COMPARISON OF INTEGRATING HETEROGENEOUS DATA: USING XML AND OLAP TECHNOLOGIES

PANKAJ NAGAR & SUMIT SRIVASTAVA

Abstract

On-Line Analytical Processing (OLAP) enables analysts to gain insight into data through fast and interactive access to a variety of possible views on information, organized in a *dimensional* model. The demand for data integration is rapidly becoming larger as more and more information sources appear in modern enterprises. In the data warehousing approach, selected information is extracted in advance and stored in a repository. This approach is used because of its high performance. However, in many situations a logical (rather than physical) integration of data is preferable. Previous web-based data integration efforts have focused almost exclusively on the logical level of data models, creating a need for techniques focused on the conceptual level. Also, integration techniques for web-based data have not addresses the special needs of OLAP tools such as handling dimensions with hierarchies. Extensible Markup Language (XML) is fast becoming the new standard for data representation and exchange on the World Wide Web. The rapid emergence of XML data on the web, e.g., business-to-business (B2B) ecommerce, is making it necessary for OLAP and other data analysis tools to handle XML data as well as traditional data formats.. Based on a real-world case study, this paper presents an approach to the conceptual specification of OLAP DB's based on web data. Unlike previous work, these approaches takes special OLAP issues such as dimension hierarchies and correct aggregation of data into account. Further the architecture issue in traditional OLAP along with that for XML technologies will be addressed further.

Keywords: Heterogeneous data structure, XML, XML Schema, X Query, Multi-Dimensional Schema