"Dimitrie Cantemir" Christian University
Knowledge Horizons - Economics
Volume 10, No. 3, pp. 68 - 76
P-ISSN: 2069-0932, E-ISSN: 2066-1061
© 2018 Pro Universitaria
www.orizonturi.ucdc.ro

BUSINESS INTELLIGENCE AND OLAP

Liviu Constantin STOICA

PhD, Academy of Economic Studies, Bucharest, email: stoica.liviu.constantin@gmail.com

Abstract *In this article, I studied Business Intelligence and Olap, that is, I have emphasized the basics of OLAP, how to deploy OLAP, metadata, the multidimensional model, and the cube. In terms of OLAP deployment, we have studied the most important server implementations that are available (MOLAP, ROLAP, HOLAP), while also emphasizing DOLAP. In addition to OLAP, we have also studied the most commonly used ETL metadata and tools. In the last part we studied the multidimensional model and cube model by realizing a vision of the DB2 Cube and MQT.*

Key words:
Database, concept, cube, tool, model
JEL Codes:
M1

1. INTRODUCTION TO OLAP

The goal of the information systems is the possibility of collecting, organizing and effectively operating the data masses available for a company. Over the years, technologies have evolved from simple reporting systems to Integrated Business Systems Intelligence while companies have made efforts to make effective use of business information. During the course of this development, a sophisticated development of Data Source Source, Data Transform, and Data Loading into Target Systems - known as ETL tools has been observed.

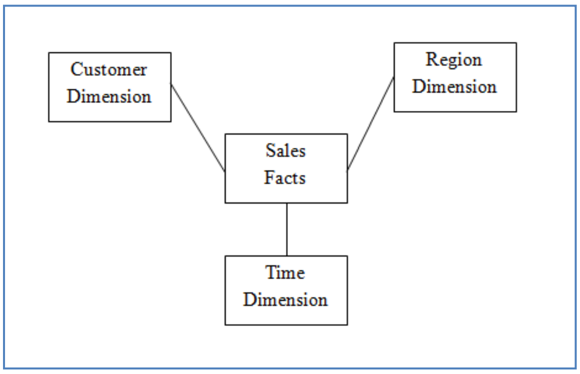

Figure 1 - Representation of the multidimensional model

The most important tools are used to query the data and have been developed for:

- be able to work with different data structures,

- keep up with the development of Web-based technologies,
- to meet the increasing demand for information analysis.

The database technologies have undergone a similar change, full of improvements that seek to meet customer inquiries in terms of information analysis.

Management systems that incorporate relational databases often have to solve large-volume data conflicts and provide users with quick answers to their queries.

The higher the tables, the more complex the data model, the more complex the attempt to provide acceptable responses over time. Additionally, exposing a data model complex to a company analyst brings other problems. An information analyst knows the questions he needs to put in the database. However, if the dataset structure is such that they do not understand how to formulate the query in a way understood by the database, they will have difficulty defining the queries, and productivity will be low.

The growth, enhancement of multidimensional data models has been seen as an attempt by data modelers to structure data in a way that is more easily understood by the information analyst. A multidimensional data model is typically targeted to a particular part of the firm, such as the sales model or the financial model.

The central point of this multidimensional model is the facts table. The facts table is the company's unit of measurement, such as quantities, monetary units, and reports are applicable in certain departments of the firm. The facts table unites, links a number of dimensional tables. These dimensional tables reflect the different ways in which user requirements can be analyzed. For example, for analyzing the facts table, we are interested in customer sales by region and by month. In figure1 a simple example of such a model, "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, Ralph Kimball, John Wiley, 1996".

2. ONLINE ANALYTICAL PROCESSING (ONLINE ANALYTICAL PROCESSING)

The Online Analytical Processing Concept is conceptual to be able to analyze business-related facts through multiple business dimensions. Online Analytical Processing is the concept that is being exploited with Online Analytical Processing (OLAP) technology. Using OLAP technologies, the company's measurement units can be analyzed by size.

Each dimension is typically expressed as a hierarchy. For example, the Time dimension could be expressed as an Hierarchy Year, Quarter, Month, Week. Queries are then an expression of units of measure (or facts) for an annuity part of a multidimensional database. The term part is used to describe the range of facts that all possible queries can access at a certain level per dimension for all the set of dimensions. An example is represented in figure 1.2 where four dimensions can be observed: Time, Store, Client, and Product.

