INFO-51 UDK: 371:659.3:007
Primljeno/Received: 2012-05-10 Authors Review/Pregledni rad

OLAP TOOLS IN EDUCATION

Danijela Subotić, Patrizia Poščić, Vanja Slavuj
Department of Informatics, University of Rijeka, Rijeka, Croatia

Abstract
Business intelligence systems are becoming a growing trend on the market nowadays. Reason for this is the increasing amount of information generated and collected within the company and from many external sources. In order to remain competitive on the market, it is very important to collect and select relevant information in time, from which a business decision can be made. OLAP (On Line Analytic Processing) systems represent one part of business intelligence systems. Their goal is analytic processing and data extraction for business decision-making. Given the increasing demands of the market, it is important to prepare students to work with these tools as early as in the classroom. In this paper we present an overview of the development and characteristics of OLAP systems and a method of their implementation for educational purposes. We compare the characteristics of two selected OLAP tools used in the classroom.

1. Introduction
Information is the key element of today's business. To get the information we need, first we have to collect data - facts, numbers or text - that can be processed by a computer. Relationships among data give us information, which we then convert into knowledge. That is why many business organizations collect and store vast amounts of data in different formats and in different databases. However, there is a problem of finding meaning in that vast amount of data and transforming it into useful information. The solution to this problem can be found in the implementation of business intelligence systems (henceforth BI) and Online Analytical Processing (henceforth OLAP) tools. BI systems enable business organizations to isolate large sets of data from various sources and convert them into meaningful information. OLAP tools are a part of business intelligence systems. The goal of OLAP tools is analytical processing and data extraction for the purpose of business decision making. With the help of OLAP tools, business organizations can accurately know the state within the company and all its parts, taking into consideration the conditions of the market. These tools can reveal hidden patterns in the collected data and, thus, help managers and analysts in making good business decisions.
The paper gives an overview of the concept and structure of an OLAP system, explains the OLAP schema and the difference between OLAP and OLTP systems, shows a common architecture and types of OLAP systems, presents reasons for introducing students to OLAP systems and compares two selected OLAP tools that are being used in teaching a graduate course at the Department of Informatics, University of Rijeka.

2. OLTP systems
Most business organizations use Online Transaction Processing (OLTP) systems in order to collect and store records generated by their everyday operations. These systems execute transactions: they insert, update or delete groups of records simoultaneously. For example, the information system of a food store chain inserts and updates information on prices, supplies and costs of merchandise, and does it very quickly. This is because they do not want customers waiting in line while the system updates the prices and stocks. Hence, OLTP systems have to collect, update, store, retrieve and archive information about the transactions /**1**/. These transactional systems are designed to manage raw data on daily business operations that require effectiveness and immediate transaction processing at the lowest level of detail.
However, the design of an OLTP system that allows exceptionally fast and accurate storing of transactions, creates problems in analyzing data in the database. This is because the OLTP database contains a large number of tables which often have multiple relationships with other tables in

the database. Because of this database complexity, the problem of understanding the database arises: we simply do not know where to look for the information we need.

The figure below shows a simplified example of an entity-relationship model /**2**/, which serves as the basis for building such a transactional database.

