Chapter VI

Some Issues in Design of Data Warehousing Systems

Ladjel Bellatreche and Kamalakar Karlapalem
University of Science Technology, Clear Water Bay Kowloon, Hong Kong, People's Republic of China

Mukesh Mohania
Western Michigan University, USA

INTRODUCTION

Information is one of the most valuable assets of an organization, and when used properly can assist intelligent decision-making that can significantly improve the functioning of an organization. Data warehousing is a recent technology that allows information to be easily and efficiently accessed for decision-making activities. On-line analytical processing (OLAP) tools are well studied for complex data analysis. A data warehouse is a set of subject-oriented, integrated, time varying and non-volatile databases used to support the decision-making activities (Inmon, 1992).

The conceptual architecture of a data warehousing system is shown in Figure 1. The data warehouse creation and management component includes software tools for selecting data from information sources (which could be operational, legacy, external, etc., and may be distributed, autonomous and heterogeneous), cleaning, transforming, integrating and propagating data into the data warehouse. It also refreshes the warehouse data and meta-data when source data is updated. This component is also responsible for managing the warehouse data, creating indices on...
data tables, data partitioning and updating meta-data. The warehouse data contains
the detail data, summary data, consolidated data and/or multidimensional data.

The meta-data is generally held in a separate repository. The meta-data contains
the informational data about the creation, management and usage of the
data warehouse. It serves as a bridge between the users of the warehouse and the
data contained in it. The warehouse data is also accessed by the OLAP server
to present the data in a multidimensional way to the front-end tools (such as
analytical tools, report writers, spreadsheets and data-mining tools) for analysis
and informational purposes. Basically, the OLAP server interprets client que-
ries (the client interacts with front-end tools and passes these queries to the
OLAP server) and converts them into complex SQL queries required to access
the warehouse data. It might also access the data from the primary sources if the
client’s queries need operational data. Finally, the OLAP server passes the
multidimensional views of data to the front-end tools, and these tools format the
data according to the client’s requirements.

There are two approaches to creating the warehouse data - bottom-up and top-
down. In a bottom-up approach, the data is obtained from the primary sources based
on the data warehouse applications and a profile of the likely queries which is
typically known in advance. The data is then selected, transformed, and integrated
by data acquisition tools. In a top-down approach, the data is obtained from the
primary sources whenever a query is posed. In this case, the warehouse system
determines the primary data sources in order to answer the query. These two
approaches are similar to eager and lazy approaches discussed in Widom (1995).
The bottom-up approach is used in data warehousing because user queries can be
answered immediately and data analysis can be done efficiently, since data will
always be available in the warehouse. Hence, this approach is feasible and improves
the performance of the system. Another approach is a hybrid approach, which
combines aspects of the bottom-up and top-down approaches. In this approach,
some data is stored in a warehouse, and other data can be obtained from the primary
sources on demand (Hull and Zhou, 1999).
Semantic Integrity Constraint Checking for Multiple XML Databases
www.igi-global.com/article/semantic-integrity-constraint-checking-multiple/3360?camid=4v1a