SAS Data Warehouse and its Usage in Government Public Sector

Snezana Savoska[1], Violeta Manevska[2], Silvana Kolevska[3],

[1] Faculty of Administration and Management of Information systems, University „St.Kliment Ohridski", Partizanska bb, 7000 Bitola, R.of Macedonia , www.famis.edu.mk

Abstract: Data Warehouses are one of the most complex IT issues whose implementation demands much organizational preparation. In fact, its successful implementation demands specified levels of organizational culture and management's full support. Nevertheless, when data warehouses are owned by the government, a very professional approach in the creation of data warehouses, as well as usage of procedures for creating specifically prepared information for dissemination in the public sector, are imperative. Our research on the application of data warehouses in the Republic of Macedonia was focused on the biggest business companies, some governmental institutions and higher education institutions. We found out that there is an increasing usage of data warehouses especially in the business sector. However, the most complex data warehouse in Macedonia is the one with the State Office of Statistics (SOS). This is the backbone of the Macedonian public administration which aims at gathering, analyzing and dissemination of relevant information for entities which are living and working in the Republic of Macedonia. This implies that it is a very important state institution accountable for collecting and storing official information with a very serious approach towards data warehousing information for statistical data analysis. The serious role which has been attributed to this state institution requires a professional software tool and that is why SAS Data integration services tool has been selected.

Keywords: Data Warehouse**,** State Office of Statistics, Data visualization.

1 Introduction

Data Warehouses are one of the most complex IT issues which demand much organizational preparation for its implementation. First of all, the organizational culture must be minimal on the defined level and, also, there must be management's full support for successful implementation. When the government is the owner of a data warehouse, there must be a very professional approach in the data warehouse creation and also the usage of procedures for the creation of specifically prepared information for the public sector. Our research on the usage of data warehouses in the Republic of Macedonia was focused on the biggest business companies, as well as some governmental and higher education institutions. We found DW solutions from

the biggest IT global company as Oracle, SAS, SAP, and Microsoft. There is an increasing application of data warehouses especially in the business sector, where we have identified usage of OLAP multidimensional databases and Analysis services as pivot tables and as end-user's reporting tools.

2   The usage of DW in the public administration

However, the most complex data warehouse in Macedonia was located in the State Office of Statistics (SOS). SOS is the cell of the Macedonian public administration which aims at gathering, analyzing and dissemination of all relevant information for entities which are living and working in the Republic of Macedonia. This fact implies that this is the most important state institution accountable for collecting and storing of official information. It has an extremely serious approach in collecting, cleansing, analyzing and storing information for statistical data analysis. This serious role attributed to this state institution requires a professional software tool for DW.
Data which are included in SOS DW usually are taken from many administrative sources; areas which gain data on the basis of statistical research carried out by SOS, as well as data from the Statistical Register in SOS.

3   The data sources for State Office of Statistic of DW

The administrative sources comprise the Central Register of the Republic of Macedonia (subjects' final finance reports); Public Revenue Government Service (taxes information database), Ministry of Finance (public revenue and expenditure – treasure account), National Bank of Macedonia (monetary statistics), Health Insurance Fund (health insurance contributions) as well as Pension and Disability Insurance Fund (pension and disability insurance contributions).
The statistical research comprises the areas of foreign trade, internal trade, industry, construction and hospitality, agriculture, household consumption surveys, price statistics and labor force survey.
 As a source of statistical registers SOS uses the Register of business entities, which is used in the stages of data preparation for loading data in the DW.
The entire operation of the DW is carried out in the SAS DW environment. The source of data for DW are previously taken from defined data sources and prepared with Microsoft Visual Studio.NET (Visual basic and ACCESS) or some other procedures for data preparation. The tool used for importing data from the prepared data sources (saved in the auxiliary lactation) is SAS Data integration studio (Figure 1). It is a very powerful tool which provides working with many data repositories located in data servers or external sources. Procedures made for this purpose are used for incorrect and missing data; their transformation; conversion of characters written with the Macedonian language support in ACSII codes and are prepared in format apposite for its first loading in the DW (first run). These procedures are created with SAS BPM [3, 4] tools with manual code corrections and they aren't completely