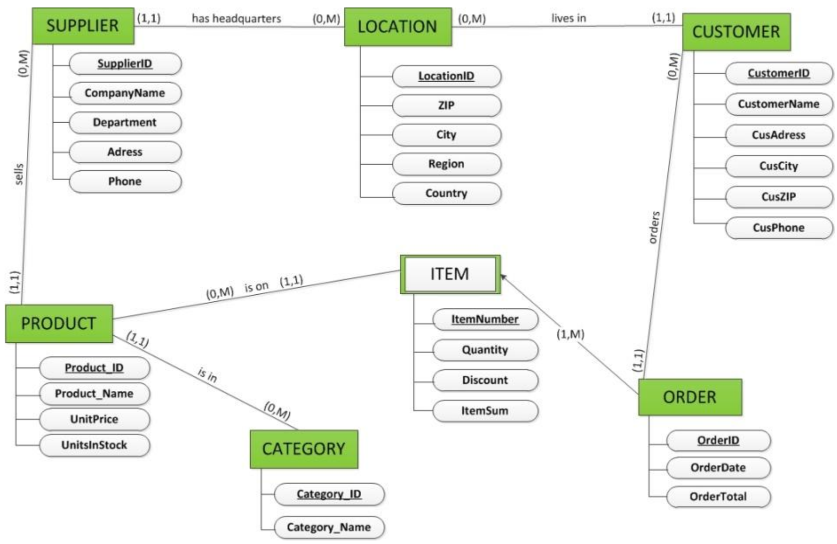

Figure 1. Entity-relationship model

During daily operation, OLTP systems are constantly updating and refreshing their data. It makes no sense to analyze data that is constantly changing because it is difficult to obtain the correct result. Also, OLTP database stores individual records. These records presents a problem for analysts because they use summarized data (total invoice amount, subtotals, quantity of items, etc.) in order to respond to business queries.
In other words, a system is needed that will extract data from the OLTP database, combine it to make totals and subtotals, and display it in a way in which the user can identify "smart" business decisions. OLAP systems are the solution to this problem.

3. OLAP systems

3.1. Defining OLAP

The term OLAP (Online Analytical Processing) originated from the British mathematician E. F. Codd, who laid the foundations of the relational calculus and presented the relational data model in 1970. The relational data model forms the basis for the most popular type of databases today: relational databases. The name OLAP itself points to the difference between such systems and the already well-known tools for online transaction processing (OLTP). In 1993, Codd defined a broad set of criteria to be met by a database in which or over which OLAP tools will be applied /**3**/.
OLAP is an information system for fast, consistent and interactive access and manipulation of multidimensional historical data that comes from different sources and is stored in a data warehouse. Data warehouse is a database structured in a way suitable for analysis. OLAP functionality is achieved through the possibility of multidimensional analyses of collected corporate data, which include: modeling by using data dimensions and hierarchies, analysis of trends over a period of time, projection of data using what-if scenarios, subsets of data, drilling to higher or lower levels of data detail, etc. OLAP systems are usually implemented in a client-server environment.

3.2. The link between OLTP and OLAP systems

OLAP and OLTP databases differ in several aspects. An OLAP database is usually kept separately from an OLTP database in order to provide good performance of the transactional database and ensure that the OLAP database receives only historical business data. Data in the transactional (OLTP) database is constantly changing, which is not the case with the analytical (OLAP) database. In the OLAP database, data always remains the same and, therefore, is called historical data. Once data has entered the OLAP database, it is not possible to make changes (update and delete) to that data. It is only possible to query that data, perform various mathematical operations on it and use it to make better business decisions.

Between transactional databases and databases intended for analytical processing there is a data warehouse. A data warehouse contains a copy of transactional data specifically structured for querying and reporting /**4**/. In other words, data warehouse is a database structured in a way suitable for performing various analyses. It is designed according to the principles of dimensional modelling. While modelling, we divide the existing OLTP system into business processes (eg. sales) and dimensions that describe these business processes (eg. time, customer, product, vendor, etc.). The selected dimensions determine the level of detail (hierarchy) in which we monitor a business process. Along with the selection of business processes, dimensions and level of detail, we also determine measurable facts that we want to know and follow (eg. income, sales quantity, costs, etc.).

3.3. OLAP schema

OLAP databases contain less tables than transactional databases and are based on a different scheme. Using dimensional model, OLAP databases reduce the number of relationships (connections between tables) to a minimum. The following figure shows what would the entity-relationship model from Fig. 1 look like if transformed into a dimensional data model.

