# Preparation of data for visualization using SQL Server 2008

Emilija Taleska[1], Snezana Savoska[2], Jasmina Nedelkoska[3]

***Abstract*** **-** Analysis of a huge amounts of data that flow into databases, successfully dealt with methods of data visualization that graphical representations of data and information provide better awareness of the management and making better decisions. Information describes things that are sometimes abstracting in schematic format that includes analysis of attributes and variables such units of information.
The purpose of this paper is to study the practical concepts and algorithms for preparing the data for visualization and creating visual representations of large amounts of data. In this paper will be given attention to the way in which data is transformed from the original format into a format suitable for their further processing using algorithms.

***Key words –*** Data visualization, database, algorithms, preparation for visualization

## I. INTRODUCTION

The data are required and important component in daily life, business, medicine, administration, science, etc. However, management and access to data is not always easy. Daily data are stored in different formats, relational tables and databases. To be able to analyze them it is necessary to use appropriate methods for data visualization, which facilitates the presentation and understanding of data and information. This enables increasing efficiency in the work of analysts and managers.

The design of the database is the process of producing a detailed data model for the database. It can be conceptual, logical and physical. The model contains all the needed logical and physical design choices and physical storage of all parameters needed to generate a design in a Data Definition Language, which can then be used to create a database. The full data model contains detailed attributes for each entity within the database management system (DBMS).

[1] Emilija Taleska is with Faculty of Adminstration and manadment of Information Systems, Bitola, Macedonia
[2] Snezana Savoska is with Faculty of Adminstration and manadment of Information Systems, Bitola, Macedonia
[3] Jasmina Nedelkoska is with Faculty of Adminstration and manadment of Information Systems, Bitola, Macedonia

## II. PARADIGMS ASSOCIATED WITH DATABASE MANAGEMENT SYSTEMS

"The main purpose of a database management system (DBMS) is to bridge the gap between information and data - the data stored in memory or on disk must be converted to usable information.

The basic processes that are supported by a DBMS are:

- Specification of data types, structures and constraints to be considered in an application.
- Storing the data itself into persistent storage.
- Manipulation of the database.
- Querying the database to retrieve desired data
- Updating the content of the database

A database is a model of a real world system. The contents (sometimes called the extension) of a database represent the state of what is being modeled. Changes in the database represent events occurring in the environment that change the state of what is being modeled. It is appropriate to structure a database to mirror what it is intended to model."[1]

Databases can be analyzed at different levels:

- Conceptual Modeling. These concepts allow application modeling the world in terms that are independent of the particular data (logical) model. Conceptual models provide a framework for the development of a database schema from the top to the bottom in the process of database design. This model examines the entity-relationship and object-oriented model to represent conceptual modeling. The entity-relationship model is widely used and the object-oriented model is gaining more acceptances of non-traditional databases.
- Logical Modeling. At the logical level, the conceptual schema is translated into the data model of a particular DBMS. A logical model is described as a set of relatively simple structures. In addition to the data representation, a DBMS needs to specify the handling of data, which is done through queries and other operations in a data manipulation language.
- Physical Modeling. Al the physical level, a DBMS is responsible for:
  - Storage. The representation of the efficient organization of data into a persistent secondary drive.

- Access Methods. Organization of data to accelerate data recovery through defining data structures or index.
- Processing queries. The set of operations for responding to a query. Those transactions define algorithms that make use of access methods.
- Query Optimization. Evaluation strategy for query processing.
- Concurrency and recovery. Strategy for managing concurrent access to data and resources from multiple users and recovering the database after a system failure.

Data visualization is the graphical representation of abstract information for two purposes: the sense-making (also called data analysis) and communication. Important stories live on our data and data visualization is a powerful means to discover and understand these stories, and then present them to others. The information is summarized in describing things that are not physical. Statistical information is abstract. Whether the case of sales, the incidence of the disease, athletic performance, or anything else, although not in the physical world, you can still show visually, but for this we have to find a way of shaping it has none. [2]

In the modern operating environment, any serious application must simultaneously confront for at least four paradigms:

- Paradigms associated with data or data paradigms, which means that data is used by a single application.
- Resources paradigm that involves everything about resources that may be required of the application in data processing (type of resources and the manner of their usage).
- Communication paradigms, which can be divided into three subsets: communication with the database, communication and coordination of components (object modules) that make up the application, and finally, communication with other utilities (services). These are paradigms that provide data access only by authorized person and no one else.
- Security paradigms. Security paradigms involve collection of measures and procedures that must be completely observed, so that an individual can work with data.

