CASE STUDY: OLAP CUBES IN LABOUR INSPECTORATE

Bologa C., G. Faur, P. Bresfelean
Babes Bolyai University of Cluj Napoca, Romania, Business Information Systems Department, Faculty of Economics, Teodor Mihali 58-60, 400591, Cluj Napoca, Romania
e-mail: cristian.bologa@econ.ubbcluj.ro

Abstract. *This paper is intended as a case study of data analysis by means of an OLAP system with real data of the Area Labour Inspectorate of Alba County to prove both the efficiency and the need to create such a system. Our intention is to prove the beneficial role of OLAP technology implementation in public institutions, at any level which may involve a managerial act.*

Keywords: OLAP technology, public institutions

INTRODUCTION

In public sector, especially in public institutions subordinated to ministries, information analysis is currently performed empirically without any analysis procedure which should be based on the use of some models or even the use of collected data analysis tools. At the same time, the issue of data turning into conclusive information for a public institution does not benefit from an efficient development framework, on the contrary, the idea of systemized data analysis lacks completely in the case of most public institutions.

A summary research conducted by us at the level of four types of institutions subordinated to the Ministry of Labour, Family and Equality of Chances, namely: the General Directorate of Labour and Social Protection, the National Unemployment Office, County State Pension office, Area Labour Inspectorates outlined the fact that none of these institutions either owns or have ever implemented a collected data analysis model in 36 counties, although such data are more than significant as each of these institutions has several million entries on a monthly basis at the national level.

It shows that most public institutions subordinated to the Ministry of Labour do not have a data analysis system, although they have been collecting such data on an organized manner for more than 10 years, and the number of records altogether exceeds the number of thousands of million records. Moreover, none of the institutions comprised in this case study owns or has a Data Warehouse, DW implementation project. We have also noticed that none of the institutions subject to our research have ever used an OLAP tool for data analysis and there is no clear view at the management level of these institutions with respect to the creation of department Data Marts.

I. CASE STUDY: OLAP CUBES AND LABOUR INSPECTION

Dimensions represent an essential and distinct concept in multi-dimension databases. The most important purpose of multi-dimension modelling is the use of dimensions to supply as much context as possible for facts.

Although the cube term leads to the idea of three-dimensionality, that is the existence of three dimensions, in fact, most cubes that can be seen in practice have 4 to 12 dimensions.

To make a brief summary of OLAP-related means and benefits we will try to define such a cube by means of Microsoft SQL Server 2005 Enterprise edition tool.

The analysis shall be made on an operational relational database relating to the collection of information on self-assessment forms concerning the commission payable by every private employer to Area Labour Inspectorate in the country (to which it belongs) for the records and the certification of Employment books.

Database operates similarly in all the counties all throughout the country as both the database and fat-client applications managing it are elaborated by the author. Databases are elaborated by means the SQL Server 2000 Standard edition at the level of each county. These databases are replicated at the Area Labour Inspectorate by means of an explicit architectural pattern for a better management of any possible errors. Basically, this database comprises the payments made by economic agents within every county both through the institution cashier's office and through payment orders within the bank. Moreover, the database comprises self-assessment forms submitted by economic agents. The database at the national level is created in PostgresSql where the self-assessment form table consists of about 12 million entries.

The analysis shall be carried out by importing the central database on Postgres server in SQL Server 2005 Enterprise Edition and processing was operated on the entire set of entries existing as of July 31st 2008. The part of the database schema on which analysis is to be operated is shown in figure 1. We shall consider the **fact** table as 'com_declun' because the database is relational, and the county code identifier is included in companies' table.