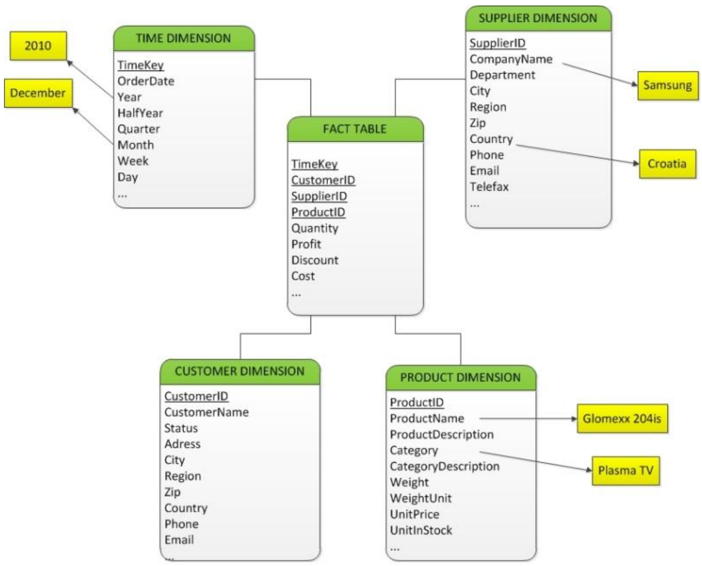

Figure 2. Dimensional model – star schema

Dimensional model scheme from Fig. 2 is called star schema. A star schema is a common schema of storing data in a data warehouse that got its name after its shape. The central table in a star schema is the fact table. It contains keys of all the other tables that are associated with it, as well as summarized measurable data, such as the *overall expences for all the milk in stock* or *the number of Samsung Glomexx 204is plasma TVs sold in December 2010 in Croatia*.

Tables associated with the fact table in the schema are called dimension tables. Each dimension table is defined by its primary key. Dimension tables contain additional data and descriptions that give meaning to the summarized data in the fact table. Given that the OLAP database is a "historical database", the answers to questions such as "*When were the most plasma TVs sold?* or "*What is the total quarterly/semiannual/annual profit in the cosmetics department?*" are certainly of interest to us. Accordingly, we can conclude that the key dimension in the OLAP database is that of time and each star schema, along with the fact table, must include the time dimension table. Elements (data) in the dimension table are called members and can be grouped into a hierarchical structure only if the members themselves allow it. For example, the dimension of time could have the following hierarchy:

Year → Half year → Quarter → Month → Week → Day → Hour

The hierarchical structure is based on a classic parent-child relationship (Half year is the child of Year, Quarter is the child of Half year, etc.). Every time we want to search the database, we must go over this hierarchical structure. Relationships between dimension tables and fact tables allow us to view data across any number of dimensions and through any number of hierarchical structures. For example, we may need the information about the "*overall profit made by selling Dorina chocolate in the year 2011 in Rijeka*" or the "*total expences and the number of alcoholic beverages sold in the first quarter of 2012 in Zagreb*". It is because of the simple star schema design (smaller number of tables and table relationships) that these queries are easier to write and are executed faster.

Dimensional model can also be presented in a snowflake schema. The star schema implies that each dimension in the model is fully denormalized, which is not the case with the snowflake schema. An example of the snowflake schema is given in Figure 3.

Figure 3. Dimensional model – snowflake schema

3.4. ETL processes

When a star schema is created, it is necessary to move all the data from the transactional database into the OLAP database (we have already emphasised that these two databases are separated). We must extract data from the transactional database, convert it into appropriate form and fill the tables in the OLAP database. This process is called ETL (Extract, Transform and Load) process /**5**/. ETL transfers data from transactional databases into a data warehouse. During that transfer, only data which we consider important and interesting enough to store in the data warehouse is selected. Next, this data has to be properly checked, modelled and reorganized in order to be accessed quickly and easily. When transfering data into the data warehouse, it is necessary to integrate and check the data coming from various sources. Time dimension of data access is especially important. That is why we have to model the data so that certain states of business processes, for specific time periods, could be adjusted (summarised, aggregated, and that other calculations could be made). Thus formed (transformed) data is entered into the data warehouse. The transfer of data from transactional systems into a data warehouse is done periodically (daily, weekly, monthly), usually after working hours so as not to burden the transactional system.

