Chapter 8.8
Evaluating XML–Extended OLAP Queries Based on a Physical Algebra

Xuepeng Yin
Aalborg University, Denmark

Torben Bach Pedersen
Aalborg University, Denmark

ABSTRACT

In today's OLAP systems, physically integrating fast-changing data (e.g., stock quotes) into a cube is complex and time-consuming. The data is likely to be available in XML format on the World Wide Web (WWW); thus, instead of physical integration, making XML data logically federated with OLAP systems is desirable. In this article, we extend previous work on the logical federation of OLAP and XML data sources by presenting simplified query semantics, a physical query algebra, and a robust OLAP-XML query engine, as well as the query evaluation techniques. Performance experiments with a prototypical implementation suggest that the performance for OLAP-XML federations is comparable to queries on physically integrated data.

INTRODUCTION

Online Analytical Processing (OLAP) technology enables data warehouses to be used effectively for online analysis, providing rapid responses to iterative complex analytical queries. Usually an OLAP system contains a large amount of data, but dynamic data (e.g., stock prices) is not handled well in current OLAP systems. To an OLAP system, a well-designed dimensional hierarchy and a large quantity of pre-aggregated data are the keys. However, trying to maintain these two factors when integrating fast-changing data physically into a cube is complex and time-consuming, or even impossible. However, the advent of XML makes it very possible that the data is available in XML format on the WWW; thus, making XML data accessible to OLAP systems is greatly needed.

Our overall solution is to logically federate the OLAP and XML data sources. This approach decorates the OLAP cube with virtual dimensions, allowing selections and aggregations to be performed over the decorated cube. In this article, we describe the foundation of a robust federation query engine with the focus on query evaluation, which includes the query semantics, a physical algebra, and query evaluation techniques. First, a query semantics that simplifies earlier definitions (Pedersen, Riis, & Pedersen, 2002) is proposed. Here, redundant and repeated logical operators are removed, and a concise and compact logical query plan can be generated after a federation query is analyzed. Second, a physical query algebra, unlike the previous logical algebra, is able to model the real execution tasks of a federation query. Here, all concrete data retrieval and manipulation operations in the federation are integrated. This means that we obtain a much more precise foundation for performing query optimization and cost estimation. Third, the detailed description of the query evaluation introduces how the modeled execution tasks of a query plan are performed, including the concrete evaluation algorithms and techniques for each physical operator and the general algorithm that organizes and integrates the execution of the operators in a whole plan. In addition, algebra-based query optimization techniques, including the architecture of the optimizer, cost estimation of physical operators, and plans, also are presented. Experiments with the query engine suggest that the query performance of the federation approach is comparable to physical integration.

There has been a great deal of previous work on data integration; for instance, on relational data (Hellerstein, Stonebraker, & Caccia, 1999; IBM, n.d.; Oracle, 2005), semi-structured data (Chawathe et al., 1994; Nicolle, Yétongnon, & Simon, 2003), a combination of relational and semi-structured data (Goldman & Widom, 2000; Lahiri, Abiteboul, & Widom, 1999), a combination of object-oriented and semi-structured data (Bae, Kim, & Huh, 2003), and an integration of remote data warehouses by hypermaterialized views (Triantafillakis, Kanellis, & Martakos, 2004). However, none of these handles the advanced issues related to OLAP systems (e.g., automatic and correct aggregation and dimensions with hierarchies). Some work concerns integrating OLAP and object databases (Gu, Pedersen, & Shoshani, 2000; Pedersen, Shoshani, Gu & Jensen, 2000), which demands rigid schemas (i.e., data is represented by classes and connected by complex associations). In comparison, using XML as data source, as we do, enables the federation to be applied on any data as long as the data allows XML wrapping, greatly enlarging the applicability. Our pervious work (Pedersen et al., 2002) presents a logical federation of OLAP and XML systems, where a logical algebra defines the query semantics and a partial straightforward implementation. In comparison, this article takes a further step and introduces a physical query algebra, and the query evaluation and optimization techniques are implemented in a robust OLAP-XML federation query engine.

This article makes the following novel contributions, which are extensions to the most related work (Yin & Pedersen, 2004). First, a robust OLAP-XML federation query engine is presented along with its GUI. Second, query evaluation is described in detail. Here, the evaluation of the physical operators, including component data retrieval, component query construction, and optimization of constructed queries, is presented. In addition, an example is given to show the query evaluation process step-by-step. Third, we also show the cost estimation algorithm of physical plans and how the costs of the physical operators are approximated.

The article is organized as follows. We first introduce an example cube and XML document used in the illustrations. Then, an overview of the overall architecture of the federation and the federation query language is given. We define the formal representation of the federation and the federation queries using the logical algebra.

Related Content

Ensuring Customised Transactional Reliability of Composite Services
Sami Bhiri, Walid Gaaloul, Claude Godart, Olivier Perrin, Maciej Zarembaand Wassim Derguech (2013).
Innovations in Database Design, Web Applications, and Information Systems Management (pp. 203-232).
www.irma-international.org/chapter/ensuring-customised-transactional-reliability-composite/74394

The Need to Consider Hardware Selection when Designing Big Data Applications Supported by Metadata
Nathan Regola, David A. Cieslakand Nitesh V. Chawla (2014). *Big Data Management, Technologies, and Applications (pp. 381-396).*
www.irma-international.org/chapter/the-need-to-consider-hardware-selection-when-designing-big-data-applications-supported-by-metadata/85464

Handling Imbalanced Data With Weighted Logistic Regression and Propensity Score Matching methods: The Case of P2P Money Transfers
Lavlin Agrawal, Pavankumar Mulgundand Raj Sharman (2024). *Journal of Database Management (pp. 1-37).*
www.irma-international.org/article/handling-imbalanced-data-with-weighted-logistic-regression-and-propensity-score-matching-methods/335888

Using Ontology Languages for Conceptual Modeling
Palash Bera, Anna Krasnoperovaand Yair Wand (2010). *Journal of Database Management (pp. 1-28).*
www.irma-international.org/article/using-ontology-languages-conceptual-modeling/39114

Operators for Multidimensional Aggregate Data
Maurizio Rafanelli (2003). *Multidimensional Databases: Problems and Solutions (pp. 116-165).*
www.irma-international.org/chapter/operators-multidimensional-aggregate-data/26967