OLAP Technology in the Enterprise Information Structure

A. Gumarova, G. Kamalova

***Abstract*—** Enterprise information systems contain applications designed for complex analysis of data, their dynamics, trends, etc. To make any management decision, you must have the necessary information. To do this, you need to collect this data from all the information systems of the enterprise, bring it to a common format, and then analyze it.

The purpose of the work is to develop an information system using a multidimensional database to improve the efficiency of data processing and analysis by sharing existing information resources of the institution with new information systems

This paper describes a solution to the problem of analyzing large amounts of data using OLAP technology. The purpose of data warehouses is to provide users with information for statistical analysis and management decision — making. Data warehouses provide high-speed data acquisition, the ability to obtain and compare so-called data slices, as well as consistency, completeness and reliability of data. OLAP (On-Line Analytical Processing) is a key Component of building and applying data warehouses. This technology is based on the construction of multidimensional data sets — OLAP cubes, whose axes contain parameters, and cells-aggregate data that depends on them. Applications with OLAP functionality provide the user with analysis results in an acceptable time, performs logical and statistical analysis, supports multi-user access to data, performs a multidimensional conceptual representation of data, and has the ability to access any necessary information.

It also discusses the basic principles of logical organization of OLAP cubes, their main terms and concepts used in multidimensional analysis. As a result, we determine what different types of hierarchies are in the dimensions of OLAP cubes.

***Keywords*—** OLAP cube, centralized storage, database server, ETL, OLAP technology, OLTP, multidimensional.

I. Introduction

Using the OLAP system allows you to automate the strategic level of organization management. OLAP (Online Analytical Processing – real-time analytical data processing) is a powerful technology for data processing and research. Systems built on the basis of OLAP technology provide almost unlimited opportunities for compiling reports, performing complex analytical calculations, building forecasts and scenarios, and developing a variety of plan options.

Full-fledged OLAP systems appeared in the early 90's, as a result of the development of information systems for decision support. They are designed to convert various, often disparate, data into useful information. OLAP systems can organize data according to a certain set of criteria. However, it is not necessary that the criteria have clear characteristics.

OLAP systems have found their application in many issues of strategic management of the organization: business performance management, strategic planning, budgeting, development forecasting, financial reporting, performance analysis, simulation of the external and internal environment of the organization, data storage and reporting.

F. A. Gumarova, Higher School of Information Technologies of the Institute of Economics, Information Technologies and Professional Education of the Zhangir Khan West Kazakhstan Agrarian Technical University; Uralsk, Republic of Kazakhstan (e-mail: g_ainura_91@mail.ru).

S. G. Kamalova, Higher School of Information Technologies of the Institute of Economics, Information Technologies and Professional Education of the Zhangir Khan West Kazakhstan Agrarian Technical University; Uralsk, Republic of Kazakhstan (e-mail: gokhakam@gmail.com).

II. THE STRUCTURE OF OLAP SYSTEM

The OLAP system is based on processing multidimensional data arrays. Multidimensional arrays are designed so that each element of the array has many relationships with other elements. To form a multidimensional array, the OLAP system must get source data from other systems (for example, ERP or CRM systems), or through external input. The user of the OLAP system receives the necessary data in a structured form in accordance with their request. Based on the specified procedure, you can imagine the structure of the OLAP system.