3.5. OLAP cube

Finally, we are left to create an OLAP "cube". Given the multidimensional OLAP data model, a table is represented as a cube. The OLAP cube aggregates the facts from each level of every dimension of an OLAP schema. We use the term "cube" because it best describes the resulting data. If we were going to create a cube based on our star schema (Fig. 2), we would take the expences, profit, quantity, discounts and other facts from the fact table, and combine them with the city, city and department, department and year and all the other possible dimensions and hierarchical levels. Suppose, for example, we want to observe sales according to suppliers, time and types of products; in that case, the data in the data warehouse would have to form a three-dimensional cube, and the sales would be seen through these three dimensions. Thus, we would get such data structure as shown in Figure 4.

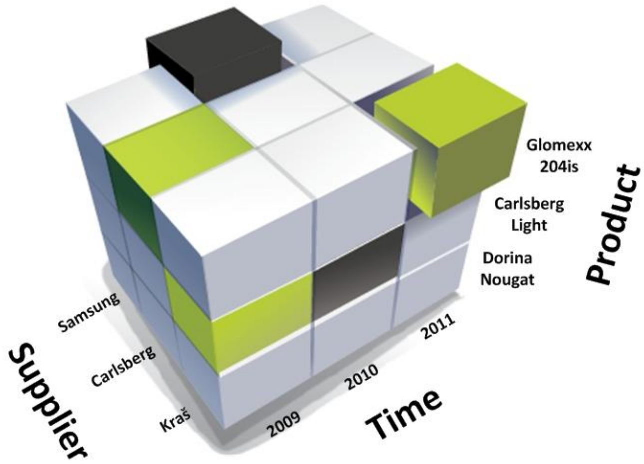

Figure 4. OLAP cube

Since the cube contains all our data in an aggregate form, all the answers (numeric data) are already in there, even before we make a query. This represents an advantage of OLAP cubes: no matter how complicated our question is, we will get the answer in record time (within seconds). Also, OLAP cubes can be viewed using any compatible OLAP tool.

3.6. Methods of OLAP analysis

There are two main methods or types of views of data during data analysis: Slice and Dice Method and Pivoting or Rotation Method. In the Slice and Dice Method, slicing represents the horizontal cross-section of the cube, while dicing represents the vertical cross-section of the cube. Also, an OLAP cube can be rotated in different ways around its axis. The rotation of a table so that the data is moved from rows to columns and vice versa is called pivoting /**6**/.

Drilling Method is also used and it comes in two varieties: drilling up and drilling down. Drilling is used to bring the analysis to a higher (eg. first we look into daily sales and then weekly, monthly and anual sales) or a lower level (eg. first we look into annual sales and then monthly, weekly and daily sales) in the hierarchy, within a particular dimension.

We can see that OLAP tools encompass a wide range of options, from simple searches, navigation and calculations, to more serious analyses such as time series analysis and complex modelling. In this way, these tools can control the whole hierarchical chain, which represents the most important part of making informed business decisions.

3.7. Architecture of OLAP systems

OLAP tools usually consist of OLAP server and OLAP client /**7**/ . The OLAP server is located between the user and the data warehouse and stores data in the form of multidimensional tables or cubes. The OLAP client is located on the user's computer and allows the user to query data from tables and cubes and to get graphic and tabular reports. An example of the architecture of OLAP systems is shown in Figure 5.