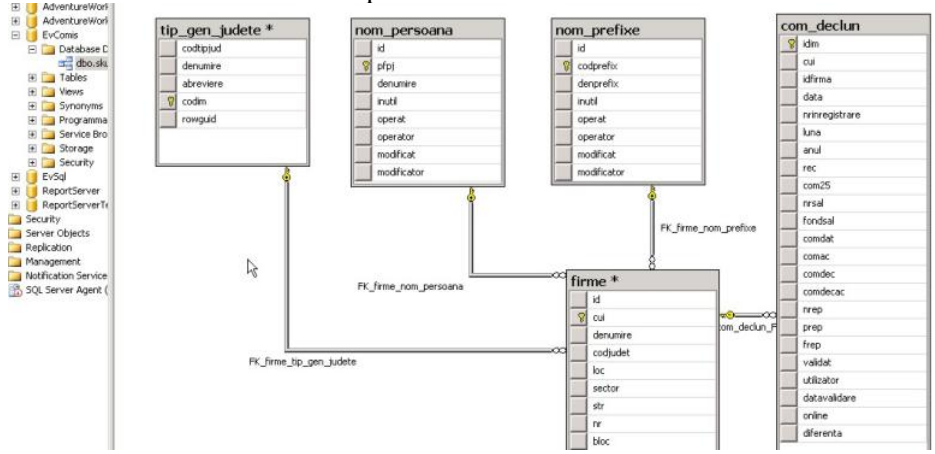


Fig. 1. Part of database schema for analysis

Moreover in companies' table, the following are put into relations: company's type (physical person or body corporate), organization form: trading company, public limited company, bank, etc.). Table's structure comprises also a logic-type field 'com25' identifying trading companies paying a 0.25%- commission of wage fund by means of the „true" value while the remaining companies pay a 0.75%- commission of this fund. To define the cube we shall use a snow flake schema related to fact table only by „CUI" relation standing for the economic agent's Tax Identification Number. All other tables in the image shall be considered dimensions. We shall also add "time" dimension in relation to fact table.

The schema of the cube for analysis is exemplified in figure 2. One can notice the fact table with a yellow header and snowflake architecture around „companies" dimension, while dimensions are represented with a blue header.

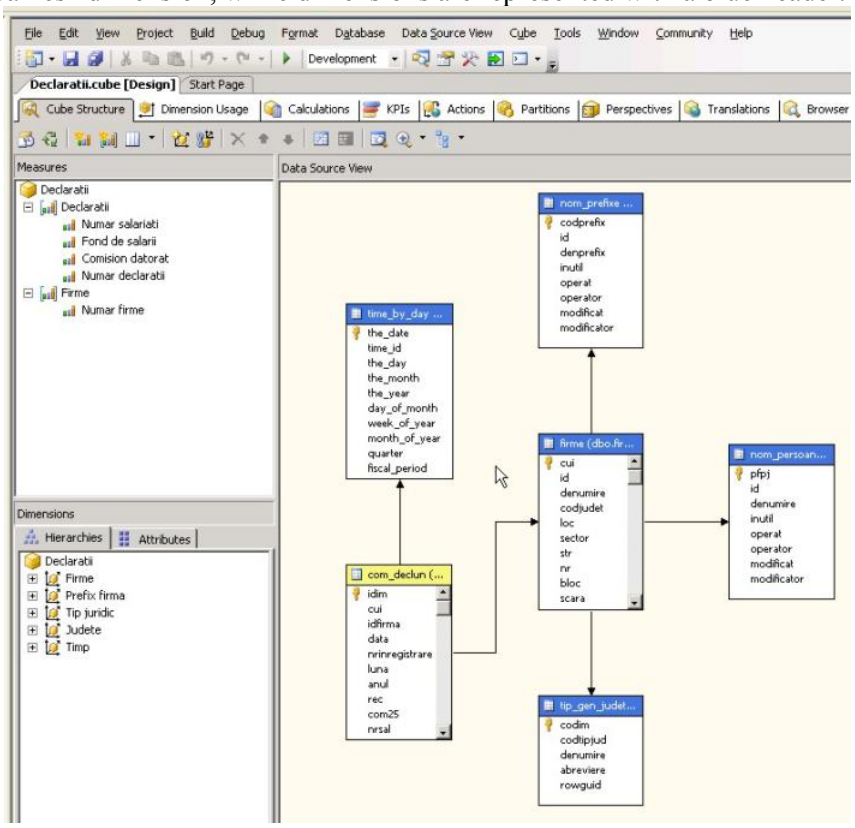


Fig. 2. Schema of the cube for analysis with fact table and dimension tables

Cube processing method is a multidimensional online analytical processing (MOLAP) generating a storage space for the efficiency of operations carried out on the cube. The processing on SQL Server 2005 is shown in figure 3.

To process this cube we resorted to a Compaq ProLiant Server with 1 GB RAM memory and two XEON processors of 2.20 GHz. Processing was made in parallel, the SQL server was set to use both processors. The number of entries involved was of 11,914,402.

The measures established for the cube were the following: number of companies having submitted self-assessment forms, number of self-assessment forms, number of employees of each company, wage funds, payable commission.

The dimensions established are as follows: time measure, related to the date when the self-assessment form has been submitted, the county where the self-assessment form was submitted, company's legal status, company's type of organization and type of payable commission (0.25% or 0.75%).

We have to mention that cube processing took 58 minutes on the server with the given specifications and all processors functioned at a maximum capacity of 100%, as one can see in figure 4. We have to mention that two queries were necessary for cube processing due to the measure „Number of companies" which is a calculated dimension. This is calculated by a distinct selection of tax identity number in the fact table, which also involves the „companies" dimension; thus selection is made twice.