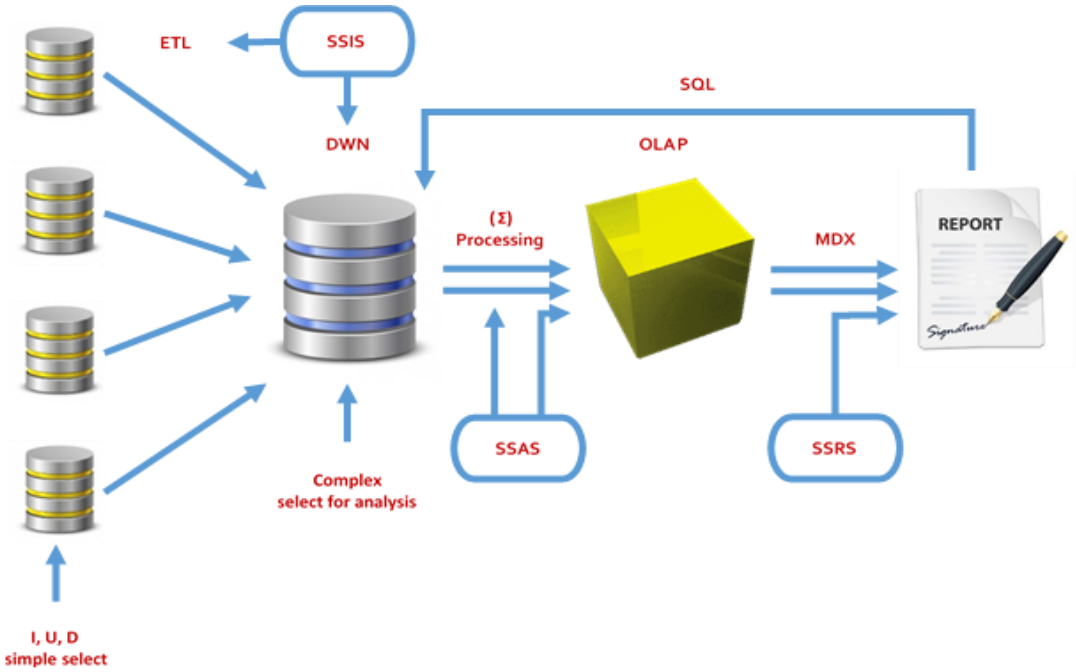

Fig. 1 The structure of olap system

In general, the structure of an OLAP system consists of the following elements.

- Database. The database is a source of information for the OLAP system. The type of database depends on the type of OLAP system and the algorithms of the OLAP server. Generally used a relational database, multidimensional database, data warehouse, etc.
- OLAP server. It provides management of a multidimensional data structure and the relationship between the database and users of the OLAP system.
- Custom application. This element of the OLAP system structure manages user requests and generates the results of accessing the database (reports, graphs, tables, etc.)
  - Depending on the way data is organized, processed, and stored, OLAP systems can be implemented on users ' local computers or using dedicated servers.
  - There are three main ways to store and process data:
- Locally. Data is placed on users ' computers. Data is processed, analyzed, and managed on-premises. This structure of the OLAP system has significant disadvantages related to data processing speed, data security, and limited use of multidimensional analysis.
- Relational database. These databases are used when the OLAP system works together with the CRM system or ERP system. Data is stored on the server of these systems in the form of relational databases or data warehouses. The OLAP server accesses these databases to form the necessary multidimensional structures and perform analysis.
- Multidimensional databases. In this case, the data is organized as a special data storage on a dedicated server. All data operations are performed on this server, which converts the source data into multidimensional structures. Such structures are called an OLAP

cube. Data sources for forming an OLAP cube are relational databases and / or client files. The data server performs preliminary data preparation and processing. The OLAP server works with an OLAP cube without direct access to data sources (relational databases, client files, etc.).

Types of OLAP

There are different types of OLAP models that vary in the way the data is stored, the way access to the data is organized, etc. We can distinguish three main online analytical processing models:

ROLAP

Relational OLAP, or ROLAP, stores all data, including aggregations, in relational databases rather than cubes, and doesn't use pre-computation. ROLAP works with SQL tools that send queries directly to the relational database. However, SQL tools do it in a way that creates an appearance of OLAP operations such as slicing, dicing, and drill down.

This approach is the best option for enterprises with large data warehouses. It heppens for three main reasons:

- offers better scalability;
- short load times;
- it is good at processing text information. However, the relational database needs specifical design (an OLTP database won't do), SQL doesn't support a lot of complex calculations, and it is often requires to create aggregate tables to improve query execution. Overall, ROLAP tools usually have lower performance than those on MOLAP base.