automated (because it is impossible to make them automated to a high degree). They are presented with modeled software business processes (Figure 2).
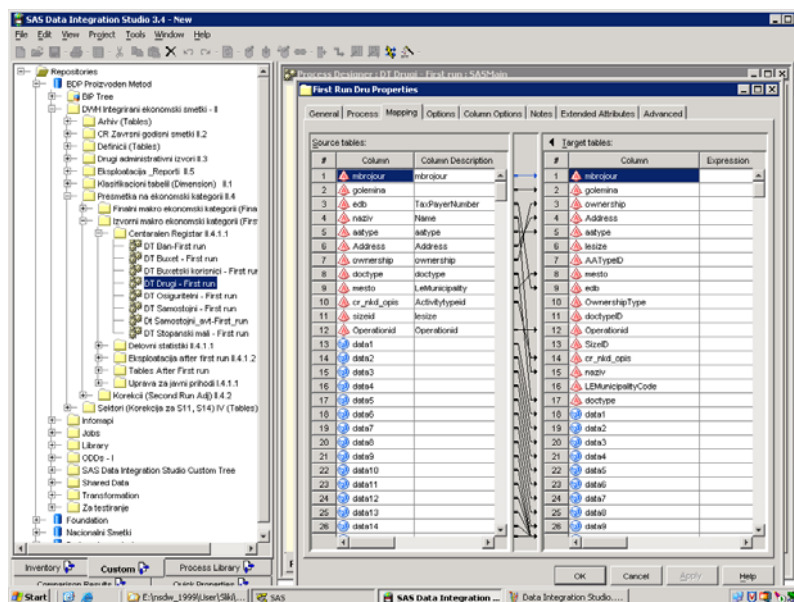

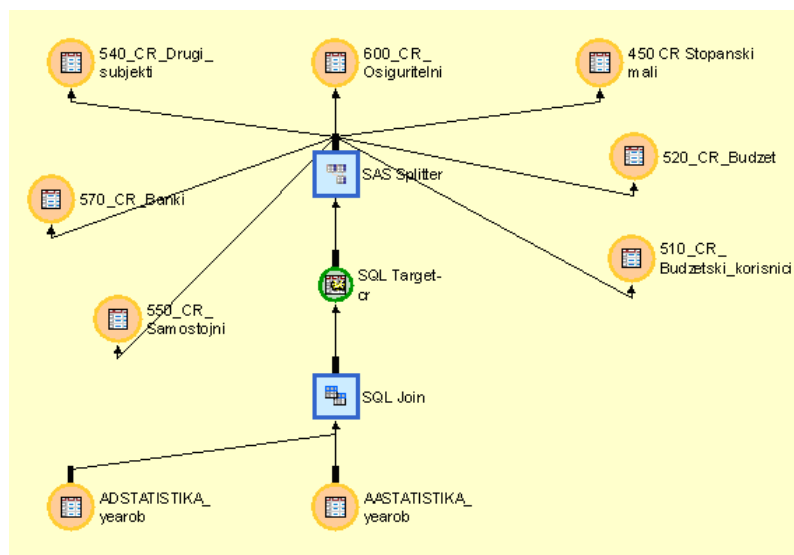**Fig. 1.** SAS Data Integration services screenshot

Fig. 2. Modeled software business processes which are objects with embedded program code

4 Metadata of SAS Data Warehouse

The metadata can be different depending on data types – source data, its cleansing and transformation, intermediate data and target data, its aggregations and the manner of preparation for the final run processes. The data can be described in different manners. In this case, all metadata descriptions, source data transformations, business processes preparation for the different phases are located in the Excel file named Cathegories.xls. In this complex file, there are business logic procedures, statistical analysis demands and algorithms for demanded aggregated tables creation and reports located (Figure 3). Moreover, this file contains intermediate calculations' results. Some of the products of these calculations are located in the temporary file directories whereas the other ones are located in annual data processing procedures and annual data warehousing locations.

Fig. 3. Excel file Cathegories.xls which is the core file for defining metadata and business processes in the DW

5 The stage of Data Transformation and DW updating

In the stage of DW updating with the business entities' final financial reports, the procedure includes merging data for each entity with fact and dimensional tables of DW. The data from the database of the Central Registry are obtained in the ACCESS

