What is OLAP – On-Line Analytical Processing?

Margus Tammeraja

Published in: "Ärielu" no 11, 1997

Every manager understands that he would be able to make more adequate decisions faster if he had not to waste his time on looking for the relevant data. All large and medium businesses use some business software - this means there is a computer somewhere where a large volume of data is stored. Still finding the necessary information is time consuming and irritatingly complicated. The managing director of a large company stated that his six analysts spend 80% of their time on looking for and collecting the data and a mere 20% on analysing. If we could reverse these percentages, we could save hundreds of thousands. Unfortunately this is not possible, as transaction processing in business software is operation based. Therefore in most companies we see the following picture:

- information is presented in the format of standard reports (printed on paper)
- data needs to be copied to a work sheet for analysis
- standard reports do not reflect the changing needs of the business
- finding the data that is not available in the standard reports is very complicated
- analysts spend too much of their time on gathering the data, which leaves them little time for thorough analysis
- different people get different results for the same query
- access to information for analysis and decision making is secondary as compared to everyday work
- it is complicated to use the necessary data to get a good picture of the business

Ideally analysts and decision makers would be able to work in a real-time situation and “play” with the data. This would let them be far more efficient. Let us discuss why it has always been so complicated to access the information and how modern technology brings the ideal closer to us.

Business Software does not Support Decision Making

The main reason why it is difficult to get the information from most business information systems is their function to support everyday business operations. Business software is meant for on-line transaction processing (OLTP). It is important to process orders, deliver goods or create invoices. Decision support is secondary for these systems. An OLTP and a decision support system differ radically. Therefore it would be complicated and also uneffective to create a system that would serve both aims. To intend to use an OLTP system for analysis is as good as to hammer screws into the wall. The following table presents the basic differences between an OLTP and a decision support system:

OLTP	**Decision support**
Save transactions	Presents information
Separate transactions (e.g an invoice)	Many-sided overviews

Only recently entered data is available	Transactions of several years are available on-line
Limited periods of time	Unlimited periods of time
Constant changes	Periodical changes
Only actual data	Actual data, budgets, prognoses and 'what-if'data

Analysts and Business Managers have their Specific Needs

To illustrate, let us compare the different approaches of a sales clerk and the marketing manager of a confectionary factory towards data processing. The clerk's task is to enter all sales orders, to check whether there is sufficient stock to satisfy these orders and to keep an eye on deliveries. For the clerk, to do his job efficiently, the database needs to contain real time detailed information about all standing orders and stock levels, in order to prevent a situation where goods are promised to one customer and sent to another one. Every now and then he needs to take a look at transactions entered a week or a month ago.

The marketing manager, on the other hand, needs to use the data only occasionally, but he then needs answers to questions like "What were the actual and forecasted wholesale figures of chocolate bars to St Petersburg during the last 12 quarters?". To answer this simple question the system needs to go through thousands or even hundreds of thousands of database transactions, which takes up more of the resources of the software system than the clerk uses during a whole month. The marketing manager does not need information about the current situation as he analyses the trends over longer periods. Lacking the transactions of a couple of hours or even days has no considerable effect on the overall trends. The manager normally needs a reliable answer in a couple of hours or by the next day. However, it is not possible to give a reliable answer if the data, the information is based on, keeps changing all the time.

Limitations of Business Software

OLTP systems (i.e business software) are mostly made for normalised databases. Normalising is a process where transactions are cut down into essential parts and saved into separate database tables. E.g different parts of an invoice (invoice header, customer contact data, item information, financial transaction, etc) can be located in 5 or 10 data tables. The idea of normalising is to save data just once (systems, where e.g invoices and their financial transactions are duplicated in different tables, sales and nominal ledger reports can give different results). Normalising makes the database easier to maintain and brings many advantages for an OLTP system. However, a normalised database is intuitively comprehensible for a business manager.

