

Online Analytical Processing Systems

Rebecca Boon-Noi Tan

Monash University, Australia

INTRODUCTION

Since its origin in the 1970's research and development into databases systems has evolved from simple file storage and processing systems to complex relational databases systems, which have provided a remarkable contribution to the current trends or environments. Databases are now such an integral part of day-to-day life that often people are unaware of their use. For example, purchasing goods from the local supermarket is likely to involve access to a database. In order to retrieve the price of the item, the application program will access the product database. A database is a collection of related data and the database management system (DBMS) is software that manages and controls access to the database (Elmasri & Navathe, 2004).

BACKGROUND

Data Warehouse

A data warehouse is a specialized type of database. More specifically, a data warehouse is a "repository (or archive) of information gathered from multiple sources, stored under a unified schema, at a single site" (Silberschatz, Korth, & Sudarshan, 2002, p. 843). Chaudhuri and Dayal (1997) consider that a data warehouse should be separately maintained from the organization's operational database since the functional and performance requirements of online analytical processing (OLAP) supported by data warehouses are quite different from those of the online transaction processing (OLTP) traditionally supported by the operational database.

OLAP Versus OLTP

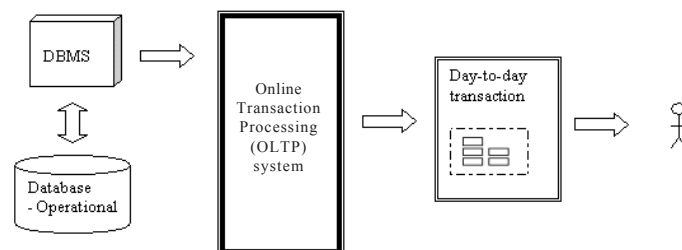
Two reasons why traditional OLTP is not suitable for data warehousing are presented: (a) Given that operational databases are finely tuned to support known OLTP workloads, trying to execute complex OLAP queries against the operational databases would result in unacceptable performance. Furthermore, decision support requires data that might be missing from the operational databases; for instance, understanding trends or making predictions requires historical data, whereas operational databases store only current data. (b) Decision support usually requires consolidating data from many heterogeneous sources: these might include external sources such as stock market feeds, in addition to several operational databases. The different sources might contain data of varying quality, or use inconsistent representations, codes and formats, which have to be reconciled.

Traditional Online Transaction Processing (OLTP)

Traditional relational databases have been used primarily to support OLTP systems. The transactions in an OLTP system usually retrieve and update a small number of records accessed typically on their primary keys. Operational databases tend to be hundreds of megabytes to gigabytes in size and store only current data (Ramakrishnan & Gehrke, 2003).

Figure 1 shows a simple overview of the OLTP system. The operational database is managed by a conventional database management system (DBMS). OLTP is designed for day-to-day operations. It provides a real-time response. Examples include Internet banking and online shopping.

Figure 1. OLTP system



Online Analytical Processing (OLAP)

OLAP is a term that describes a technology that uses a multi-dimensional view of aggregate data to provide quick access to strategic information for the purposes of advanced analysis (Ramakrishnan & Gehrke, 2003).

OLAP supports queries and data analysis on aggregated databases built in data warehouses. It is a system for collecting, managing, processing and presenting multidimensional data for analysis and management purposes (Figure 2). There are two main implementation methods to support OLAP applications: relational OLAP (ROLAP) and multidimensional OLAP (MOLAP).

ROLAP

Relational online analytical processing (ROLAP) provides OLAP functionality by using relational databases and familiar relational query tools to store and analyse multidimensional data (Ramakrishnan & Gehrke, 2003). Entity Relationship diagrams and normalization techniques are popularly used for database design in OLTP environments. However, the database designs recommended by ER diagrams are inappropriate for decision support systems where efficiency in querying and in loading data (including incremental loads) are crucial. A special schema known as a star schema is used in an OLAP environment for performance reasons (Martyn, 2004). This star schema usually consists of a single fact table and a dimension table for each dimension (Figure 3).