MOLAP

Multidimensional online analytical processing (MOLAP) is the traditional OLAP model. All data, including aggregations, is stored in multidimensional data cubes. The data in the cubes is pre-computed, which ensures fast query performance.

The multidimensional model makes it easier to locate cells in a cube than in a table, which, along with powerful techniques and calculation algorithms, allows retrieving and analyzing data significantly faster than when working with relational databases.

At the same time, since the data in the cubes is pre-calculated and thus harder to update, such systems are not very scalable. Besides, transferring data from a relational database to a MOLAP engine can lead to data redundancy.

HOLAP

Hybrid OLAP (HOLAP) combines features of the two previous approaches in order to provide fast query processing in combination with high scalability. In this OLAP data model, a relational database and multidimensional cubes divide the data. There are two approaches as to how to divide the data:

- Horizontal partitioning implies that some data slices that require faster query processing are stored in MOLAP. And ROLAP stores the others;

- In the vertical partitioning, aggregations are stored in MOLAP and detailed data – in ROLAP.

Major vendors now support this approach in one form or another.

Other types

There are four more OLAP models we want to touch upon:

DOLAP

Desktop OLAP, or DOLAP, implies that users can download a part of the cube on their computer and work with the data locally. The main advantage of such applications is that they are easy and cheap to deploy and maintain. The biggest drawback, however, is their very limited functionality.

WOLAP

WOLAP includes web-based online analytical processing tools that do not provide options for data download or local access to data. Such solutions have low deployment cost as all the client needs is Internet connection and a web browser. However, just as desktop OLAP, they have limited functionality. Besides, all major providers now offer web-only access to data as a feature of their much broader solutions.

RTOLAP

The main feature of Real-Time OLAP (ROLAP) is that aggregations are not stored in relational tables or cubes, but are calculated in-memory during the query execution.

DMR

This term is frequently meets in the context DMR vs OLAP. However, Dimensional modeling of relational data source (the full for DMR) is a form of ROLAP. DMR includes the representation of relational data sources using OLAP cubes, therefore DMR is closely related to the latter. DRM also uses MDX-based queries that promotes powerful analysis. It helps organizations to get a reliable and effective solution for tracking various aspects of a particular business, including profitability portfolios, cost analysis and anomalies analysis and business fraud.

OLAP vs. OLTP

Online analytical processing was created due to the inability of OLTP systems to process complex queries fast enough (OLTP stands for online transactional processing). Since then, the two technologies have been contrasted to each other constantly. Here is a short overview of the main differences between OLAP and OLTP tools.

Difference:	OLAP	OLTP
Queries	Works with a relatively small amount of very complicated read-only queries	OLTP processes a large volume of simple queries of different kinds, such as insert, update, read, and delete queries.
Users	Tools of OLAP are mainly used by data analysts, managers and decision-makers.	These systems are used by database administrators and IT professionals.
Data	OLAP uses consolidated historical data organized according to a star schema or a snowflake schema. OLAP data updates rarely and the updating process takes a lot of time.	OLPT uses operational data stored in many tables. System's data updates regularly and fast.
Data representation	Multidimensional.	Flat two-dimensional.

Size of the database	As OLAP databases have to store consolidated data, they are significantly larger than OLTP ones, containing millions of record and occupying up to 1 TB as opposed to thousands of records and the size of up to 1 GB of the OLTP databases.	
Purpose	The main purpose of OLAP is data analysis and business intelligence operations scaffolding planning, problem-solving and decision-making processes in the company.	OLTP is used for running and controlling ongoing business tasks.

OLAP Operations

The data in cubes is organized in dimensions that consist of hierarchies of multiple levels. Such data structure lays down the foundation for operations of online analytical processing tools that facilitate data analysis. The main OLAP operations include: roll up, drill down, slice, dice, and pivot.