database tables which are saved in the source data folders and then they undergo a procedure for data transformation in the format comprehensive for SAS DW. These dimensional tables contain the territorial classification which is valid for the Republic of Macedonia; the institutional sector to which the data belong; Classification of National Activities (CNA) etc. This data are the base for new calculations of the variables, called categories for GDP calculation (gross value, intermediate consumption, value added, number of employees, assets employed, depreciation etc). The schematic display of the processing of source data until reaching final calculated data is shown on Figure 4, whereas the grouping of aggregated data according to sector affiliation is presented on Figure 5.
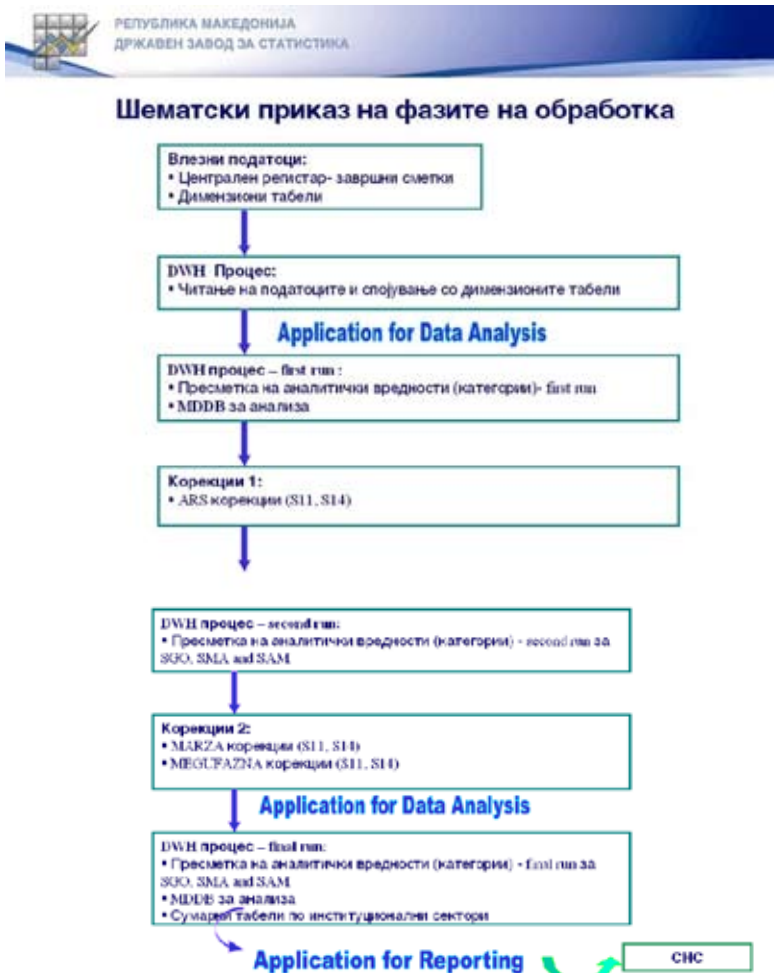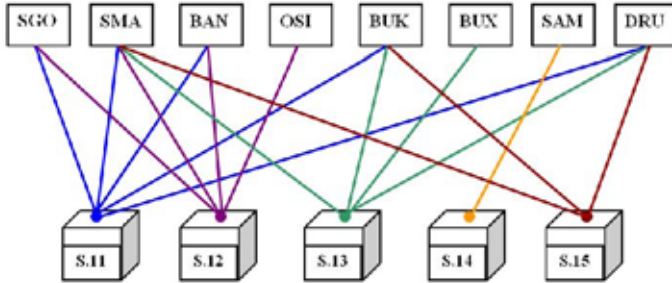

Fig. 4. Phases in the process of data preparation for DW loading stage

Fig. 5. Aggregated tables according to state institutional sector

SAS DW owns many tools which have different functions. The SAS ETL tool possesses many other parts whose licenses are sold separately. The SAS products witch are used in SOS are: SAS Base; SAS Data integration studio; SAS Adjustment analysis; the tools Corrections and Reporting (Figure 6).

Fig. 6. Screenshot of Adjustment Analysis tools, Corrections – Data Analysis

Dissemination of data prepared in DW is another aspect [5] of the end user's data usage. For this purpose, the most suitable tool is the SAS Reports tool (Figure 7). These tools provide selection of many dimensions, time periods and lists of offered reports.
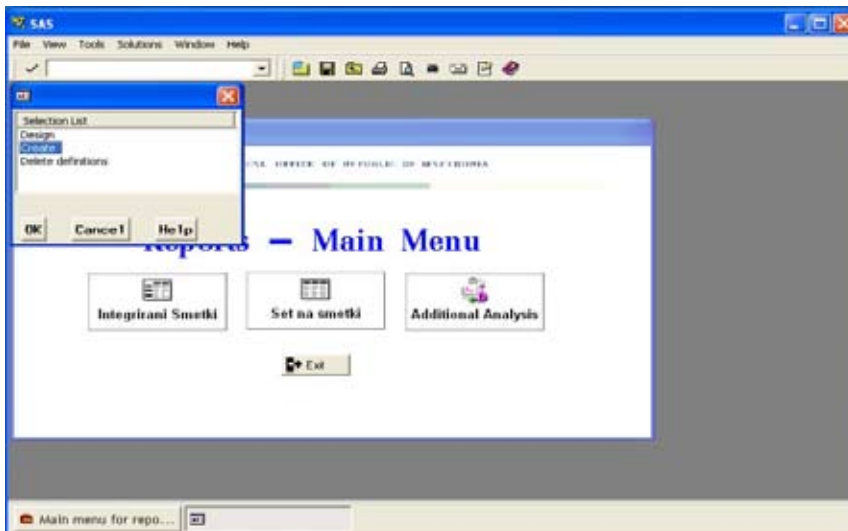
Fig. 7. Screenshot of SAS Reports main menu

