Concept Based Design of Data Warehouses: The DWQ Demonstrators

M. Jarke[1], C. Quix[1], D. Calvanese[2], M. Lenzerini[2], E. Franconi[3], S. Ligoudistianos[4], P. Vassiliadis[4], Y. Vassiliou[4]

[1] Informatik V, RWTH Aachen, Germany, {jarke,quix}@informatik.rwth-aachen.de

[2] Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Italy

[3] Dept. of Computer Science, Univ. of Manchester, United Kingdom

[4] Dept. of Electrical and Computer Eng., National Technical University of Athens, Greece

Abstract

The ESPRIT Project DWQ (Foundations of Data Warehouse Quality) aimed at improving the quality of DW design and operation through systematic enrichment of the semantic foundations of data warehousing. Logic-based knowledge representation and reasoning techniques were developed to control accuracy, consistency, and completeness via advanced conceptual modeling techniques for source integration, data reconciliation, and multi-dimensional aggregation. This is complemented by quantitative optimization techniques for view materialization, optimizing timeliness and responsiveness without losing the semantic advantages from the conceptual approach. At the operational level, query rewriting and materialization refreshment algorithms exploit the knowledge developed at design time. The demonstration shows the interplay of these tools under a shared metadata repository, based on an example extracted from an application at Telecom Italia.

1 Overview of the Demonstration

The demonstration follows roughly the 6-step process indicated in figure 1. In the middle of the figure, light-grey boxes indicate conceptual models which for the designer look like extended ER diagrams but are internally represented in a description logic formalism. The light-grey boxes represent logical (relational or semi-structured or multi-dimensional) schemata. The corresponding tools are organized around the DWQ metadata framework presented in [3] and implemented using the ConceptBase system.

Step 1 and 2: The DWQ approach to source integration is concept-centered, symmetric and incremental: Conceptual models of the sources and of the enterprise model are defined initially independently and related by explicit interschema assertions about which subsumption reasoning can be done. Thus, both a global-as-view and a local-as-view approach to data warehouse design can be suppored.

3. This conceptual modeling approach has been extended to the case where concepts are organized into aggregates along multiple dimensions [2]. Multi-dimensional views over the enterprise conceptual model can be defined to express the interests of DW clients, without losing the advantages of consistency and completeness checking as well as semantic optimization provided by the conceptual modeling approach.

4. These "conceptual data cubes" can be implemented by a MOLAP or ROLAP data model. We follow the latter approach which requires a careful design of an OLAP relational algebra, together with the corresponding rewritings to underlying star schemata [5].

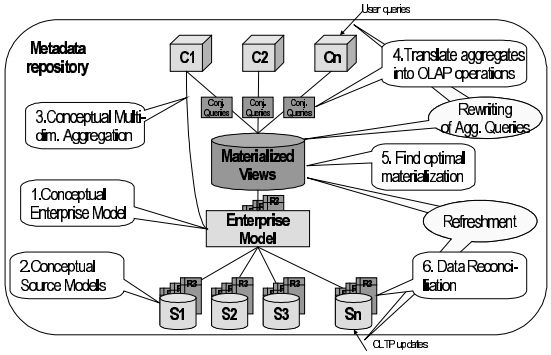

Figure 1: Structure of the demonstration

5. The mapping of multi-dimensional aggregates to ROLAP queries creates a set of view definitions. The materialization of these queries would be the optimal solutions in a query-only DW with hardly any updates. The direct mapping of the enterprise conceptual model to a relational database (ODS) would conversely be best for a very update-intensive DW. Typical DWs have less extreme usage patterns and therefore require a compromise between these two view materialization strategies. We demonstrate a combinatorial optimization algorithm for solving this problem [4].

6. The resulting optimal design is now implemented by data integration and reconciliation algorithms again derived from the conceptual perspective. The views to be materialized are initially defined over the ODS relations; there can be several qualitatively different, possibly conflicting ways to actually materialize these ODS relations from the existing sources which are generated by a further set of rewritings derived from steps 1 and 2 [1].

References

[1] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, R. Rosati: A principled approach to data integration and reconciliation in data warehousing. *Workshop on Design and Management of Data Warehouses (DMDW'99)*, 1999.

[2] E. Franconi, U. Sattler: A data warehouse conceptual data model for multidimensional aggregation. *Workshop on Design and Management of Data Warehouses (DMDW'99)*, 1999.

[3] M. Jarke, M.A. Jeusfeld, C. Quix, P. Vassiliadis: Architecture and quality in data warehouses: An extended repository approach. *Information Systems*, 24(3), pp. 229-253, 1999.

[4] D. Theodoratos, T. Sellis: Data warehouse configuration. *23^{rd} Conf. Very Large Databases (VLDB)*, pp. 126-135, 1997.

[5] P. Vassiliadis: Modeling multidimensional databases, cubes and cube operations. *10^{th} Conf. Scientific and Statistical Database Management (SSDBM)*, pp. 53-62, 1998.