Curso de Engenharia de Sistemas e Informática - 5º Ano

Anexo à Ficha T. Prática n.º 1

True Relational OLAP:

The Future of Decision Support

By Michael J. Saylor, Manish G. Acharya, and Robert G. Trenkamp

Article published in **DATABASE JOURNAL**
November-December 1995

Once upon a time, in a business very very close, an executive met a leading management guru and asked: "Oh guru, how will I ensure success for my company?" And the guru replied "Information." And then the executive asked "Oh guru, what should I fear most?" And the guru replied "Data."

Introduction

Today an organization needs to utilize all the information available to create and maintain a competitive advantage. Businesses are being bombarded with immense quantities of data from internal systems, customers, and external data services, and the race is on to utilize this data to make the correct decision before the competitors do. Is the two-for-one sale the best promotional strategy? What inventory level is appropriate for Brazil? What is my optimal product portfolio? So organizations have adopted the mantra of Data Warehousing and are implementing these warehouses with the fervor once reserved for CASE and AI, but data warehousing is not decision support. There are a set of analysis filters that data needs to pass through before it becomes information, and this leads to a set of requirements for the data analysis tools.In this article, we will define these requirements, consider the traditional approaches to data access, and outline the new hybrid data access approach that is revolutionizing the decision support/data warehousing marketplace.

Decision Support Requirements

To meet the rapidly growing data delivery and analytical demands of today's increasingly sophisticated users, decision support systems (DSS) face a stringent set of requirements. Users want analytical complexity without sacrificing usability; they want flexibility without sacrificing performance; and they want scalability without sacrificing functionality. Their demands can be summarized into the seven requirements detailed below.

• Sophisticated Analytical Functionality - To determine the optimal portfolio mix or the most effective promotion or the cost-optimized production schedule, users need advanced analytical capabilities that enable them to make informed decisions. For example, to determine purchase quantities for white dresses, a merchandising buyer must analyze sell-through, year-over-year sales, and sales contribution for white dresses. DSS solutions must be able to perform sophisticated analytical calculations in a timely manner.

• Support for a Wide Range of Users - The same data needs to be analyzed differently by different classes of users within an organization. Often, opening the data warehouse to users throughout an organization means opening the data warehouse to users of all skill levels and technical ability. Decision support systems must scale to meet both high level executive needs and detailed ad hoc query requirements.

• High End-User Productivity - Decision support systems should allow users to spend their time analyzing information and making decisions, not defining and retrieving information from the data warehouse. Subsequently, DSS solutions should provide the ability to view data multidimensionally, share work with others and integrate with existing productivity tools like spreadsheets and electronic mail systems.

• Support for Very Large Data Warehouses - The amount of information available to organizations is increasing at a tremendous rate. Data warehouses that contain eighteen months of item-level data, market information from external data sources, and characteristic information about customers, suppliers and competitors can quickly grow to hundreds of gigabytes, or even terabytes in size. An informed decision, by definition, is one made by considering all the information available, and effective DSS solutions must be capable of analyzing information at any level of detail throughout the entire data warehouse.

• Support for a Large Number of Users - To counter competitive threats and focus promotion, sales and service initiatives require large-scale deployment of decision support systems. Deployment to hundreds (and thousands) of users allow businesses to provide people throughout an organization with the ability to make the "right" decisions at the right time. DSS applications need to be architected to support large numbers of users; for example, they must provide facilities for multi-level security, query governing, and background processing.

• Easy to Build and Maintain - Information systems and data delivery requirements are continually being refined as the business climate changes. Consequently, decision support systems must be easy for organizations to develop and maintain.

• Support for an Open Architecture - Decision support systems should be based on an open design, conform to industry standards, and integrate cleanly with the existing and future IS infrastructure. Changes in other systems, such as a data warehouse platform or a database driver, should not affect the functionality of the DSS.

Ad hoc Query Tools and Report Writers

For years, companies have been building general purpose decision support systems with relational databases and ad hoc query/reporting tools. Through graphical front-ends using point-and-click construction techniques, these ad hoc query/reporting tools automate the task of writing the SQL (Structured Query Language) used to access relational databases.