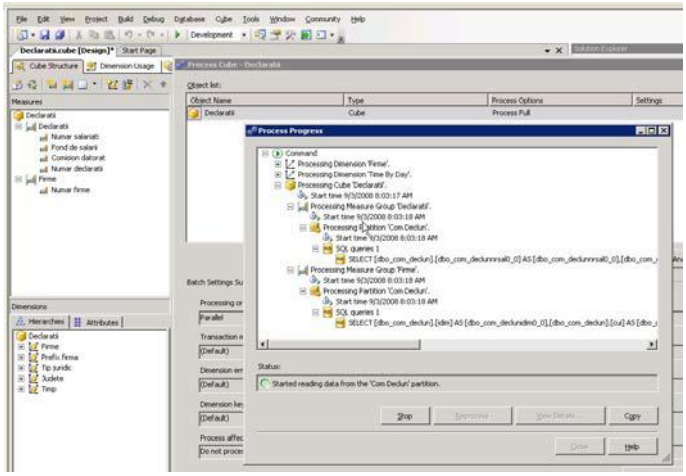


Fig. 3. Cube processing in SQL Server 2005

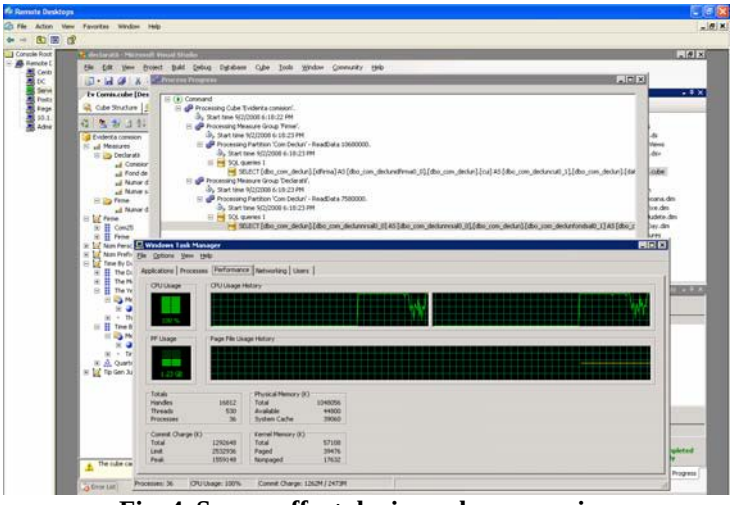


Fig. 4. Server effort during cube processing

As one can see, processing effort requires huge system resources and therefore the physical machine on which analysis is performed should be an extremely powerful one, this cube being extremely simple but there are much more dimensions, especially measures, in real life. Moreover, processing can be carried out by means of several fact tables, and therefore we may conclude that OLAP server requires an extremely powerful machine, at national level at the least. This machine may be an 8-16 processor server or, preferably, a mainframe.

Figure 5 exemplifies drill-down operation on "time" dimension on the processed cube.

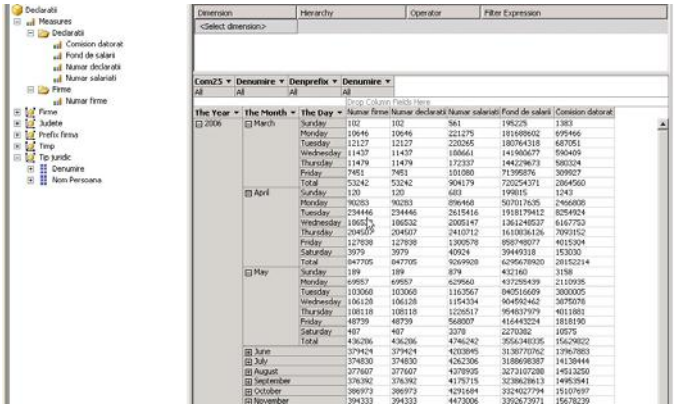


Fig. 5. The Cube obtained and a drill-down operation in "time" dimension

Figure 6 exemplifies slice operation by dimension "county".