Knowledge Horizons - Economics
Volume 10, No. 3, pp. 68 – 76, © 2018 Pro Universitaria

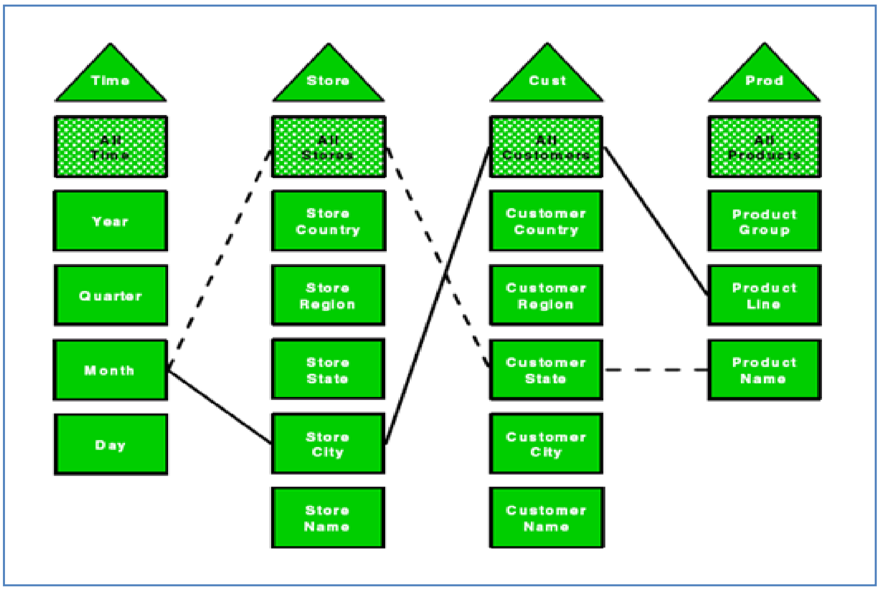

Figure 1.2 - Database slices

The table of facts itself is not represented in this chart. The solid line and dashed line represent two different data parts. Continuous queries are part of the month's database, store cities, all customers. For example, the query may be for sales in May of a specific product line for all customers consolidated by the store's city. The data line represented by a broken line represents the months, all stores, the sales status, and the level of a particular product.

3. OLAP IMPLEMENTATIONS

The term OLAP [1] is generally the term includes a number of different technologies that have been discovered to deploy an OLAP database. The most common server implementations that were currently available are: [2]

- MOLAP,
- ROLAP,
- HOLAP.

MOLAP is the preservation for Multidimensional OLAP. The database is stored in a special, proprietary structure that is optimized for multidimensional analysis.

ROLAP is Relational OLAP. The database is a standard relational database and the database model is a multidimensional model, often referred to as a star or snowflake model.

HOLAP stands for Hybrid OLAP and, as the name stands for, is a hybrid between ROLAP and MOLAP. In a MOLAP database, data is often pre-calculated, which is an advantage that a query is answered in a very short time, but the drawbacks include the time taken to calculate these pre-calculated values and the space needed to store that one. There is therefore a practical limit to the size of a MOLAP database.

In a ROLAP database, query performance will mostly be governed by SQL complexity and the number and size of the tables that are related to the query. In the context of these constraints, a ROLAP solution is generally an escalating solution.

A HOLAP database can be thought of as a virtual database in which the higher levels of the database are implemented by MOLAP, and the lower levels are ROLAP.

For example, in Figure 1.2 one or more of the product name, client name, store name, or day could be stored in a relationship while the rest of the database could be MOLAP. It can not be said that only a very low level of any size could be stored in a relational database. For example, for client size, both variables could be stored both in the customer's name and in the client's city in a relational basis.

From a user perspective, the line between what is stored in a MOLAP and what is stored in a relational basis should be unnoticed. Therefore, a HOLAP environment tries to combine the benefits of both MOLAP and ROLAP technologies. Storing them the lowest levels of a dimension in a relational basis instead of a MOLAP. Therefore, the size of the MOLAP database is reduced and therefore the time needed to perform an ante-calculation of the aggregate data is low. Queries that require data from a MOLAP section of a database will benefit from the fast performance expected with ante-calculated data. Moreover, by storing the lower levels in a relational, the database as a whole (MOLAP and relational combination) can be extended to take over the benefits of a relational basis.

The above terms are used to refer to servers based on OLAP technologies. Also, there is an acronym, DOLAP, which refers to Desktop OLAP. DOLAP gives users the ability to quickly drop small cubes running on desktop or laptop computers. All of these OLAP technologies can be implemented using the IBM DB2 family of products.