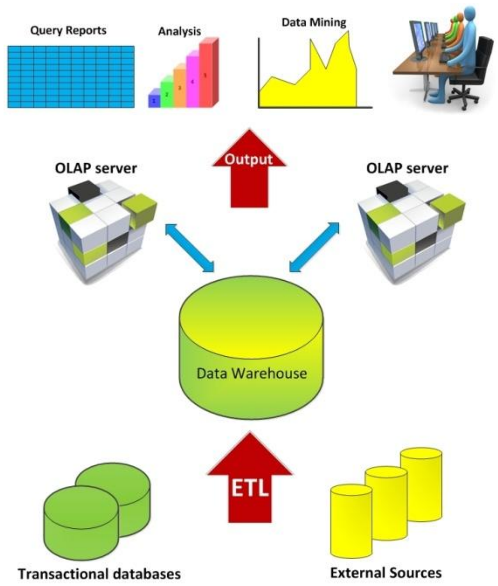

Figure 5. Architecture of OLAP system

3.8. Types of OLAP systems

There are several types of OLAP tools available on the market /**8**/:

- ROLAP tools,
- MOLAP tools,
- DOLAP tools,
- HOLAP tools.

ROLAP tools (Relational OLAP tools) use relational model as the basis for a database. Because of that, they can reflect the real world very accurately. A problem with the practical usage of these tools may arise if the manufacturers of these tools decide to implement some specifics to their product. This may lead to difficulties in applying standard query languages (eg. SQL) already familiar to users: other forms of interaction with the database are considered a shortcoming.

Multidimensional OLAP, or MOLAP, is a set of tools that use their own database with a multidimensional matrix structure. One of the characteristics of such a database is that it is not built on a common model or standard, such as the relational model. Because of that, MOLAP tools are less adaptable to changes in the size of a business system. Also, all the elements of the system are considered equal and are treated in the same way: it is not possible to identify any particular hierarchy among them. The advantage of MOLAP tools is their speed. Their shortcoming lies in their need for a large amount of memory space (because of the large number of dimensions).

DOLAP (Desktop OLAP) are OLAP tools for desktop computers. These are simple OLAP tools, with lower prices, and adapted for the use on desktop computers with modest memory capacity and performance options. DOLAP tools typically support only small applications designed to meet the needs of individual users. There are also versions of these tools for portable computers.

HOLAP tools (Hybrid OLAP) are a hybrid product with which it is possible to simultaneously carry out multidimensional analyses of data from the data warehouse and data from the management system relational database. These combines the advantages of MOLAP and ROLAP tools.

Some of the best known manufacturers of OLAP tools today are Cognos, Microsoft, Oracle, SAS, IBM, Microstrategy, Business Objects, Aplix, Hyperion Solutions, etc.

4. OLAP tools in education

Today, the implementation of business intelligence systems seems to be a general trend in business worldwide. We know that quality business decisions can be made only with the help of information. For this reason, we collect and store large amounts of data, often to the point we become overwhelmed with it. Business intelligence systems produce information based on the data obtained from internal and external sources. Information is becoming the basis for making quality business decisions which, up to recently, was done on the grounds of intuition. That is why we believe that it is important to begin preparing students to work with business intelligence systems (with OLAP tools in particular) as early as in the classroom. It is important that students become familiar with the concepts and methods of business intelligence and OLAP systems, so that when they graduate and try to find employment, they will be competitive on the labor market. Knowledge and ability to work in a business intelligence environment can be acquired by working with the selected OLAP tool.

Our course was designed as a combination of theoretical lectures and practical exercises using the computer. Through lectures, students were able to acquire some basic terms and concepts of business intelligence systems (ERP systems, data warehousing, dimensional modeling, ETL processes, OLAP and Reporting tools, etc.). During exercises, students were working in two selected OLAP tools. OLAP tools we had chosen to work in the classroom were the Palo Suite and Microstrategy Reporting Suite (both tools work on a web platform and are open source, which was the most important factor in the process of making a selection). Students, with the help of teachers, created test reports and went through the demo examples offered by each tool, thus adopting the principles of working with OLAP tools. After successfully solving demo examples, students were given the task to independently create several reports in both tools and compare the performance of both tools. Also, they had to write their own impressions of both tools.

