R&S

Responsibility and Sustainability
Socioeconomic, political and legal issues (ISSN: 2340-5813)
Vol 1(1), pp: 53-59 *Case report*
Received: 15 June 2012 / Accepted: 05 July 2012

ONLINE ANALYTICAL PROCESSING (OLAP) CUBES. A CASE STUDY IN LABOUR INSPECTORATE OF ALBA COUNTY (ROMANIA)

Cristian Bologa *(Babes Bolyai University of Cluj Napoca, Romania)* ✉
María P. García *(University of León, Spain)*
Gheorghe Faur *(Babes Bolyai University of Cluj Napoca, Romania)*
Paul Bresfelean *(Babes Bolyai University of Cluj Napoca, Romania)*

Abstract:

In a number of situations public sector information analysis uses to be performed empirically without any concrete 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.

This paper is intended as a case study of data analysis by means of an Online Analytical Processing (OLAP) system with real data from the Area Labour Inspectorate of Alba County (Romania) to prove both the efficiency and the need to create such a system. The aim is to prove the beneficial role of OLAP technology implementation in public institutions, at any level which may involve a managerial act.

Keywords: *data analysis*; *Online Analytical Processing*; *public institutions*; *Labour Inspectorate*

CUBOS DE PROCESAMIENTO ANALÍTICO ONLINE (OLAP). UN ESTUDIO DE CASO EN LA INSPECCIÓN DE TRABAJO DEL CONDADO DE ALBA (RUMANÍA)

Resumen:

En numerosas situaciones el análisis de la información del sector público suele llevarse a cabo empíricamente sin un procedimiento concreto, el cual debería basarse en el uso de modelos o de herramientas de análisis. Al mismo tiempo, lo tocante a la conversión de los datos en información apta para la toma de decisiones en una institución pública no se beneficia de un marco de desarrollo eficiente. Al contrario, la idea de un análisis sistematizado de datos falta totalmente en el caso de muchas instituciones públicas.

Este artículo se plantea como estudio de caso de análisis de datos por medio de un sistema de Procesamiento Analítico Online (OLAP) con datos reales del Área de Inspección de Trabajo del Condado de Alba (Rumanía) para probar tanto la eficiencia como la necesidad de establecer un sistema de tales características. El propósito es probar el papel positivo de la implementación de la tecnología OLAP en las instituciones públicas, a cualquier nivel que pueda implicar un acto de gestión.

Palabras clave: *análisis de datos*; *Procesamiento Analítico Online*; *instituciones públicas*; *Inspección de Trabajo*

✉ Faculty of Economics (Babes Bolyai University of Cluj Napoca), Teodor Mihali 58-60 400591-Cluj Napoca (Romania)
e-mail: *cristian.bologa@econ.ubbcluj.ro*

1. Introduction

In public sector, especially those institutions subordinated to ministries, information analysis is currently performed empirically without any concrete 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 at the level of four types of institutions subordinated to the Romanian Ministry of Labour, Family and Equality of Chances (namely the General Directorate of Labour and Social Protection, the National Unemployment Office, the County State Pension Office, and the 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, Family and Equality of Chances 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. Additionally, none of the institutions subject to our research have ever used an Online Analytical Processing (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.

2. 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. In this sense, although the cube term leads to the idea of three-dimensionality, i.e. the existence of three dimensions, in fact most cubes that can be seen in practice as having from 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.

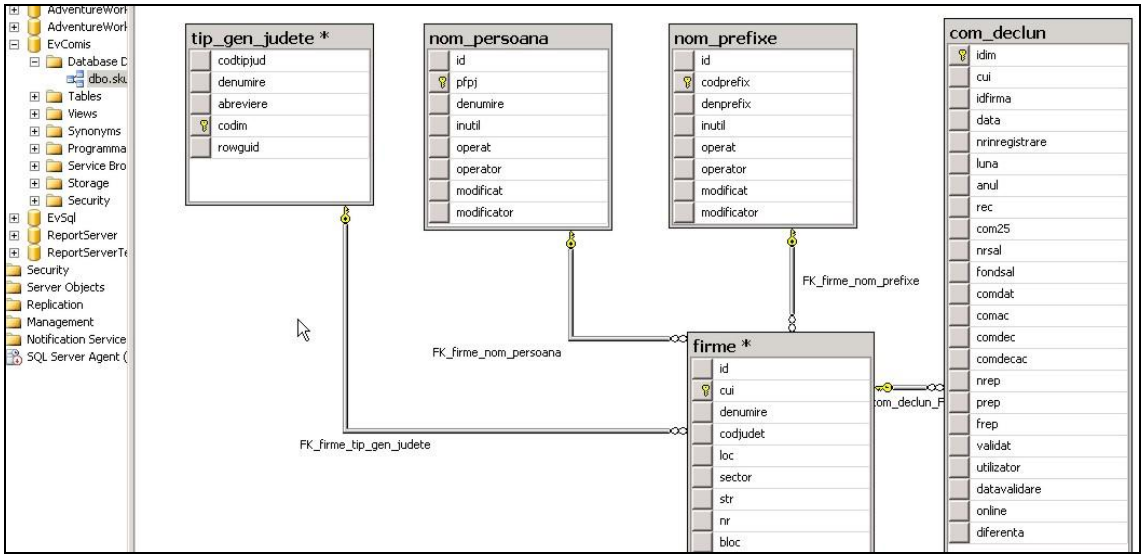
Database operates similarly in all the counties all throughout the country as both the database and fat-client applications managing it are elaborated by the author. Databases are elaborated by means the *SQL Server 2000* Standard edition at the level of each county. These databases are replicated at 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.