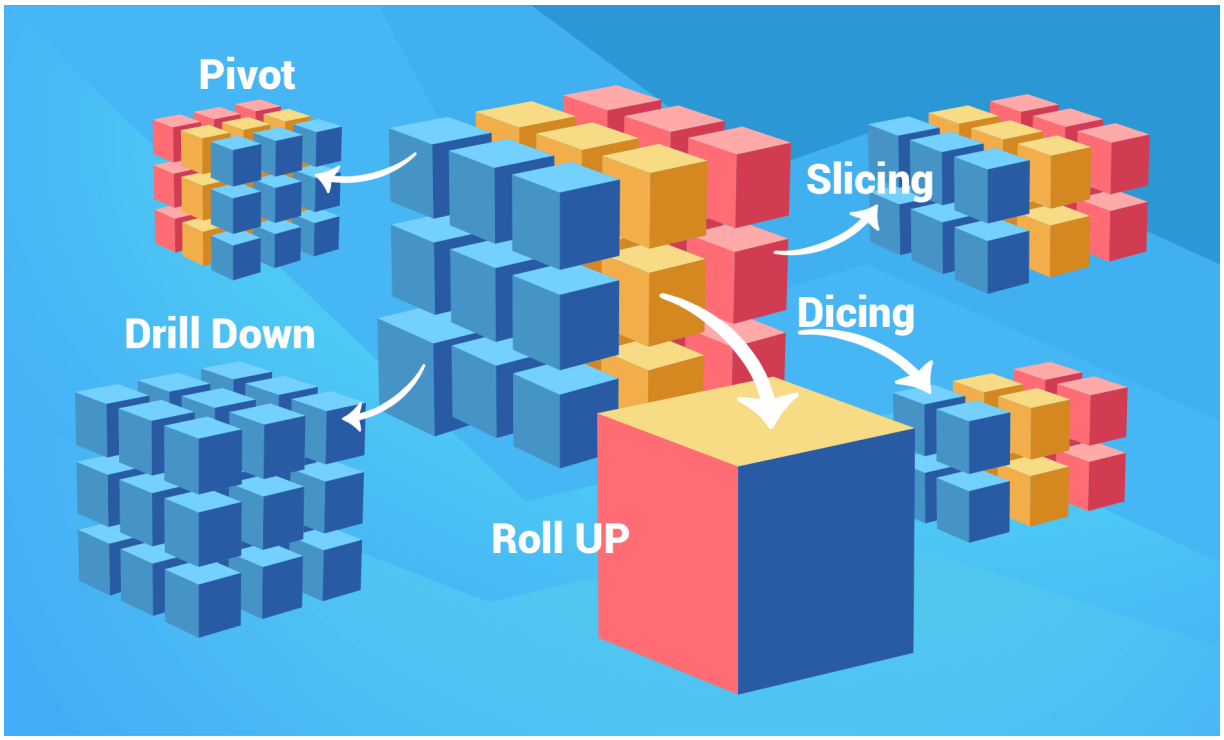

Figure 2. OLAP Operations

Roll Up

Roll Up is the aggregation of data along a hierarchy of a cube dimension. When Roll Up is performed, the data is aggregated by ascending a hierarchy.

Drill Down

Drill Down operation allows users to step down the hierarchy, from the most aggregated values to the most detailed ones. For example, “year” to “month” to “day” in the Date dimension, “country” to “province” to “city” in the Geography dimension.

Slice

The Slice operation extracts a subcube from a cube by selecting a value for its dimension. For example, we can retrieve a “slice” from a cube showing the number of internet sales in all countries for all years by choosing “Austria” as the value of the country dimension.

Dice

The Dice operation is an “extension” of the slice operation as it allows users to extract a subcube by selecting values for several dimensions. Using the example above, we can perform the dicing operation by selecting “Austria” for the country dimension and “2017” for the Date dimension.