These tools work well for basic data retrieval and frequently provide a number of data formatting and report writing options. Their GUI interfaces expand the group of users who can directly access information in the data warehouse beyond the IS department to the technically-savvy decision makers in an organization. The main advantage of the ad hoc query tools is their endorsement of an open architecture by utilizing ANSI-standard SQL. Conforming to this industry standard allows ad hoc query tools to readily adapt to changing information systems requirements.

While these report writers offer many advantages as basic reporting tools, they fall short of meeting many of the requirements organizations place on their decision support systems. Their greatest shortcoming results from the fact that end-users are still directly involved in the SQL definition process. In order to achieve this simplified ad hoc query environment, reports writers have had to sacrifice analytical complexity. To perform analyses on micro-segments of the data warehouse, a query plan would need to consist of several related SQL statements. This SQL execution plan is difficult for even experienced database professionals to write, and impossible to construct with single-statement relational report writers. As a result, relational report writers do not provide the complex analyses required by today's business users.

Data access times are not a significant issue with small data warehouses; however, as databases grow from a few gigabytes to tens and hundreds of gigabytes, the data warehouse schema needs to be optimized for VLDB in order to achieve reasonable response times. Optimization techniques such as aggregating fact tables, partitioning fact tables, and denormalizing relation tables all provide the benefit of significantly decreasing response times and the drawback of increasing the complexity of the SQL. Because point-and-click ad hoc query tools shift the burden of SQL creation to the users, there is no systematized way of generating the optimized SQL required to run against these large warehouses. Consequently, these tools inherently have difficulty achieving reasonable response times with large data warehouses.

Decision makers and executives who require access to information stored in the data warehouse possess widely varying levels of technical ability. While many vendors of report writers claim that their users are 'shielded' from the database-level complexities of the data warehouse, some technical understanding of the underlying database table structures is always required to effectively utilize any ad hoc query tool. For this reason, report writers are not a realistic data delivery solution for the less technically adept users in an organization.

Report Writer Scorecard

Sophisticated Analytical Functionality	○
Support for a Wide Range of Users	○
High End-User Productivity	◐
Support for Very Large Data Warehouses	◐
Support for a Large Number of Users	◐
Easy to Build and Maintain	◐
Support for an Open Architecture	●

None ○ Partial ◐ Full ●

Figure 1

Multidimensional OLAP

Multidimensional On-Line Analytical Processing (OLAP) is a class of systems that provides sophisticated analysis capabilities using proprietary multidimensional databases. The proprietary multidimensional databases rely on an array processing technology adapted from the engineering scientific community. Using multidimensional array structures, called data cubes or meta cubes, multidimensional databases provide the ability to organize data hierarchically in multiple dimensions for advanced OLAP analysis of predetermined micro-segments of the data warehouse. Using proven array processing techniques, multidimensional databases are capable of conducting sophisticated analyses on small data sets extracted from the data warehouse.

A benefit of the multidimensional OLAP approach is an intuitive view of the data for the end-user. Users view the data multidimensionally and think about it as they would in the context of daily business transactions. Another advantage of this approach is the advanced analytical complexity provided to the users. This increased usability combined with sophisticated analytical functionality provides two compelling benefits for using a proprietary multidimensional database as a decision support tool.

There are, however, a number of limitations to the multidimensional OLAP approach. Proprietary multidimensional databases lack the scalability and flexibility organizations now require of their decision support systems. Today a wide variety of executives, knowledge workers, and service personal need to perform OLAP analyses of any micro-segment of the market depending on a variety of business drivers. The multidimensional cube structure is inflexible and does not support the ad hoc creation of multidimensional views of the products, services, and customers with which organizations may be working. Most importantly, multidimensional databases cannot handle more than 20 or 30 gigabytes of data. This flaw in the underlying architecture removes it from consideration as a data delivery platform for the enterprise. For those organizations with small (5-10 gigabyte) warehouses or for departmental solutions, multidimensional OLAP is a viable technology. However, for enterprises looking to leverage the power of its information across its employee base, these systems are inappropriate.