## III. THE NEED FOR PREPARING DATA FOR VISUALIZATION

The data stored in a data warehouse is usually hard and is not suitable for data visualization. Therefore, to obtain visual representations the data need to be prepared. Sometimes it is important to extract the necessary data and put the data into separate repositories, even recode them and normalizing their values. But when it comes to large amounts of data, it is still necessary to aggregate forms or views (Views) to accelerate the visual representation. During the preparation of data, it is important to know the amount of data which is represented, its dimensions and data interaction. Because of all these factors, it can be said that the preparation of data for visualization is not at all an easy process. [3]

When it comes to data used in the health care and preparation for display, we can say that this is a complex process that requires a detailed knowledge of data (bottom-up analysis) and detailed knowledge of the administrators that need visual reports, because it is necessary to know the transaction information systems and to make them appropriate repositories of data that are required to manage.

In Macedonia transactional information systems are placed on multiple platforms, databases and applications. Its integration inevitably requires a uniform deposit of data that will be the basis for the visualization. Some of the data is encoded differently in different transactional systems and needed to be stored in recoded parallel codes. After defining the logical and physical model of the repository that will be used for data visualization, it is necessary to create views that will contribute to get an efficient and effective data visualization.

## IV. PROCEDURES FOR PREPARING DATA FOR VISUALIZATION

For the research that we made and received information about the needs of managers in health institutions in Macedonia, was required to obtain certain information to be presented visually through dashboards. For this purpose, have been defined procedures for the preparation of the data.

The software that is used is conceptualized in classic reports in Excel or pivot tables. The reports include analytical data. Managers and analysts themselves can choose which data they want to see in the reports. But this method of selection is used often by our site managers from our side analysts. The data which were stored data on drugs and their consumption by departments, the financial plan for the budget accounts and subaccounts, the realization of the budget accounts for certain quarters, codes for performed services, diagnoses and other statistics.

Results Messages

| | SmetkaID | KvartalID | KontoID | Planirani | Fakturirani | Plateni | Ostatok | Procent |
|---|---|---|---|---|---|---|---|---|
| 1 | 73729 | 1 | 401 | 36645819 | 0 | 0 | 36645819 | 0 |
| 2 | 73729 | 1 | 401130 | 0 | 0 | 33596562 | -33596562 | 0 |
| 3 | 73729 | 1 | 401310 | 0 | 0 | 2504124 | -2504124 | 0 |
| 4 | 73729 | 1 | 401320 | 0 | 0 | 0 | 0 | 0 |
| 5 | 73729 | 1 | 402 | 13554123 | 0 | 0 | 13554123 | 0 |
| 6 | 73729 | 1 | 402110 | 0 | 0 | 8902970 | -8902970 | 0 |
| 7 | 73729 | 1 | 402210 | 0 | 0 | 3610661 | -3610661 | 0 |
| 8 | 73729 | 1 | 402220 | 0 | 0 | 247292 | -247292 | 0 |

| | Godina | Prihodi | Suma | log | ln |
|---|---|---|---|---|---|
| 1 | 2009 | Камати и депозити | 6959 | 3,84254683649502 | 8,84779106484485 |
| 2 | 2009 | Приходи од продажба на лекови | 1013431 | 6,00579418485995 | 13,8288521616488 |
| 3 | 2009 | Приходи од спец.конс.здравствена заштита | 402637 | 5,60491368152504 | 12,9057906905979 |
| 4 | 2009 | Приходи од здравствени услуги | 6008879 | 6,77879345877087 | 15,6087487666514 |
| 5 | 2009 | Закупнина на деловен простор | 854452 | 5,93168767066044 | 13,6582156067593 |
| 6 | 2009 | Приходи од минати години и тенд. | 154000 | 5,18752072083646 | 11,9447078813958 |
| 7 | 2009 | Донации | NULL | NULL | NULL |
| 8 | 2009 | Други вонредни приходи | 3767721 | 6,57607873577514 | 15,141980867351 |

| | Godina | Trosoci | Suma | log | ln |
|---|---|---|---|---|---|
| 1 | 2009 | Исплатено бруто плати | 227278461 | 8,35655827993051 | 19,241686524104 |
| 2 | 2009 | Трошоци за превоз | 187650 | 5,2733485687491 | 12,1423338045632 |
| 3 | 2009 | Трошоци за храна | 2334922 | 6,36827237717819 | 14,6634890438163 |
| 4 | 2009 | Надомест за одвоен ... | 888135 | 5,94847898525929 | 13,6968790374464 |
| 5 | 2009 | Отпремнина за пенз... | 563730 | 5,75107114726967 | 13,2423306924513 |

Fig. 1. Display of data that are briefly generated reports

The next step is to create a repository in parallel codes that are created. But in the system as a problem arises in aligning the codes of different drugs were enrolled in the Fund for Health Insurance and pharmacies. The beginning of this year it has been overcome and the system started working properly.