4.1. Microstrategy Reporting Suite

MicroStrategy Reporting Suite is a Microstrategy's business intelligence software that is designed for various business organizations that want to start building and using reports quickly and easily, free of charge. It is a free reporting software that works on several operating systems (Windows, Unix, Linux, Solaris, HP-UX, and AIX) and with any data source. It has free online support and training and the license to use it cannot expire. It offers a comprehensive range of data visualizations, including: Gauges, Dials, Heat maps, Graphs, Tables, Scorecards, Speedometers, Traffic lights, Spark lines, Bullet graphs, Waterfalls, Graph matrices, Funnels, Bubble graphs, Bubble grids and Data clouds. Users have the ability to drill within the document, drill anywhere, or disable drilling. Drilling capabilities can be applied to a grid or a graph, which enables users to perform an investigative analysis and present the results directly on a dashboard, scorecard, or enterprise report. Elements can be removed or added dynamically and users can make their own calculations between various groups or individual items. Microstrategy Reporting Suite consists of several products and components which can be seen in Figure 6. Figure 7 shows a sample of a graphic dashboard and report.

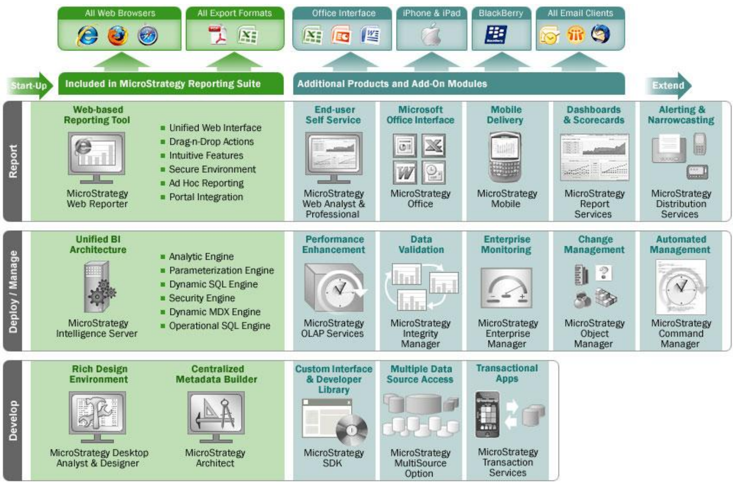

Figure 6. Microstrategy Reporting Suite components

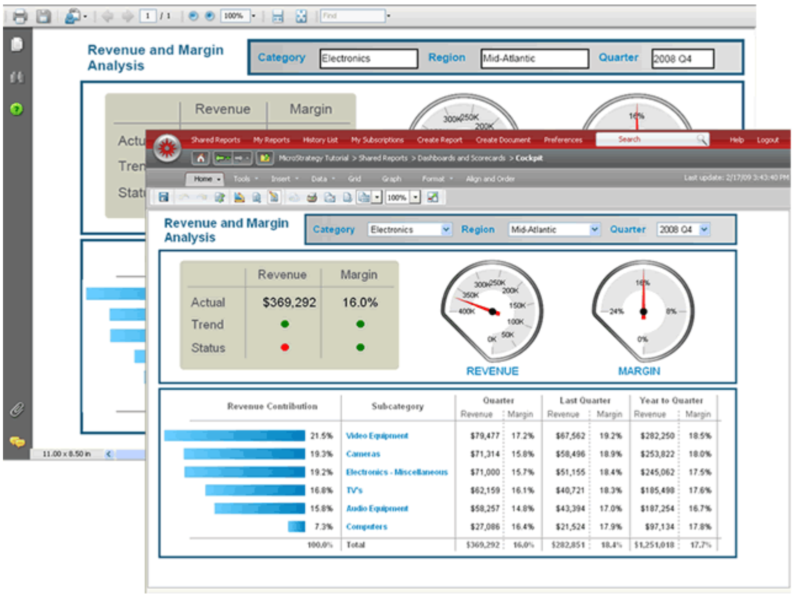

Figure 7. Microstrategy sample: graphic dashboard and report

4.2. Palo BI Suite - Palo Web

The Palo BI Suite combines four core BI applications (Palo OLAP Server, Palo Web, Palo for Excel and ETL Server) into one BI platform. Palo is an open source product which is completely free of any license fees. It was first developed in 2007 by the German-based company Jedox AG.