MOLAP is provided with the IBM DB2 OLAP Server that is an OEM (version of Hyperion Essbase) version and is separate from the DB2 database engine. HOLAP is available by using both IBM DB2 OLAP Server and IBM DB2 itself. ROLAP and DOLAP are available with DB2, and various dimensional dimensioning tools are provided with them.

4. OLAP AND METADATA

While DB2 has accommodated successfully with each of the OLAP technologies, so far it is not accommodated with Cube views, taking into account the multidimensional structure of OLAP databases and the entity entities in which this structure is found.

OLAP environment information is stored as metadata. Matadata [3] is data about data. The nature of metadata is dependent on the context in which they are used. For example, in the ETL environment, metadata can contain information about the layout of source data and target data, transformation rules, and data about how data is loaded (number of rows inserted, uninitialized number, and related time).

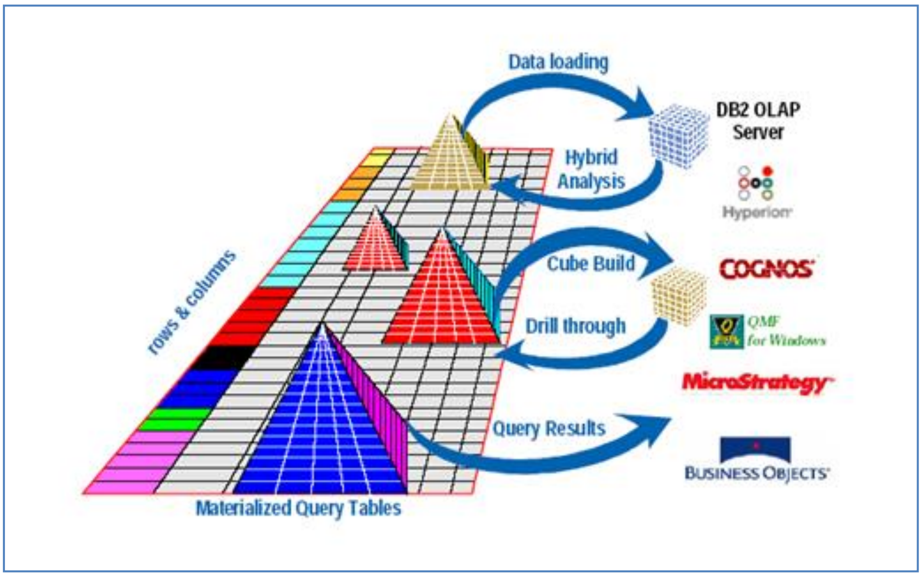

Figure 2 - DB2 Cube Views and MQTs

In the context of OLAP, the metadata will contain information about the database-based structure (typically available to the user through graphical and visual representation) so-called business metrics, measurement units that are available, and hierarchies that exist in each dimension.

The presence of metadata increases productivity because it explains the metadata user, the nature of the subject they are looking at. How a map provides a visual representation of the city, giving it a level of immediable understanding of the city's location. The same concepts can be applied to metadata.

Moreover, structured metadata can be used by applications and can increase productivity. If the metadata structure is known then we can write applications that make us use metadata. If metadata are not made available to apps, it's first necessary to discover the nature that is behind them before using them.

In a Business Intelligence implementation, there is a multitude of already defined and available metadata. ETLs may have their own metadata, the MOLAP tool may have its own meditations describing the source data that is used to load the MOLAP database and any transformation that is required to load into a MOLAP database; the tools MOLAP, ROLAP, HOLAP will also have their own metadata describing them, and the dimensions and hierarchies are represented in the database.

In the HOLAP environment, there will also be metadata describing the crossing point of what is in the MOLAP database and what is in a relational database.

All types of OLAP requests require some basic metadata: the cube model and mapping that model to a relational source. A common factor of these scenarios is that at a certain point, a query against the relational database must be generated.

When implementing an OLAP environment, there are a number of current issues an analyst needs to solve in designing an OLAP database.

a) Metadata flow

There is actually a metadata flow that traverses the product set that can be installed in the business intelligence environment. Part of the problem may be that there is no metadata flow but only isolated metadata groups that support different parts of the solution that has been deployed. In this case, the challenge is that every part of the implementation where the metadata is required, it is necessary to take steps to re-recreate the metadata. This process of loss of information and rediscovery is expensive and generates errors. This is not the only problem. If the schema changes, all tools and applications will need to be updated.

b) Optimizing and managing common summary tables

Organizing effective multidimensional data requires the ability to create and organize aggregates and index data across multiple dimensions. Aggregation organization is particularly challenging as demand grows greatly in terms of storing multiple-data data.