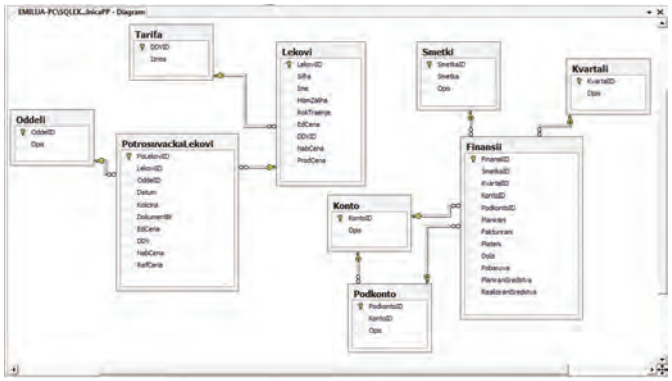


Fig. 2. Database diagram represented by the star schema

In the design of relational databases (RDBMS), normalization of databases is the process of organizing data to reduce the burden. The purpose of normalizing the database is to break the anomalies to make a smaller, well structured relationship between data. Normalization usually involves dividing large tables into small tables and define relationships between them. The task is to isolate data for gathering, deletions and changes in the field can be done at one table and, as such, to operate across the entire database, defined by relationships.

The next step in data preparation for visualization is process of creating views. By considering the original data, views depend on the visual representation and the whole process of preparation of the data is shown in Figure 3.

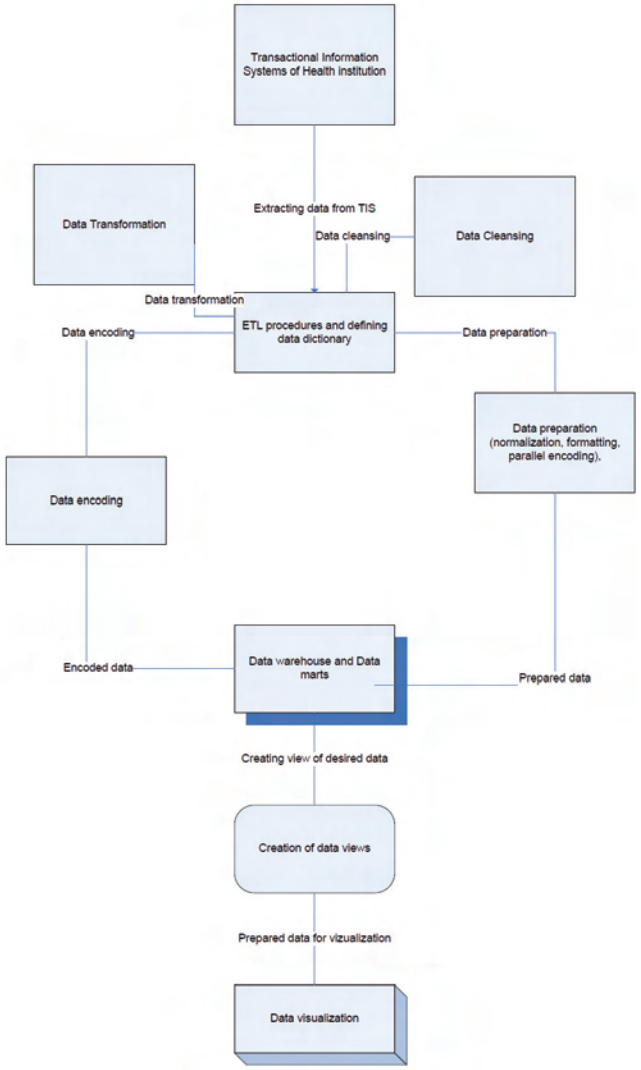


Fig. 3. The process of preparation of the data for visual representation

## V. CONCLUDING OBSERVATIONS

All these requires accessibility to databases as in Republic of Macedonia is a problem considering to outsourcing the software that is used and the lack of staff in some health institutions responsible for database administration. There are problems with inconsistency in transaction systems, with different ways of coding the entities and their attributes, and sometimes with the relations between them. In fact, have not yet been made appropriate integration of the health system and therefore it is difficult to get to the data.

This way of presenting the data is preferred by managers of health institutions, but it needs a serious approach and responsible people who will administer data warehouse and will take care of proper preparation and implementation of procedures, their automation and continuous work.

## REFERENCES

[1] Andrea Rodriguez, "Brief Introduction to Database Concepts", Department of Computer Science, 2004
[2] M. Kaufman, "Information visualization – Perception for design, 2005
[3] E. Turban, Volonino L., Information technology, Wiley, 2011
[4] A. Gunasekaran, Handbook On Business Information Systems, Word Scietific Publishing.Co.Pte.Ltd, 2010
[5] http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html