MOLAP

Multidimensional online analytical processing (MOLAP) extends OLAP functionality to multidimensional database management systems (MDBMSs). A MDBMS uses

special proprietary techniques to store data in matrix-like n-dimensional arrays (Ramakrishnan & Gehrke, 2003).

The multi-dimensional data cube is implemented by the arrays with the dimensions forming the axes of the cube (Sarawagi, 1997). Therefore, only the data value corresponding to a data cell is stored as direct mapping. MOLAP servers have excellent indexing properties due to the fact that looking for a cell is simple array lookups rather than associative lookups in tables. But unfortunately it provides poor storage utilization, especially when the data set is sparse.

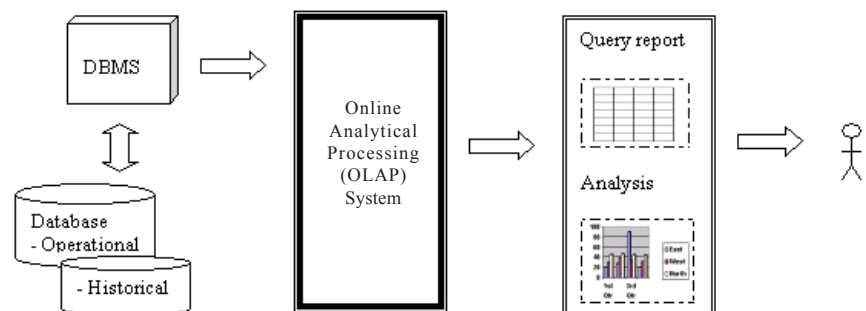
In a multi-dimensional data model, the focal point is on a collection of numeric measures. Each measure depends on a set of dimensions. For instance, the measure attribute is amt as shown in Figure 4. Sales information is being arranged in a three-dimensional array of amt. Figure 4 shows that the array only shows the values for single L# value where $L\# = L001$, which presented as a slice orthogonal to the L# axis.

Cube-By Operator

In decision support database systems, aggregation is a commonly used operation. As previously mentioned, current SQL can be very inefficient. Thus, to effectively support decision support queries in OLAP environment, a new operator, Cube-by was proposed by (Gray, Bosworth, Lyaman, & Pirahesh, 1996). It is an extension of the relational operator Group-by. The Cube-by operator computes Group-by corresponding to all possible combinations of attributes in the Cube-by clause.

In order to see how a data cube is formed, an example is provided in Figure 5. It shows an example of data cube formation through executing the cube statement at the top left of the figure. Figure 5 presents two way of presenting the aggregated data: (a) a data cube, and (b) a 3D data cube in table form.

Figure 2. OLAP system



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/onling-analytical-processing-systems/10720

Related Content

Commercial and Open Source Business Intelligence Platforms for Big Data Warehousing

Jorge Bernardino, Joaquim Lapa and Ana Almeida (2019). *Emerging Perspectives in Big Data Warehousing* (pp. 158-181).

www.irma-international.org/chapter/commercial-and-open-source-business-intelligence-platforms-for-big-data-warehousing/231012

Data Mining Techniques and Medical Decision Making for Urological Dysfunction

N. Siraam, V. Natasha and H. Kaur (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2506-2516).

www.irma-international.org/chapter/data-mining-techniques-medical-decision/7779

XML Object Identification

(2014). *Innovative Techniques and Applications of Entity Resolution* (pp. 140-170).

www.irma-international.org/chapter/xml-object-identification/103247

Data Mining with Cubegrades

Amin A. Abdulghani (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 288-292).

www.irma-international.org/chapter/data-mining-cubegrades/10609

Neural Networks for Prediction and Classification

Kate A. Smith (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 865-869).

www.irma-international.org/chapter/neural-networks-prediction-classification/10718