

University of Business and Technology in Kosovo UBT Knowledge Center

UBT International Conference

2017 UBT International Conference

Oct 27th, 3:00 PM - 4:30 PM

Decision making based on data analyses using data warehouses

Valdrin Haxhiu

University for Business and Technology, valdrin.haxhiu@ubt-uni.net

Follow this and additional works at: <https://knowledgecenter.ubt-uni.net/conference>



Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

Recommended Citation

Haxhiu, Valdrin, "Decision making based on data analyses using data warehouses" (2017). *UBT International Conference*. 82. <https://knowledgecenter.ubt-uni.net/conference/2017/all-events/82>

This Event is brought to you for free and open access by the Publication and Journals at UBT Knowledge Center. It has been accepted for inclusion in UBT International Conference by an authorized administrator of UBT Knowledge Center. For more information, please contact knowledge.center@ubt-uni.net.

Decision making based on data analyses using data warehouses

Valdrin Haxhiu

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo
valdrin.haxhiu@ubt-uni.net

Abstract. Data warehouses are a collection of several databases, whose goal is to help different companies and corporations make important decisions about their activities. These decisions are taken from the analyses that are made to the data within the data warehouse. These data are taken from data that companies and corporations collect on daily basis from their branches that may be located in different cities, regions, states and continents. Data that are entered to data warehouses are historical data and they represent that part of data that is important for making decisions. These data go under a transformation process in order to accommodate with the structure of the objects within the databases in the data warehouse. This is done because the structure of the relational databases is not similar with the structure of the databases (multidimensional databases) within the data warehouse. The first ones are optimized for transactions on daily basis like: entering, changing, deleting and retrieving data through simple queries, the second ones are optimized for retrieving data through multidimensional queries, which enable us to extract important information. This information helps to make important decisions by learning which are the weak points and the strong points of the company, in order to invest more on the weak points and to strengthen the strong points, increasing the profits of the company. The goal of this paper is to treat data analyses for decision making from a data warehouse by using OLAP (online analytical processing) analysis. For this treatment we used the Analysis Services of Microsoft SQL Server 2016 platform. We analyzed the data of an IT Store with branches in different cities in Kosovo and came to a conclusion for some sales trends. This paper emphasizes the role of data warehouses in decision making.

Keywords: multidimensional databases, multidimensional queries, OLTP, OLAP, fact table, cube, dimension

Introduction

Making decisions in order to increase the profits is an important part of activity of different companies. This becomes even more pronounced by taking in consideration that the main purpose of profitable companies is the greatest profit. Extraction of relevant information from data warehouses by analyzing data that are found within them has undergone a great development especially in the last decade of the XX century, the development of this area continues today. Companies during their work face problems. Their main problem is the decline in the results of their work. Companies in order to identify the work they perform in their various departments, maintain the data that describe that work. These data are stored in an organized form through databases. The volume of data within these databases depends on the jobs that companies perform as well as the size of the companies.

Such databases are mainly relational databases, which organize data in tables that exclude the possibility of repeating the same data (excluding redundancy). These databases are managed by Relational Database Management Systems- RDBMS and are suitable for online analytical processing – OLTP. These systems offer the possibility of using the SQL language. The data that are stored in the databases of a company can be extracted from tables and views according to certain criteria. These systems are used during the daily work of one company department. Within them are stored data on goods that a store in a store chain of a company has available, the price of each product per piece, the amount of products sold, the exact sales time, the exact time of each transaction (each action such as: the data entry, alteration and deletion), etc. From these and other data, through aggregates and analysis can be extracted important information. This information can be extracted and stored inside these systems, after they have been brought in the appropriate form, but it would not have good effect on the performance of these systems, taking into account the large number of transactions within them even though not in large volumes of data. In this way the data from which information can be extracted can be transmitted with the appropriate changes into multidimensional databases within a data warehouse, which is carried out separately from the relational database management system. An important part of multidimensional databases are OLAP cubes, which make it possible to extract data according to some of their dimensions through multidimensional or OLAP queries. Extracting data in this way represents the analytical processing of OLAP data, which enables making important decisions for a company or corporation. Data warehouses have become an inseparable part of any serious company and corporation, because they allow them to increase their results and be more competitive to their rivals under certain conditions of market competition. The data are the blood of any company or corporation, without data they have no idea where they are and what direction they go [1]. Their activity over a long period of time, the collected data present a great value to them. Behind these data, behind the figures found within them, a great asset is hidden. By using this data can be created a data warehouse. Data warehouses contain unified data from various sources, supplemented with summary information and cover a long period of time [2].