It is a common practise to use query tools and report generators on OLTP databases, but this is not an effective solution. Firstly, business managers do not have the time to acquire the necessary skills to use these rather complex tools; secondly, having acquired an understanding of it, the managers are still not able to move around in the structures of data tables to find answers to questions like the one we posed above. Consequently the "IT guys" have to create

special reports or send the data to spreadsheets.

Unfortunately such “customised” reports tend to be unsatisfactory, as it might take long to generate the report or it does not contain all the right information. Besides, generating a report with complex queries uses a lot of the system’s resources. To get an answer to the chocolate bar question (above) makes the system to look through very many database transactions and can take hours. What is even worse, is that the whole system slows down and this can seriously affect order processing and shipping of goods.

Another problem arises when you need an overview of the business to make a decision but you are yet not able to formulate the question. This can be called the process of discovering or "I will know the question as soon as I see the answer”. Traditional systems do not support discovering as they presume you know the exact question you want an answer for.

The Paradigm of Decision Support

Experts have reached the understanding that OLTP systems and tools are not sufficient for decision support. A new way of thinking has emerged, that decision makers and analysts need database tools that have specially been created to find and manipulate complex aggregated information. The classical relational database has proved to be adequate in many ways but it has never been meant to be a powerful tool for data synthesis, analysis and consolidation.

The new approach is called on-line analytical processing (**OLAP**). The main elements of OLAP' are:

Users **with no specific technical knowledge** can access data in their own computer whenever they feel the need;

Response times to queries are short (normally less than five seconds);

Data is achieved from a **central database**, users do not have to go looking for data;

A central database guarantees **safety** and **a uniform approach** to the data;

The decision support system is physically separate from the OLTP system, so that the analytical queries would not affect everyday data processing;

Users’ **access to data supports both analysis and discovery processes;**

The results are ready to be shared with other users.

We have observed the differences between OLTP and decision support systems. Let us now show why the multi-dimensional client/server architecture databases are the optimum solution for decision support systems.

Multi-Dimensional Databases (MDD)

An MDD allows users with no specific technical knowledge to make fast queries, as the data here is stored absolutely differently as compared to OLTP systems. Instead of storing data on the transaction level, an MDD presents tabular data necessary for decision support. Data is presented in the way managers usually observe the operating of their business, and not in

normalised data tables. The following diagram shows the logical structures of a relational database and an MDD:

Businessmen handle multi-dimensional information every day. The above example posed a four-dimensional question. The dimensions are products (chocolate bars), customers (wholesale buyers), areas (St. Petersburg) and time. Dimensions are just the different aspects of transactions, which business managers use to divide different transactions into groups with certain meaning. E.g as for sales, it would be wise to classify the transactions by product groups, customer categories, areas, salesmen and time. An MDD is a database that has been specially designed for the handling of classified data. Besides dimensions, business managers want to observe their data on different levels of detail. Therefore each MDD dimension can comprise different levels of detail (hierarchies).

The following table shows the connections between the dimensions and the hierarchies:

Dimensions	Levels of Hierarchy		
Product	Item Code	Product Group	Product Class
Area	District/ Town	County	Country
Customer	Customer Code	Customer Category	
Time	Day	Week	Month

The power of an MDD can be illustrated with the question of the marketing manager of the confectionery factory: "What were the actual and forecasted wholesale figures of chocolate bars to St Petersburg during the last 12 quarters?". It takes a couple of clicks to formulate this question for an MDD. In a multi-dimensional database this information is stored on just 24 records, which gives you an immediate response. We can imagine how a similar question would make an OLTP system look through thousands of records... Furthermore – if the answer called forth new questions, the MDD technology allows you to immediately look at more detailed data or change the angle. E.g the sales manager might want to look at the sales figures for chocolate in other areas or for a different customer category; or see which brand was the most popular, etc. If sales has decreased in some area, it is easy to find out, which product groups or sales channels have caused this. The sales manager can approach the data from many different angles in order to find all possible questions and reasons. The MDD technology allows the decision makers to follow their own train of thought and get a good picture of their business (process of discovering). This kind of approach is of crucial importance in the work of a manager, and it would be practically impossible if he would have to formulate each question very precisely and wait after several reports.