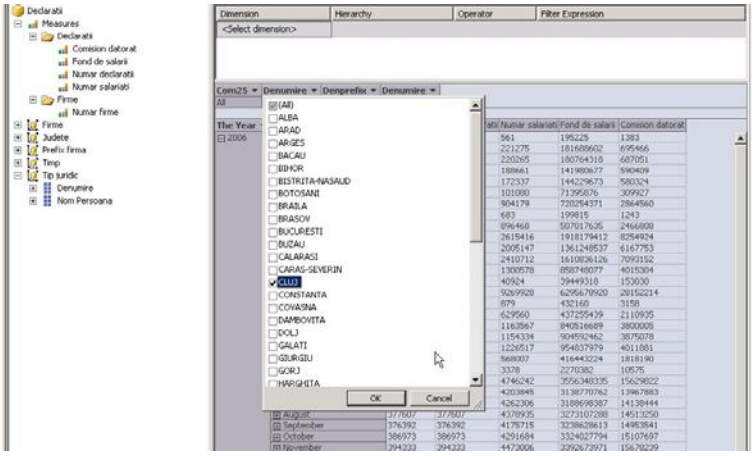
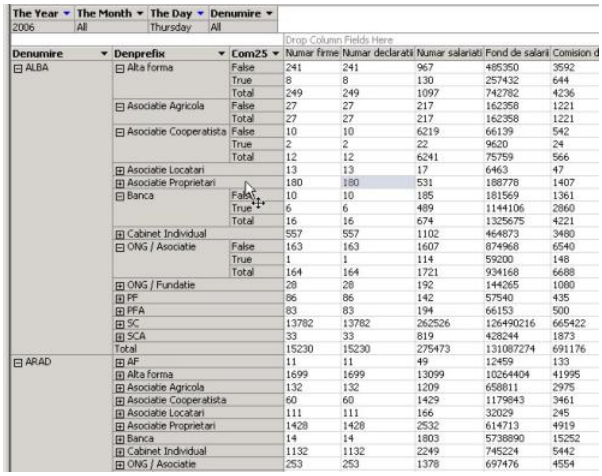


Fig. 6. Exemplification of slice operation by decreasing cube dimensionality to one slice (Cluj county)

Figure 7 outlines a cube rotate operation. One can notice drill-down operation in "county" dimension and the creation of a hierarchy including the legal status and type of commission payable to Area labour Inspectorate.



The Year	The Month	The Day	Denumire
2006	All	Thursday	All

Drop Column Fields Here

Denumire	Denprefix	Com25	Numar firme	Numar declarati	Numar salariati	Fond de salari	Comision d
ALBA	Alta forma	False	241	241	967	485350	3592
		True	8	8	130	257432	644
		Total	249	249	1097	742782	4236
	Asociatie Agricola	False	27	27	217	162358	1221
		Total	27	27	217	162358	1221
	Asociatie Cooperatista	False	10	10	6219	66139	542
		True	2	2	22	9620	24
		Total	12	12	6241	75759	566
	Asociatie Locatari		13	13	17	6463	47
	Asociatie Proprietari		180	180	531	188778	1407
	Banca	False	10	10	185	181569	1361
		True	6	6	489	1144106	2860
		Total	16	16	674	1325675	4221
	Cabinet Individual		557	557	1102	464873	3480
	ONG / Asociatie	False	163	163	1607	874968	6540
		True	1	1	114	59200	148
		Total	164	164	1721	934168	6688
	ONG / Fundatie		28	28	192	144265	1080
	PF		86	86	142	57540	435
	PFA		83	83	194	66153	500
	SC		13782	13782	262526	126490216	665422
	SCA		33	33	819	428244	1873
	Total		15230	15230	275473	131087274	691176
ARAD	AF		11	11	49	12459	133
	Alta forma		1699	1699	13099	10264404	41995
	Asociatie Agricola		132	132	1209	658811	2975
	Asociatie Cooperatista		60	60	1429	1179843	3461
	Asociatie Locatari		111	111	166	32029	245
	Asociatie Proprietari		1428	1428	2532	614713	4919
	Banca		14	14	1803	5738890	15252
	Cabinet Individual		1132	1132	2249	745224	5442
	ONG / Asociatie		253	253	1378	697476	4554

Fig. 7. Exemplification of a cube rotate operation

CONCLUSIONS

This paper focuses on the modes and the working tools to define OLAP cubes which may be helpful at any organizational level, scalable, multi-dimension cubes. Thus, department Data Marts can be created, but OLAP may very well operate directly on Data Warehouse, as working interface.

Therefore we proved that the use of OLAP technology through data management operations brings about an efficient retrieval of information from the data collected into enterprise's information systems.

ACKNOWLEDGEMENTS.The research from this present paper is a part of CNCSIS TE_316 Grant "Metode inteligente pentru fundamentarea deciziei de tranzacţionare pe bursă pe baza informaţiilor publice" (Intelligent methods for decision fundamentation on stock market transactions based on public information), manager V.P. Bresfelean, PhD.

REFERENCES

1. Hawkins, H., Data Warehousing Architecture and Implementation., Editura Prentice Hall, 1998
2. Hyperion Solutions, The Role of the OLAP Server in a Data Warehousing Solution, Hyperion Solutions, 2000
3. Inmon, W. H., Building The Data Warehouse (4th ed), Editura Wiley, 2005
4. Kimball, R., The Datawarehouse Toolkit OLAP Solutions (2nd ed), Wiley, 2002
5. MicroStrategy, The Case for Relational OLAP, MicroStrategy In, 1995
6. Thomsen, E., OLAP Solutions: Building Multidimensional Information Systems, Wiley, 2002