For the aggregation of all combinations across all sizes to provide weekly data for the past two years for 1,000 products sold to 500,000 customers, a multi-dimensional space of:

104x1000x500,000x4=208,000,000,000

sales figures.

OLAP models are very rare; they typically include several sizes and require additional storage for aggregates such as months, quarters, years, product groups, and sales regions. The space and time required to produce such aggregates can make the aggregation impractical.

The alternative for aggregating anyone is the choice where aggregates are built within the multidimensional space. In fact, the multidimensional space completely contains the same number of total values, only some of these values are pre-aggregated and stored; the rest are aggregated upon request.

However, the mission of deciding which values should be pre-aggregated is a major challenge for DBA people, and typically involves DBA in analytical work to determine whether pre-aggregation summary tables are set in a manner established by the DBA.

5. EFFICIENT MULTIDIMENSIONAL MODEL. THE CUBE MODEL

The metadata model for viewing cubes with DB2 is intelligible a model capable of shaping a wide range of schematics from simple to complex. The various OLAP tools place relative importance on the different types of objects in the OLAP model. Metadata from DB2 Cube Vids have been discovered with built-in flexibility to accommodate these differences.

The metadata model in DB2 Cub Views is at a good level, a review of what is depicted in figure 3.

Knowledge Horizons - Economics
Volume 10, No. 3, pp. 68 – 76, © 2018 Pro Universitaria

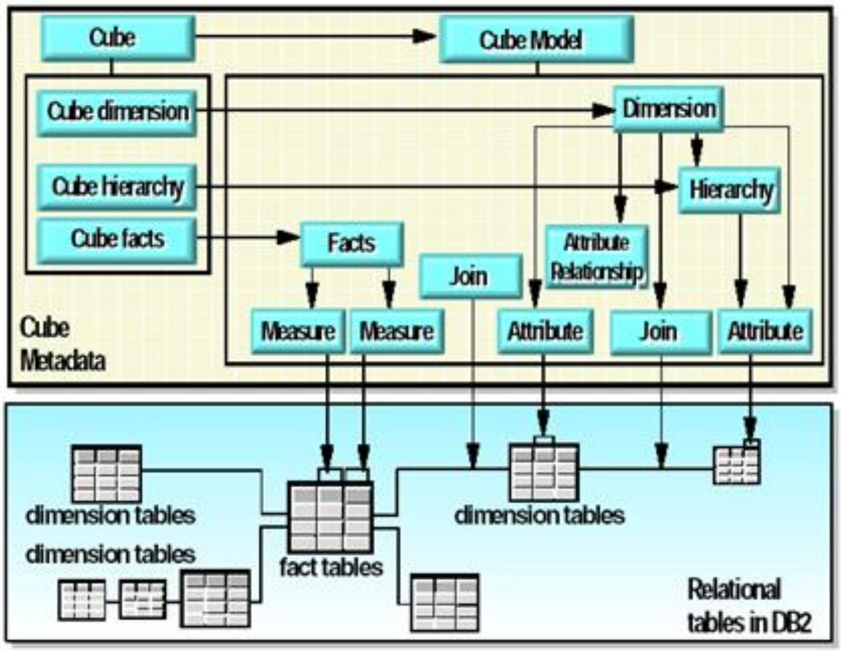

Figura 3 - DB2 Cube Views and MQTs

The figure 3 demonstrates how the cube metadata indicated at the top of the diagram maps the relational tables built into in the DB2 UDB V8.1 indicated at the bottom of the diagram.

Cube metadata define two major structures, the cube pattern and the cube. The cube model can be compared to a conceptual OLAP database. The cube model can be built in several ways. Lists OLAP metadata objects to relational structures in DB2 UDB V8.1. The metadata objects that are stored in the cube model are objects, dimensions, hierarchies, attribute measures, and attribute-to-attribute relationships.

The cube [4] is an extrapolation of the general cube pattern. It is possible to have a cube or multiple cube pattern for the cube model. A cube is the closest object to a MOLAP database. Matadata objects that are stored in the cube are:

- facts cube,
- Dimensions - cube
- Cube hierarchies.