Moreover, in companies' table *company's type* (physical person or body corporate), and *organization form* (trading company, public limited company, bank, etc.) are put into relation. 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 the fact table.

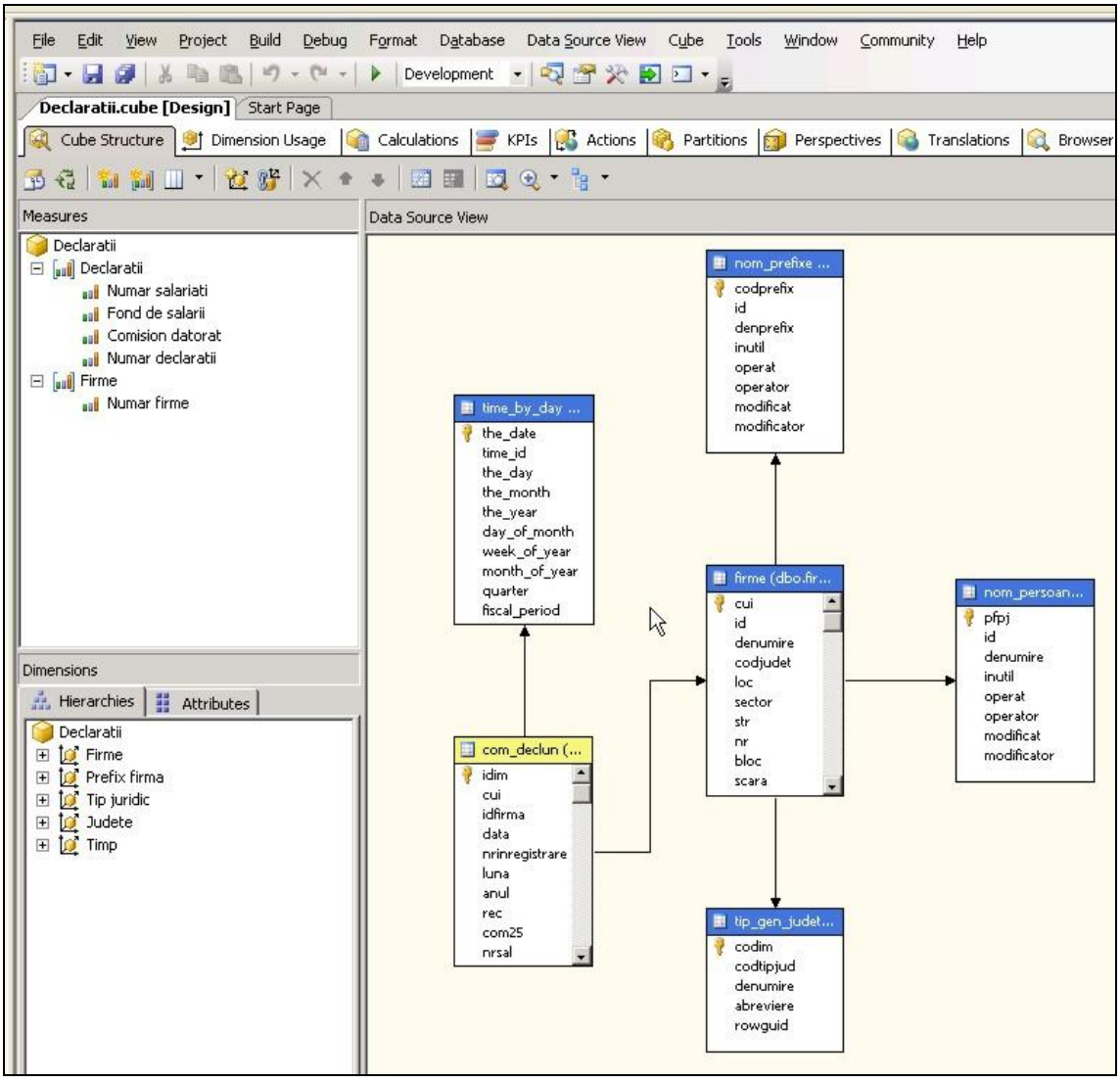
The schema of the cube for analysis is exemplified in Figure 2. One can notice that the fact table with a yellow header and snowflake architecture around *"companies"* dimension, while dimensions are represented with a blue header.