Elements of the data warehouse

Constituent elements of a data warehouse are: fact tables, dimensions, dimension tables and cubes. Fact tables are tables that contain business facts. Facts of a business are the data of that business. These data are also known as detail data or measures. Within a data warehouse, data are saved as denormalized to facilitate data analyses. In our case study we created the fact table called Sales which has information about sales amount of different personal computer parts. Dimensions are used to analyze data in a multidimensional way. Data within a data warehouse are numerical values. In order to supplement these data with additional information within a data warehouse are created dimension tables. Dimension tables enable analyzing facts by several dimensions. We implemented three dimension tables: place, product and time because we wanted to analyze data through these dimensions. Data cubes are multidimensional data structures, which enable multidimensional queries to be executed on them. Result sets returned by executing such queries may be used by analysts to analyze data. Cubes are created from fact tables and dimension tables. Data can be used from many fact tables, but usually they are from a single fact table about a certain subject, for example for the subject Sales. In Fig. 1 is represented data cube Sales.

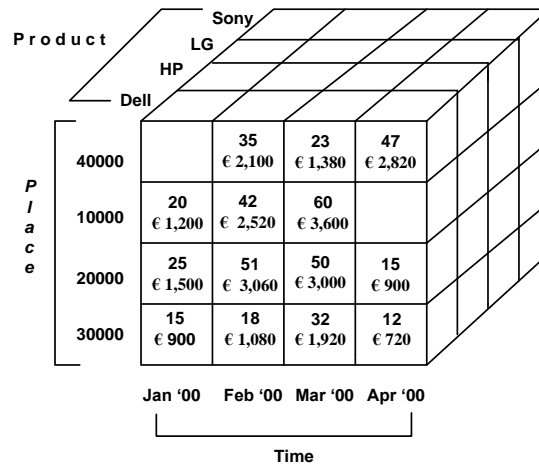


Fig. 1. Data cube Sales

The cube shown in Fig. 1 is created from table Fact_Sales. This cube has three dimensions: time dimension, place dimension and product dimension. On the front side of the cube within small squares or cells there are facts or data from columns: Sales_amount and Price. For example: values 15 and € 900 in the bottom left cell on the front side of the Sales cube represent the sales amount of Dell monitors and the price of those sales in 1st January of year 2000 in the city of Pejë.

In data cubes beside facts, there also can be saved some pre calculated values, such that when a certain query wants those values, the responses could be obtained faster. In Fig. 1, two of the cells on the front side do not have values because in the fact table there are no data for these cells.

OLAP

Analytical processing – OLAP deals with data analysis within a data warehouse. OLAP is a term about a technology that uses multidimensional data representation for fast access on strategical information for advanced analyses [1]. Analytical processing enables analyzing facts through queries. From these facts can be derived important information: knowing about different trends of a company or corporation, what products had most sales and at what place, which clients are frequent clients and what profit has a company had from them, what can be done so that a certain branch increase sales and get more profit, see how much rentable has been a branch since it was open at a certain city, etc. In Fig. 2 is shown the schema of a typical OLAP system. The focus of this paper is data analysis.

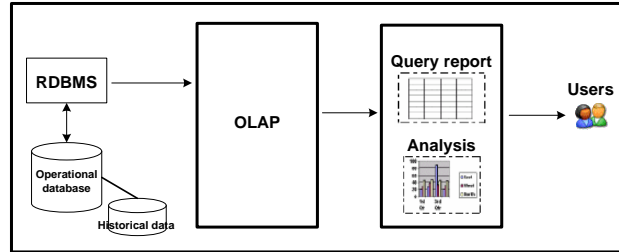


Fig. 2. Schema of a typical OLAP system

OLAP queries

Below we have represented an OLAP query written in MDX (multidimensional query expressions) language. MDX utilizes expressions composed of identifiers, values, statements, functions, and operators that Analysis Services can evaluate to retrieve an object (for example a set or a member), or a scalar value (for example, a string or a number) [3]. This query looks for sales amounts of PC monitors in the first days of January, February, March and April of the year 2000 in cities Prishtinë, Prizren, Pejë and Mitrovicë. After the SELECT statement we put the levels of dimensions through which we wanted to see the sales amount of PC monitors. After the FROM statement we put the name of the cube from which the data are extracted and after the WHERE statement we put the fact through which we extract the data from the data cube.

```
SELECT [Dim Place].[City].MEMBERS ON COLUMNS,
       [Dim Time].[Sales date].MEMBERS ON ROWS
FROM [Sales] WHERE Measures.[Sales amount]
```

Part `[Dim Place].[City]` represents the dimension *Place*, which is the X axis and part `[Dim Time].[Sales date]` represents the dimension *Time*, which is the Y axis. The data or facts about the sales amount of PC monitors are taken from *Sales_Amount* column of the table of facts *Fact Sales*.