Certain query tools are able to directly link metadata from DB2 Cub View via the DB2 Cub View API.

The metadata user interface in DB2 Cube Views is the graphical user interface on the workstation named OLAP. The OLAP Center is a Java ™ -based utility that uses DB2 UDB V8.1 OLAP is a basic utility that uses the usual classes and maintains the same look and behaves like other DB2 GUI tools. The OLAP Center can be launched by other DB2 UDB V8.1 tools. The architecture of OLAP is described in the picture.

6. OPTIMIZE SUMMARY TABLES. OPTIMIZATION ADVISOR

An analyst using OLAP query tools, capable of sharing data in different ways, in different combinations of levels within the hierarchy for each dimension. Queries that are aggregated on the database will be better resolved if the data needed for these queries have as a source the summary tables rather than the database as the source. The analyst should summarize the Optimization Advisor tables. The interface to Optmization Advisor is a wizard called OLAP Center. The main purpose of the program is to build Materialized Query Tables (MQTs) for certain query queue cubes.

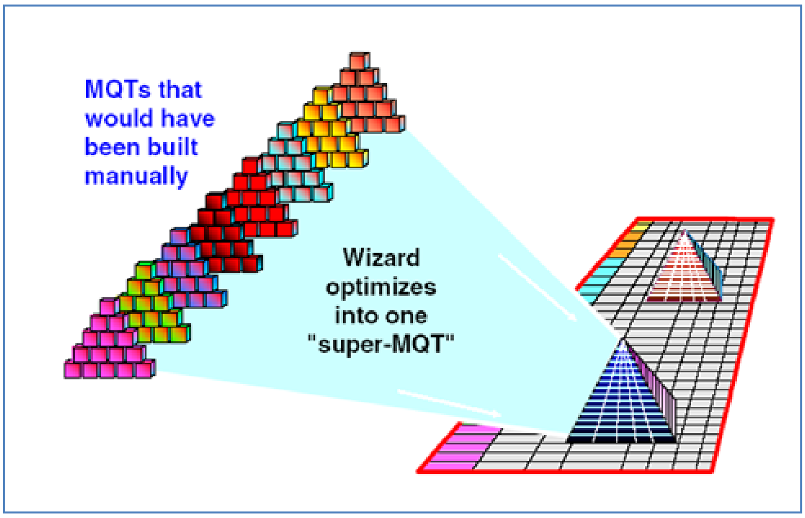

Figure 4 - Optimization MQT

Optimization Advisor takes the input values that are entered into the OLAP Center as input data and the statistics are kept in the catalog tables.
Queries that take advantage of MQT use include:

- types of ROLAP inquiries
- queries used to load MOLAP databases from DB2 UDB V8.1;
- queries generated in a HOLAP environment

7. CONCLUSIONS

Following the study on "Business Intelligence and OLAP," we note that a multidimensional data model is typically targeted at a particular part of a firm, such as a sales model or financial model. At the same time, the central point of this multidimensional model is given by the facts table. I notice that the OLAP model information is stored as a metadata, and the metadata is data about the data. At the same time, ETL tools may have their own metadata, the MOLAP tool may have its own meditations describing the source data that is used to load the MOLAP database and any transformation that needs to be loaded into a MOLAP database; MOLAP, ROLAP, HOLAP will also have their own metadata describing them, and the dimensions and hierarchies are represented in the database.

8. BIBLIOGRAPHY

[1] Alberto Abelló, Oscar Romero, Torben Bach Pedersen, "Using Semantic Web Technologies for Exploratory OLAP: A Survey", Browse Journals & Magazines > IEEE Transactions on Knowledg... > Volume: 27 Issue: 2, 2018

[2] Yansong Zhang ; Yu Zhang ; Shan Wang ; Jiaheng Lu, "Fusion OLAP: Fusing the Pros of MOLAP and ROLAP Together for In-memory OLAP", IEEE Transactions on Knowledge and Data Engineering, DOI: 10.1109/TKDE.2018.2867522, Print ISSN: 1041-4347, Electronic ISSN: 1558-2191

[3] Habermann Ted, Kozimor John, Gordon Sean, "Metadata Evaluation and Improvement: Evolving Analysis and Reporting", NTRS 2017

[4] Sunita Sarawagi, Rakesh Agrawal, Nimrod Megiddo, "Discovery-driven exploration of OLAP data cubes", International Conference on Extending Database Technology, EDBT 1998: Advances in Database Technology — EDBT'98 pp 168-182