Palo Web is a component of the Palo BI Suite and includes all the tools and modules necessary to create comprehensive planning, analysis and reporting applications. Palo Web consists of the following: Palo Spreadsheet (offers a comfortable user interface and usability known within Microsoft Excel), Palo Pivot (allows ad-hoc requests on all Databases of the Palo OLAP Data model), Palo Pivot (allows filtering of all dimensions as well as their rows and columns), Palo User Manager (manages all users and user groups within the Palo Suite allowed to access single components and databases), Palo ETL Manager (manages the definition of source and target, transformation and loading of data into Palo OLAP databases) and Palo File Manager (sets user rights for every file within the Palo suite and offers every user and every application data space in which to save files). Figure 8 shows a sample Palo report.

Figure 8. Sample Palo report

ISSN 1333-6371

4.3. Criteria for evaluation of OLAP tools

E. F. Codd suggested twelwe rules for the evaluation of OLAP tools /**9**/. In our evaluation, we will use only some of them in order to simplify the process of evaluation and modify it to suit our needs. Also, we will add some of our own criteria appropriate for the classroom environment and theaching a course.
Codd's criteria that we will use in the evaluation are:

1. Multidimensional Conceptual View – multidimensional data displayed to the end user.

Accessibility – OLAP tool must access and retrieve only the required data, without the insertion of unnecessary data, regardless of data model type and database used. Further, it needs be able to make any necessary conversion of data in order to present to users unified, coherent and consistent reports.

Consistent Reporting Performance – performance should be the same, regardless of the number of dimensions used.

Client-Server Architecture – OLAP tool must be based on client-server architecture.

Multi-User Support – multiple users should be able to work simoultaneously in the OLAP tool.

Intuitive Data Manipulation (simplicity of use) – user view of data must contain everything necessary so that the user does not have to resort to using the menu or multiple searches of user interface.

Flexible Reporting – users should be allowed to present their data in a manner in which they wish to.

Unlimited Dimensions and Aggregation Levels – there should not be a limit on the number of dimensions and levels of aggregation in a model. Every serious OLAP tool should be able to accommodate at least 15 to 20 dimensions in the common analytical model.

Here is a list of our own criteria that we will use in evaluation:

1. Graphic representation of the reports - visual appeal of the report
2. Presence and quality of demo examples - checking whether they are (and how many) demo examples offered by the tool and how detailed they are
3. Reporting speed - how much time passes from the moment user sends a request for creating a report, until the final report is presented on the screen

4.4. Microstrategy Reporting Suite vs. Palo Suite

After a semester of working with both tools and completing the task of independently generating several reports, we asked our students to express their opinion of these tools. Taking into consideration their opinion, we came to a conclusion that both tools meet the set criteria, as is presented in Table 1. Report generating speed is great for both tools: reports are generated in a very short time. Further, both tools are free of charge, which is extremely important to us. Both tools are fairly simple to use, although a larger number of students (around two-thirds) said that Palo Suite is more intuitive and easier to use. They consider Palo Suite to be simpler than Microstrategy Reporting Suite: we ascribe that to the differences in structure and to the smaller number of components. As the advantage of Microstrategy Reporting Suite students pointed out a large number of demo tables and examples through which users can learn, try out options and generate reports. Running on various operating systems and working with all data sources is said to be yet another advantage of Microstrategy Reporting Suite. Finally, students consider it to have better support and a larger number of free online tutorials and webinars.
However, as a tool that better suits teaching purposes, they chose the Palo Suite. They consider it smaller and simpler, more intuitive and more detailed, and, overall, more appropriate for learning and classroom usage.