Figure 1. Part of database schema for analysis



Source: own elaboration

Figure 2. Scheme of the cube for analysis with fact table and dimension tables



Source: own elaboration

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.

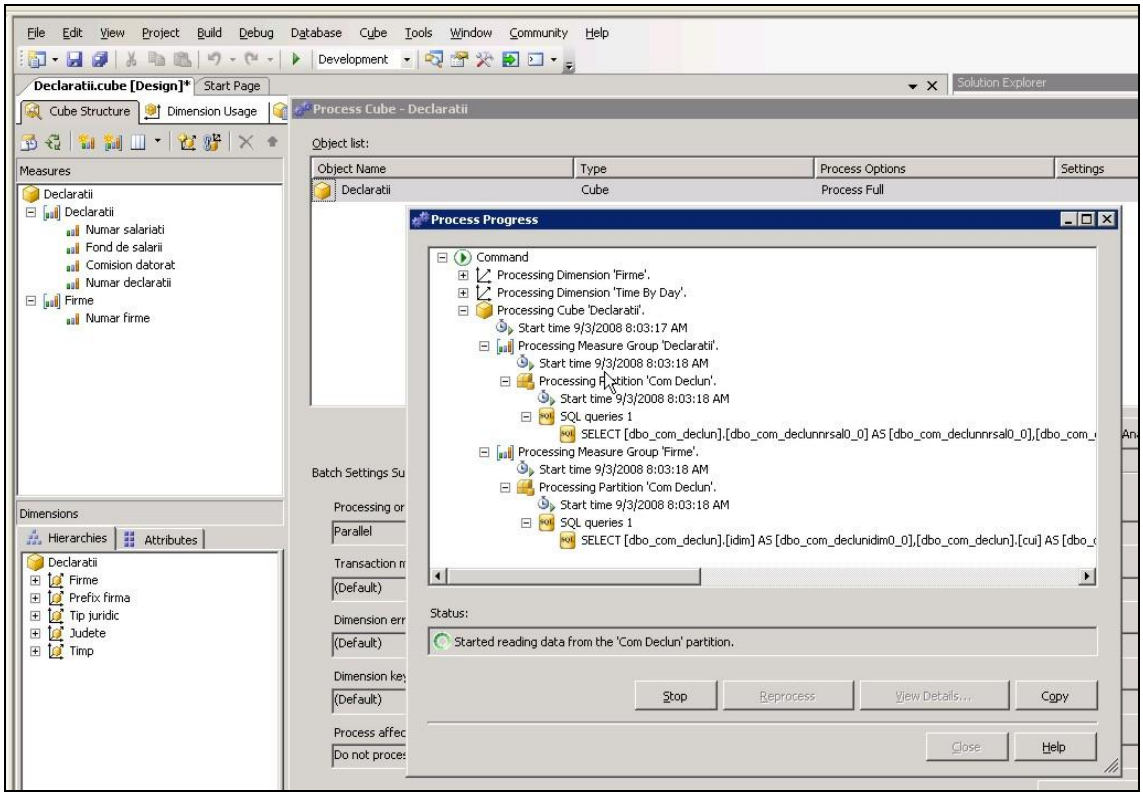
On the one hand, the *measures* established for the cube were the following the number of companies having submitted self-assessment forms, the number of self-assessment forms, the number of employees of each company, the wage funds, and the payable commission.

On the other hand, 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, the company's legal status, the company's type of organization, and the 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 it can be seen in Figure 4. Additionally, 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, and thus selection is made twice.