Pivot

Pivot operation figuratively rotates the cube in order to provide a view of the OLAP data from an alternative perspective. This operation allows analyzing the data from different perspectives, making it one of the most important OLAP features for creating analytical reports.

Application of OLAP

As we have already mentioned above, the main asset of online analytical processing is its powerful data analysis capabilities. It determines extensive use of OLAP in business intelligence and reporting applications.

Standard examples of OLAP applications include marketing analysis and planning, sales analytics, business modeling and forecasting, financial reporting and other spheres. Online analytical processing tools are widely used in data mining, providing insights into the company’s data trends and delivering information to the management used in the company's’ problem solving and decision-making processes.

Software product

Within the framework of OLAP technology, various groups of specialists have developed a large number of software products that implement a multidimensional data model. These products include OLAP Option to Oracle Database by Oracle, Microsoft Analysis Services by Microsoft, Palo by Jedox, and Mondrian by Pentaho. The usability of the end system, its scalability, performance, and functionality depend on the tools designed for computer-aided design of such systems.

III. Conclusion

OLAP technology is a relevant and popular research topic, and its practical results are widely used.

Despite a fairly long history of research, there are still no uniform terminology standards, data transmission standards, query language, and cube formation. Growing volumes of corporate data increase the importance of analysis tools, most of which are based on OLAP principles, and therefore the problems of choosing the optimal storage and processing schemes for OLAP cubes are relevant. Integration of various sources of information for analysis gives rise to new applications of analytical technologies, for example, for the analysis of geo-positioning data (GPS) in relation to financial information, spatioOLAP technologies are used. Budgeting tasks that require combining the input speed of transactional systems and the analytical

Capabilities of OLAP are a special class of systems whose algorithmic base is just being created.

ACKNOWLEDGMENT

This publication is the result of the project implementation ERASMUS+ ACeSYRI: Advanced Centre for PhD Students and Young Researchers in Informatics reg.no. 610166-EPP-1-2019-1-SK-EPPKA2-CBHE-JP

REFERENCES

[1] Barseghyan A. Data analysis methods and models: OLAP и Data Mining (CD-ROM) / A. Barseghyan, M. Kupriyanov, V. Stepanenko, and others.. -, 2017. - 336 p.
[2] Berger A.Microsoft SQL Server 2005 Analysis Services. OLAP and multidimensional data analysis / A. Berger. –Saint Peterburg, 2018. - 147 c.
[3] Ovsjanitskaja L. Data mining as a component of pedagogical management // Education and science. Proceedings of the Ural branch of the Russian Academy of education. 2013. № 10. P. 80–90.
[4] P'janina S. About one approach of monitoring the business processes of the University // Eastern European journal of advanced technologies. 2011. № 2. T. 4. P. 6–8.
[5] www.microsoft.com/en-us/sqlserver/solutions-technologies/business-intelligence/analysis.aspx
[6] https://olap.com/learn-bi-olap/tutorials/
[7] https://galaktika-soft.com/blog/overview-of-olap-technology.html
[8] https://www2.cs.sfu.ca/CourseCentral/459/han/papers/chaudhuri97.pdf
[9] Erik Thomsen Microsoft® OLAP Solutions; - Moscow, 2014. – p. 528 .
[10] Weinberg G.M. Psychology of computer programming; - , 2014
[11] Ye N. The Handbook of Data Mining; - , 2014.
[12] Guo Y., Grossman R. (eds.) High Performance Data Mining: Scaling Algorithms, Applications and Systems; - , 2013.
[13] Resnick M. Turtles, termites and traffic jams. Explorations in massively parallel microworlds; - 2014.
[14] Jesse Russel Palo (OLAP database); Book on Demand-Moscow, 2012. – p. 82.
[15] SAS Institute SAS(R) 9.1.3 OLAP Server; Book on Demand-Moscow, 2010. – p. 88.