Chapter LVII
Online Analytical Processing and Data–Cube Technologies

Lixin Fu
University of North Carolina, Greensboro, USA

Wenchen Hu
University of North Dakota, USA

INTRODUCTION

Since the late '80s and early '90s, database technologies have evolved to a new level of applications: online analytical processing (OLAP), where executive management can make quick and effective strategic decisions based on knowledge in terms of queries against large amounts of stored data. Some OLAP systems are also regarded as decision support systems (DSSs) or executive information systems (EIS). The traditional, well-established online transactional processing (OLTP) systems such as relational database management systems (RDBMS) mainly deal with mission-critical daily transactions. Typically, there are a large number of short, simple queries such as lookups, insertions, and deletions. The main focus is transaction throughput, consistency, concurrency, and failure recovery issues. OLAP systems, on the other hand, are mainly analytical and informational. OLAP systems are usually closely coupled with data warehouses, which can contain very large data sets that may include historical data as well as data integrated from different departments and geographical locations. So the sizes of data warehouses are usually significantly larger than common OLTP systems. In addition, the workloads of OLAP are quite different from those of traditional transaction systems: The queries are unpredictable and much more complicated. For example, an OLAP query could be, "For each type of car and each manufacturer, list market share change in terms of car sales between the first quarter of 2005 and the first quarter of 2006." The purpose of these queries is not for the daily operational maintenance of data; instead, it is for deeper knowledge from data used for decision support.

Closely related to OLAP technologies are data-cube technologies. A data cube is a data model that allows data to be viewed in multiple dimensions. For example, in a large retail database (e.g., of Wal-Mart), there may be several dimensions such as time, location, store, and product that track particular sales of certain dollar amounts. We may need to know the sales by time; by location; by store; by categories; by time and location; by store, time, and category; and so forth. Actually, we may navigate by any combination of the dimensions. Furthermore, these dimensions may be hierarchical (e.g., time with day-month-year hierarchy, location with city-state-region hierarchy, product with categories, and so on). Data cubes will allow users to navigate different parts of the data at different granularity levels. For instance, users can report sales grouped by month and state, or query the total sales of product category *electronics* in the southeastern region.

OLAP and data-cube technologies have broad applications. With such decision support systems in place, enterprises can gain competitive advantages by making timely and informative business decisions based on the embedded patterns computed from large data sets. Without data warehouse and OLAP systems, the data would be separated and may have different formats. It is also very hard to maintain the currency of the databases. To answer an analytical query, we may have to coordinate with different databases stored and run on different systems. In coupling with formatting problems, data analysts may have to do it manually, which greatly reduces the overall system performance. Even if we can somehow evaluate the complex queries, the evaluation and coordination process may interfere with the existing operational OLTP systems. In view of these difficulties, the data warehousing and data-cube technologies are needed for satisfying modern data analysis requirements.

Both the total revenues in this industry and the number of vendors have experienced explosive growth in the recent past years. In academia, OLAP and data cube pose new challenges and many interesting research problems. One of the challenges is that the queries can be very complex in contrast to normal short queries in relational database systems. For example, in car sales data warehouses, one might want to know which venders or types of cars are sold fast in terms of their market share change in the past 5 years. To answer such queries, we often need to scan a large part of the overwhelming data sets. It is very challenging to evaluate complex queries against large data sets with hundreds of gigabytes (1 gigabyte = 10^9 bytes) or terabytes (1 terabyte = 1,000 gigabytes) of data. Another challenge is the notorious problem of "the curse of dimensionality." That is, when the number of dimensions of data increases, the queries may be more complex and harder to evaluate. No existing method or system so far performs efficiently on large data sets of high dimensionality.

BACKGROUND

An OLAP system is usually deployed as a part of a greater warehousing system. Data from different sources are extracted, cleaned, transformed, integrated, and loaded into a central place called a data warehouse (Chaudhuri & Dayal, 1997). The centralized data can further be partitioned into subsets of data corresponding to different departments. These smaller data sets are called data marts that are suitable for particular queries. The back-end processing steps are important because the source data more than likely contain noisy and even incorrect data, and the formats and schemas of different sources are often different. The data quality and structure of the integrated data from different sources are critical in whether we can perform meaningful data analysis. The OLAP server is a front-side data analysis tool. OLAP works on top of the warehouse database to provide sophisticated querying capabilities.

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/online-analytical-processing-data-cube/21284

Related Content

A Comprehensive Framework Towards Information Sharing Between Government Agencies
Akhilesh Bajajand Sudha Ram (2007). *International Journal of Electronic Government Research (pp. 29-44).*
www.irma-international.org/article/comprehensive-framework-towards-information-sharing/2029

A Reference Architecture for Context-Aware Intelligent Traffic Management Platforms
Zeenat Rehena, Marijn Janssenand Samiran Chattopadhyay (2018). *International Journal of Electronic Government Research (pp. 65-79).*
www.irma-international.org/article/a-reference-architecture-for-context-aware-intelligent-traffic-management-platforms/226268

Conquering the Invasion of Technology: E-Government and Property Tax Collections in the Stonewall County Tax Commissioner's Of?ce
Saundra J. Reinke (2007). *Case Studies on Digital Government (pp. 195-200).*
www.irma-international.org/chapter/conquering-invasion-technology/6193

A Location-Aware Architecture for an IoT-Based Smart Museum
Giuseppe Del Fiore, Luca Mainetti, Vincenzo Mighali, Luigi Patrono, Stefano Alletto, Rita Cucchiaraand Giuseppe Serra (2016). *International Journal of Electronic Government Research (pp. 39-55).*
www.irma-international.org/article/a-location-aware-architecture-for-an-iot-based-smart-museum/162737

The Techonomic Divide
Alex Pettitand Anthony Caranna (2007). *Case Studies on Digital Government (pp. 314-326).*
www.irma-international.org/chapter/techonomic-divide/6201