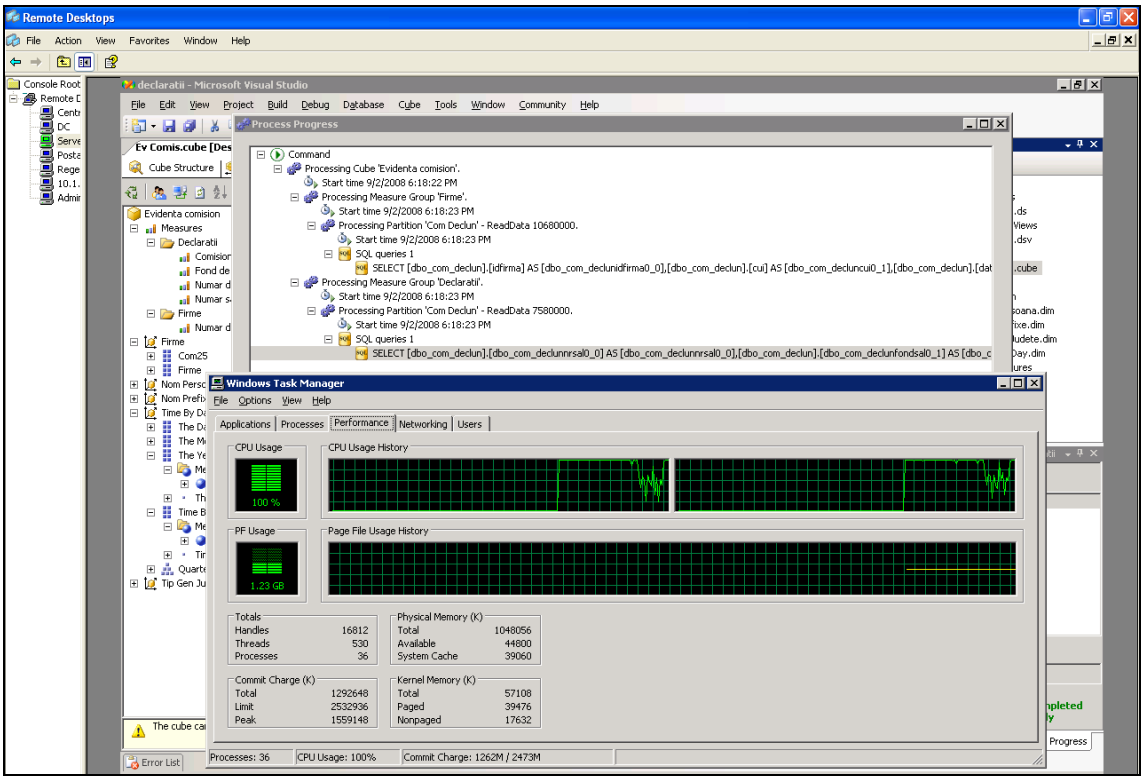
As one can see, processing effort requires huge system resources and therefore the physical machine on which analysis is performed should be an extremely powerful one. However, even when this cube appears as extremely simple, there are much more dimensions to be considered –especially measures– in real 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 3. Cube processing in *SQL Server 2005*



Source: own elaboration

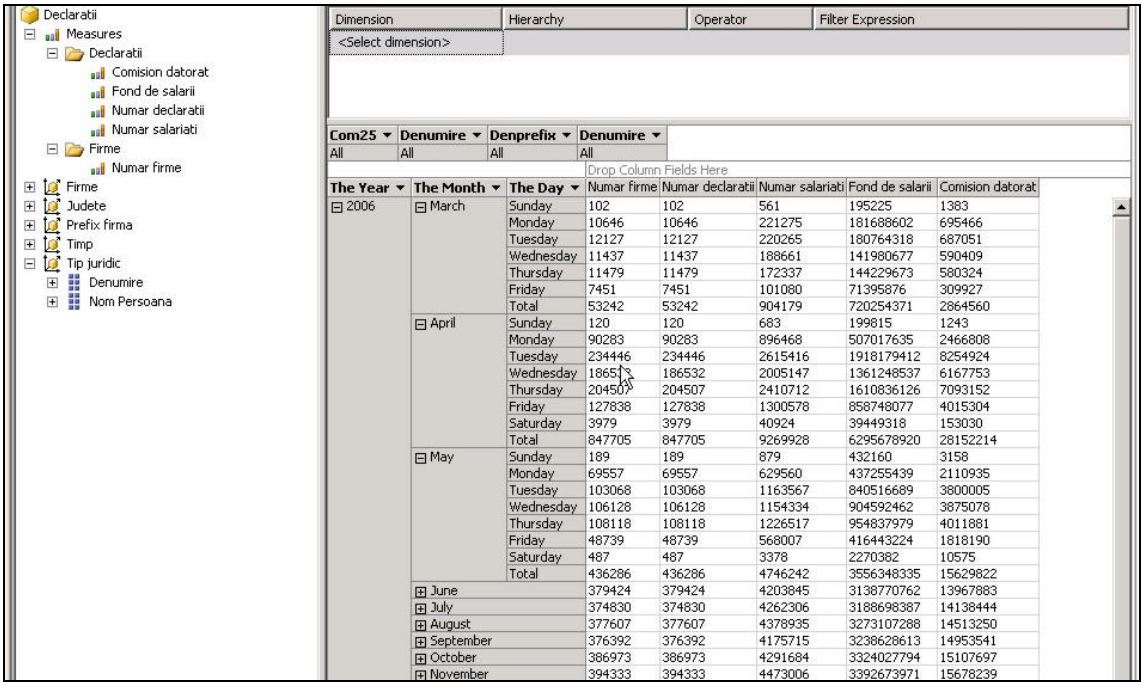
Figure 4. Server effort during cube processing