By executing this query we obtained the results which are represented in Table 1.

Table 1. The result of the execution of the MDX query

| | All | Mitrovicë | Pejë | Prishtinë | Prizren |
|---------------------|-----|-----------|------|-----------|---------|
| All | 445 | 105 | 77 | 122 | 141 |
| 2000-01-01 00:00:00 | 60 | (null) | 15 | 20 | 25 |
| 2000-02-01 00:00:00 | 146 | 35 | 18 | 42 | 51 |
| 2000-03-01 00:00:00 | 165 | 23 | 32 | 60 | 50 |
| 2000-04-01 00:00:00 | 74 | 47 | 12 | (null) | 15 |

In Table 1, in the columns there are the names of the cities: Mitrovicë, Pejë, Prishtinë dhe Prizren, while in the rows there are the first days of months: January, February, March and April of the year 2000. In this table there is the sales amount of monitors in these cities and months. We have the total sales of monitors sold in every city, the total for all cities, the total of monitors sold in the first day of every month taken into consideration and the total for every of the first days of these months. These totals were calculated by the system.

Based on the results in Table 1 we could do these analyses: from the results table we can see that monitors were mostly sold in the city of Prizren (it is about DELL monitors and the total 141), when it comes to the beginning of the months, most monitors were sold on March 1st of the year 2000 (165 in total), there were no sales of monitors in the city of Mitrovicë on January 1st in 2000, there were no sales of monitors in the city of Prishtinë on January 1st in 2000, etc. The data that we used to make such analyses are within a single table, Table 1. The question then arises: could it be possible to get Table 1 by executing a single SQL query? The answer is no. I have got this answer by trying to write a single SQL query to get Table 1, but it was not possible. I got the data of Table 1, but in 4 separate tables! I had to write and execute 4 SQL queries to get these tables of data. I had to calculate the total 445 by writing this SQL query:

```
SELECT SUM (Sales_amount) AS [Total]
FROM [Computers_DW].[dbo].[Fact_Sales]
```

To get the total of sales for each of the cities: Prishtinë, Prizren, Pejë and Mitrovicë I had to write this query:

```
SELECT [Postal_Code], SUM(Sales_amount) AS [Total]
FROM [Computers_DW].[dbo].[Fact_Sales] GROUP
BY [Postal_Code]
```

In order to get the total of sales for the beginning of every of the months taken into consideration I had to write this query:

```
SELECT [Nr_Date], SUM (Sales_amount) AS [Total]
FROM [Computers_DW].[dbo].[Fact_Sales]
GROUP BY [Nr_Date]
```

To get the sales amount for the beginning of each of the months taken into consideration: January, February, March and April of year 2000, I had to write this query:

```
SELECT [Sales_date], [City], [Sales_amount] FROM
[Computers_Dw].[dbo].[Dim_Time],
[Computers_Dw].[dbo].[Dim_Place],
[Computers_Dw].[dbo].[Fact_Sales] WHERE
[Computers_Dw].[dbo].[Dim_Time].[Nr_Date]=
[Computers_Dw].[dbo].[Fact_Sales].[Nr_Date] and
[Computers_Dw].[dbo].[Dim_Place].[Postal_code]=
[Computers_Dw].[dbo].[Fact_Sales].[Postal_code]
GROUP BY [Sales_date], [City], [Sales_amount]
```

MDX is suitable for multidimensional data analysis. SQL requires more work to accomplish the same job.

Conclusion

Understanding trends by conducting multidimensional data analysis through OLAP analysis can be very useful. Knowing the direction a company is going is very important for staying in the business and competing with other businesses. Different products may have different sales trends. Analyzing sales data by their dimensions enables seeing what has happened with sales during a period of time. A certain company may have branches in different cities or regions. OLAP analysis makes it possible to analyze data from many points of views to see for example: what were the sales of a certain product, at a certain point in time and at a certain place. Based on the information gained from data analyses, companies may make decisions about their ongoing activities, they may for example decide to shut down stores on certain branches

because their sales revenue was not satisfactory or they may take measures to make things get better. The tools used for OLAP analysis in this paper were sufficient enough for our case study.

References

1. . John Wang: Encyclopedia of Data Warehousing and Mining, Volume 1 & Volume 2, Idea Group Inc., (2006)
- 2.
3. . Raghuram Ramakrishnan & Johannes Gehrke: Database Management Systems, second edition, McGraw-Hill Higher Education, (2002)
4. <https://docs.microsoft.com/en-us/sql/mdx/multidimensional-expressions-mdx-reference>