Proprietary multidimensional databases are as their name implies, proprietary systems that do not conform to industry database standards. This lack of an open architecture is a drawback to the multidimensional approach, because organizations are forced to commit resources to support a non-standard system. As the information systems infrastructure, data warehouse, and DSS reporting requirements evolve, organizations frequently face difficulties integrating an existing proprietary multidimensional database with other information systems.

While multidimensional databases provide sophisticated analytical processing capabilities, they are not capable of scaling to meet the complete set of requirements organizations place on their decision support systems. Users must constrain their analyses to predefined data cube subsets of the data warehouse. Populating and defining these data cube structures is a time consuming, IS resource-intensive procedure. Finally, a proprietary architecture must be endorsed as an organizational standard for decision support. For these reasons, multidimensional databases do not provide a realistic decision support solution.

Multidimensional OLAP Scorecard

Sophisticated Analytical Functionality	●
Support for a Wide Range of Users	◐
High End-User Productivity	●
Support for Very Large Data Warehouses	○
Support for a Large Number of Users	◐
Easy to Build and Maintain	○
Support for an Open Architecture	○

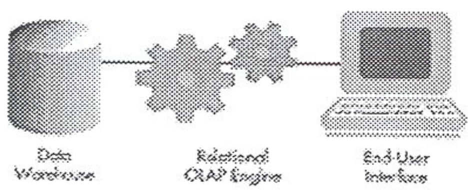
None ○ Partial ◐ Full ●

Figure 2

Relational OLAP

Relational On-Line Analytical Processing (OLAP) decision support systems were designed to remove the technical constraints imposed by the traditional data access methods, without removing the benefits provided by these methods. These systems combine the direct data access functionality of the report writers with the advanced analytical capabilities of the multidimensional OLAP systems to provide an open, scalable data delivery architecture.

Relational OLAP Architecture



Relational OLAP provides sophisticated analysis capabilities directly with a relational warehouse.

Figure 3

The relational OLAP system utilizes a dynamic OLAP Engine to generate the SQL queries used to access the relational warehouse. End-users are provided with an intuitive multidimensional analysis interface similar to those found in the proprietary multidimensional database systems. Through this interface, users can define complex multidimensional analyses which are dynamically compiled into SQL statements by the relational OLAP engine. A well architected relational OLAP solution should meet all the requirements of DSS. The following points will outline how relational OLAP meets these requirements and will provide an analytical framework with which to evaluate relational OLAP tools.

• Sophisticated Analytical Functionality - The dynamic DSS engine differentiates the relational OLAP systems from the basic report writers by removing end-users from the SQL generation process. It enables the advanced analytical reporting functionality found in many multidimensional OLAP systems without requiring a proprietary data store. It does this by generating multi-pass SQL execution plans database. (Some of the more sophisticated Relational OLAP tools provide "switches" which enable the engine to produce database-specific SQL. While this functionality enhances performance, organizations should be secure in their database platform choice before utilizing this functionality.)

Relational OLAP systems are a hybrid approach combining the best of ad hoc query tools and multidimensional OLAP systems. By considering the demands of the modern data warehouse during the design process, these tools meet the requirements of modern decision support and provide a flexible and robust platform for growth.

Relational OLAP Scorecard

Sophisticated Analytical Functionality	●
Support for a Wide Range of Users	●
High End-User Productivity	●
Support for Very Large Data Warehouses	●
Support for a Large Number of Users	●
Easy to Build and Maintain	●
Support for an Open Architecture	●

None ○ Partial ◐ Full ●

Figure 4

Summary

The quantities of information available to corporations from internal systems, customers, and external data services are increasing by orders of magnitude. Organizations who can effectively manage and make informed decisions based on these vast quantities of information will realize a tremendous competitive advantage in the marketplace. Traditional methods of decision support fall short of meeting the requirements organizations will place on their information delivery systems. Ad hoc query tools are limited to basic reporting and require technically adept end-users. Multidimensional databases are proprietary systems not capable of handling more than 20-30 gigabytes of information. Relational OLAP tools provide a solution by combining the direct data access functionality of the ad hoc query/report writers with the analytical sophistication of the multidimensional OLAP tools. Relational OLAP provides an open, scalable architecture that meets all the requirements for decision support.