Table 1. Comparison of tools by set criteria

Criteria	Microstrategy	Palo
Multidimensional Conceptual View	✓	✓
Accessibility	✓	✓
Consistent Reporting Performance	✓	✓
Client-Server Architecture	✓	✓
Multi-User Support	✓	✓
Intuitive Data Manipulation (Simplicity of use)	✓	✓
Flexible Reporting	✓	✓
Unlimited Dimensions and Aggregation Levels	✓	✓
Graphic representation of the reports	✓	✓
Presence and quality of demo examples	✓	✓
Reporting speed	✓	✓

5. Conclusion

Today's market is characterized by great dynamism and competitiveness. The amount of business data is growing on daily basis. Business organizations are becoming more and more reliant on collecting information in order to respond to business requirements and to predict the situation on the market. They either survive or fail, depending on the quality, flexibility and speed of their information systems. In order to successfully make appropriate business decision, it becomes necessary for them to implement a business intelligence system and OLAP tools that generate "smart, tailor-made" business reports. The quality of business decisions made with the help of business intelligence systems is considerably higher than of those made on the basis of intuition. Because of this growing trend of relying on business intelligence systems and because of the constant demand for educated and competent workforce, we began to familiarize students with the concepts of business intelligence and OLAP systems. Students were required to work with two OLAP tools, followed by evaluation of both tools. During the course, students had the opportunity to familiarize themselves with the tools' interfaces, understand the concepts of OLAP reporting, generate different types of reports and compare the performance and functions of both tools they used. They also got an insight into the structure and working mode of selected OLAP tools. This knowledge gives them great flexibility and makes them more competitive on the market.

Notes

/1/ Claybrook, Billy G., OLTP: Online Transaction Processing Systems. John Wiley and Sons, Inc., New York, 1992.
/2/ Modelling according to: Pavlić, M., Informacijski sustavi. Odjel za Informatiku, Sveučilište u Rijeci, Rijeka, 2010
/3/ Codd E.F., Codd S.B., Salley C.T., Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. Codd & Date, Inc., 1993., Volume 32, Pages 31.
/4/ Kimball Ralph, Ross Margy, Thornthwaite Warren, Mundy Joy, Becker Bob, The Data Warehouse Lifecycle Toolkit. John Wiley and Sons, Inc., Indianapolis, 2008.
/5/ Kimball Ralph, Caserta Joe, The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. John Wiley and Sons, Inc., Indianapolis, 2004.
/6/ Celco Joe, Joe Celko's Analytics and OLAP in SQL, Morgan Kaufmann, San Francisco, 2006.
/7/ Inmon William H., Building the Data Warehouse. John Wiley and Sons, Inc., Indianapolis, 2005.
/8/ Ibidem
/9/ Codd E.F., Codd S.B., Salley C.T., Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. Codd & Date, Inc., 1993., Volume 32, Pages 31.

OLAP ALATI U NASTAVI

Danijela Subotić, Patrizia Poščić, Vanja Slavuj
Odjel za informatiku, Sveučilište u Rijeci, Rijeka, Hrvatska

Sažetak
U današnje doba sve veći trend u poslovanju postaje implementacija sustava poslovne inteligencije. Razlog tomu je rastuća količina informacija koja se stvara i prikuplja unutar samog poduzeća te iz brojnih vanjskih izvora. Kako bi poslovanje ostalo konkurentno, važno je na vrijeme prikupiti i odabrati odgovarajuće informacije na temelju kojih se zatim donose poslovne odluke. OLAP (On Line Analytic Processing) sustavi predstavljaju jedan dio sustava poslovne inteligencije. Cilj im je analitička obrada, procesiranje i izvlačenje podataka za potrebe poslovnog odlučivanja. S obzirom na sve veću potražnju, već u nastavi je važno početi pripremati studente za rad s navedenim alatima. U radu ćemo prikazati pregled razvoja i karakteristika OLAP sustava te način njihove implementacije u nastavni proces. Usporedit ćemo karakteristike dva odabrana OLAP alata korištena u nastavi.