6   Publication of data from SSO Data Warehouse

However, the publication of the prepared data in the DW is not a DW administrator's task [1]. Dissemination of public data is entrusted to the SOS sector for public relation whose mission is to provide public access to SOS data via SOS Web site (http://www.stat.gov.mk). The web site is a great combination of information intended for public usage. The data from DW can be seen with the choice "Databases" where we can find statistical analysis of data of interest combined in tabelar format and appropriate visual format. The user access to the web viz browser with a modern HCI provides selection of desired data, sectors, geographic location and format of report. When the desired selection is filled, the data are extracted from DW flat and dimensional tables (mostly from aggregated tables previously prepared in the DW depending on suitable procedures). There is a tool offered for specific data visualization format PC-AXIS for Windows, which can be downloaded from the site and which provides a special format compatible with EU statistics standard tools which can give maps, graphs, diagrams and tables [2]. Figure 8 presents an instance of a screenshot of some statistical population data (visual region representation, statistical data for specific characteristic and thematic scope of categories which are gained from selections www.stat.gov.mk).
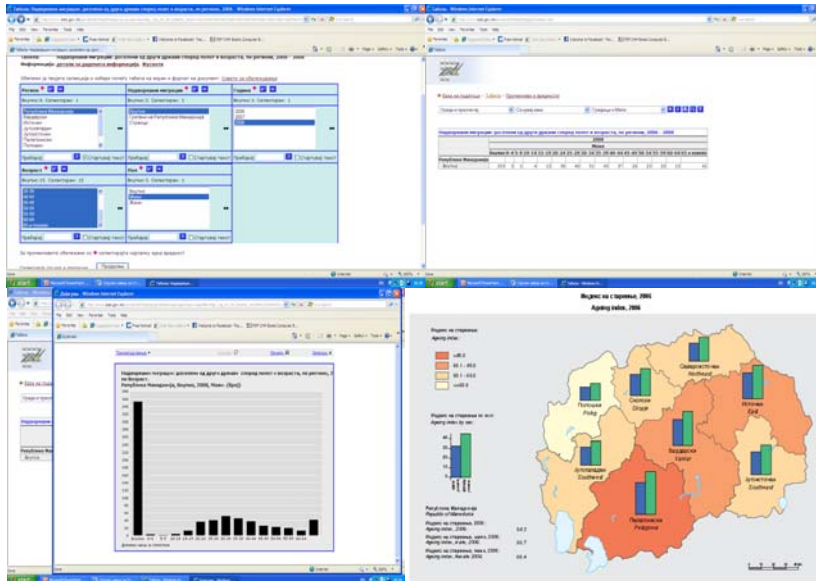
Fig. 8. Example of SOS web site – step by step

7 Conclusion

When talking about most complex data warehouse in Macedonia, we think about SAS DW. It is located in the State Office of Statistics (SOS) and is a cell of the Macedonian public administration. Its purpose is gathering, analyzing and dissemination of all relevant information for entities which are living and working in the Republic of Macedonia. This serious role attributed to this state institution requires a professional software tool for DW and also professional tools for data dissemination via web site.

References

1. D.Keim, H.P.Kriegel, M.Ankerst, Recursive Patterns: A Technique for Visualisation Very Large Amounts of Data, Proceedings of IEEE Visualization '95, 1995, pp. 279-286
2. J.Becker, D.B¨urkle, R.T.Happe, T.Preußer, M.Rumpf, M.Spielberg, R.Strzodka, Aspects on Data Analysis and Visualization for Complex Dynamical Systems, 2000
3. www.sas.com, SAS® OLAP Server, www.datawarehouse.com
4. www.datawarehouse.com, SAS Extends Support for Open Metadata With Meta Integration Technology, Inc. , 7.3.2006
5. J.Yang, et al. , "Value and Relation Display for Interactive Exploration of High Dimensional Datasets", Proceedings of InfoVis '04, Austin, TX, Oct. 2004, pp. 73-80
6. www.stat.gov.mk, 2010, Web site of State Office of Statistics