http://www.informatik.uni-trier.de/~ley/db/conf/sigmod/sigmod95-16.html).

The second paper, by Dimitri Theodoratos, "Detecting Redundant Materialized Views in Data Warehouse Evolution," addresses the problem of determining materialized views that may become redundant after a number of changes have been made to the enclosing set of such views. The author provides an extremely clear presentation of query evaluation AND/OR dags to characterize materialized views, and derives a method for detecting redundant views.

The third paper by T.B. Pedersen, C.S. Jensen, and C.E. Dyreson, "A Foundation for Capturing and Querying Complex Multidimensional Data," makes two valuable contributions. First, the paper compares thirteen existing multidimensional OLAP data models from the research literature, using eleven carefully designed modeling requirements. (Actually, one of the multidimensional models compared is Ralph Kimball's Star Schema, included for completeness although it is based on SQL and is not truly representative of the OLAP model). The authors show that each one of the models considered have some shortcomings with respect to their requirements. Second, the paper proposes a multidimensional data model and algebraic query language that addresses their eleven requirements.

Acknowledgements

It has been a pleasure for us to act as Co-Editors for this special issue. We would like to thank the reviewers who helped us in this effort.

References

[CCS93] E.F. Codd, S. B. Codd, and C.T. Salley, "Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate," 1993, Available from http://www.hyperion.com/downloads/OLAPCoddWP.pdf.

[GMR95] Ashish Gupta, Inderpal Singh Mumick, and Kenneth.A. Ross, "Adapting materialized views after redefinitions," SIGMOD Conference, 1995, 211-222.

[KIMB96] Ralph Kimball, "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses," John Wiley & Sons, 1996.

[M20487] "MODEL 204 Architecture and Performance," P.E. O'Neil, 2nd Intl. Workshop on High-Performance Transactions Systems, September 1987, published in Springer-Verlag Lecture Notes in Computer Science 359, 1989, pp. 40-57.

[WEST01] "Data Warehousing Using the Wal-Mart Model," Paul Westerman, Academic Press/Morgan Kaufmann, 2001.