Chapter 2.8
Physical Modeling of Data Warehouses Using UML Component and Deployment Diagrams:
Design and Implementation Issues

Sergio Luján-Mora
University of Alicante, Spain

Juan Trujillo
University of Alicante, Spain

ABSTRACT

Several approaches have been proposed to model different aspects of a Data Warehouse (DW) during recent years, such as the modeling of a DW at the conceptual and logical level, the design of the ETL (Extraction, Transformation, Loading) processes, the derivation of the DW models from the enterprise data models, and customization of a DW schema. At the end of the design, a DW has to be deployed in a database environment, requiring many decisions of a physical nature. However, few efforts have been dedicated to the modeling of the physical design of a DW from the early stages of a DW project. In this article, we argue that some physical decisions can be taken from gathering main user requirements. In this article, we present physical modeling techniques for DWs using the component diagrams and deployment diagrams of the Unified Modeling Language (UML). Our approach allows the designer to anticipate important physical design decisions that may reduce the overall development time of a DW, such as replicating dimension tables, vertical and horizontal partitioning of a fact table, and the use of particular servers for certain ETL processes. Moreover, our approach allows the designer to cover all main design phases of DWs from the conceptual modeling phase to the final implementation. To illustrate our techniques, we show a case study that is implemented on top of a commercial DW management server.

INTRODUCTION

Data warehouses (DW) provide organizations with historical information to support a decision. It is widely accepted that these systems are based on multidimensional (MD) modeling. Thus, research on the design of a DW has been addressed mainly from the conceptual and logical point of view through multidimensional (MD) data models (Abelló, Samos & Saltor, 2001; Blaschka, Sapia, Höfling & Dinter, 1998). However, to the best of our knowledge, there are no standard methods or models that allow us to model all aspects of a DW. Moreover, as most of the research efforts in designing and modeling DWs have been focused on the development of MD data models, the attention to the physical design of DWs from the early stages of a DW project has been very little. Nevertheless, the physical design of a DW is of a vital importance and highly influences the overall performance of the DW (Nicola & Rizvi, 2003) and the following maintenance; moreover, a well-structured physical design policy can provide the perfect roadmap for implementing the whole warehouse architecture (Triantafillakis, Kanellis & Martakos, 2004).

Although in some companies the same employee may take on both the role of DW designer and DW administrator, other organizations may have separate people working on each task. Regardless of the situation, modeling the storage of the data and how it will be deployed across different components such as servers and drives helps to implement and maintain a DW. In traditional software products or transactional databases, physical design or implementation issues are not considered until the latest stages of a software project. Then, if the final product does not satisfy user requirements, designers do a feedback, taking into consideration (or at least bearing in mind) some final implementation issues.

Nevertheless, due to the specific characteristics of DWs, we can address several decisions regarding the physical design of a DW from the early stages of a DW project, with no need to leave them until the final implementation stage. DWs, mainly built for analytical reasons, are queried by final users trying to analyze historical data on which they can base their strategy decisions. Thus, the performance measure for DWs is the amount of queries that can be executed instead of the amount of processes or transactions that it supports. Moreover, the kinds of queries on DWs are demonstrated to be much more complex than the queries normally posed in transactional databases (Kimball, 1996; Poe, Klauer & Brobst, 1998). Therefore, poor performance of queries has a worse impact in DWs than in transactional databases. Furthermore, the set of OLAP (Online Analytical Processing) operations that users can execute with OLAP tools on DWs depends very much on the design of the DW (i.e., on the multidimensional model underneath) (Sapia, 1999; Trujillo, Palomar, Gómez & Song, 2001).

Based on our experience in real-world DW projects, physical storage and query performance issues can be discussed in the early stages of the project. The reason is that in DW projects, final users, analysts, business managers, DW designers, and database administrators participate at least in the first meetings. Therefore, we believe that some decisions on the physical design of DWs can be made in the beginning. Some examples of these decision are as follows: (1) the size and the speed of the hard disk needed to deal with the fact table and the corresponding views; (2) a coherent partitioning of both fact and dimension tables based on data and user requirements; (3) the estimation of the workload needed and the time boundaries to accomplish it. Based on our experience, we believe that making these decisions in the early stages of a DW project will reduce the total development time of the DW.

At this point, we must point out that we are not suggesting that the conceptual modeling of a DW take into account physical issues. Instead, we advocate that the physical aspects and following implementation details from the conceptual

Related Content

An Intelligent Support System Integrating Data Mining and Online Analytical Processing

Rahul Singh, Richard T. Redmondand Victoria Yoon (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2964-2977).*

www.irma-international.org/chapter/intelligent-support-system-integrating-data/7815

Spectral Methods for Data Clustering

Wenyuan Li (2005). *Encyclopedia of Data Warehousing and Mining (pp. 1037-1042).*

www.irma-international.org/chapter/spectral-methods-data-clustering/10749

Data Warehousing and OLAP

Jose Hernandez-Orallo (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 169-178).*

www.irma-international.org/chapter/data-warehousing-olap/7639

Mining Historical XML

Qiankun Zhaoand Sourav Saha Bhowmick (2005). *Encyclopedia of Data Warehousing and Mining (pp. 800-804).*

www.irma-international.org/chapter/mining-historical-xml/10706

Entity Resolution on Graph Data Set

 (2014). *Innovative Techniques and Applications of Entity Resolution (pp. 171-194).*

www.irma-international.org/chapter/entity-resolution-on-graph-data-set/103248