Source: own elaboration

In this sense, Figure 5 exemplifies the drill-down operation on *"time"* dimension on the processed cube, while Figure 6 exemplifies the slice operation by the dimension *"county"*. Later on, Figure 7 outlines a cube rotate operation, where one can notice a drill-down operation in the *"county"* dimension and the creation of a hierarchy including the legal status and type of commission payable to the Area labour Inspectorate.

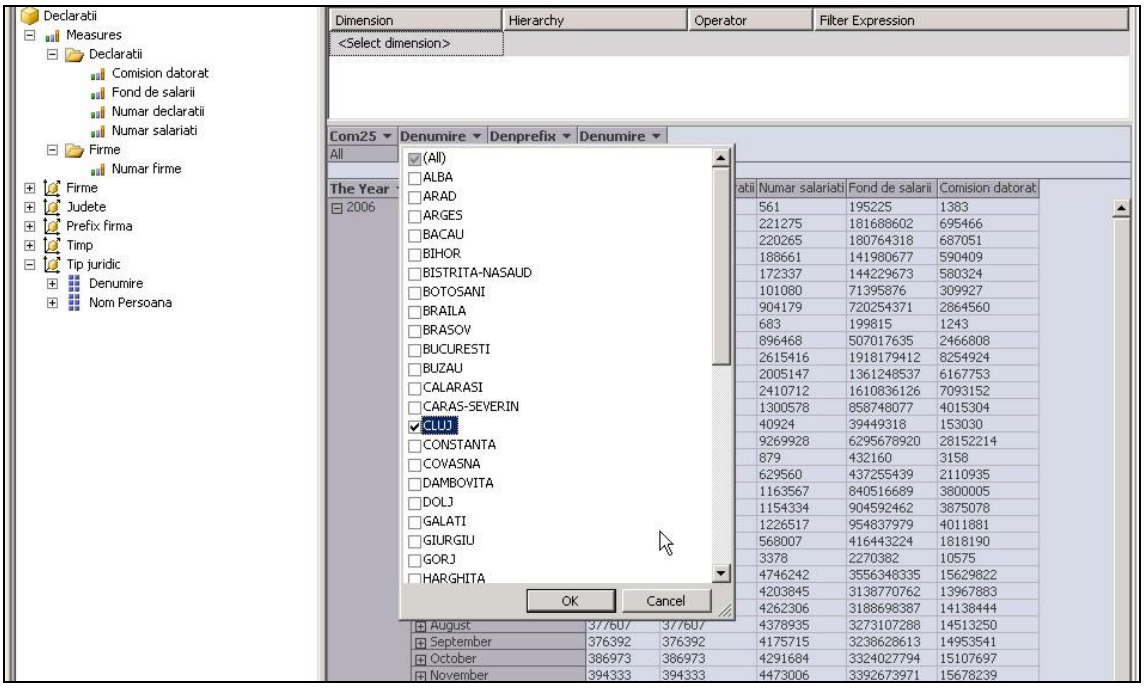
Figure 5. The Cube obtained and a drill-down operation in "time" dimension



The Year	The Month	The Day	Numar firme	Numar declaratii	Numar salariati	Fond de salarii	Comision datorat
2006	March	Sunday	102	102	561	195225	1383
		Monday	10646	10646	221275	181688602	695466
		Tuesday	12127	12127	220265	180764318	687051
		Wednesday	11437	11437	188661	141980677	590409
		Thursday	11479	11479	172337	144229673	580324
		Friday	7451	7451	101080	71395876	309927
		Total	53242	53242	904179	720254371	2864560
	April	Sunday	120	120	683	199815	1243
		Monday	90283	90283	896468	507017635	2466808
		Tuesday	234446	234446	2615416	1918179412	8254924
		Wednesday	186532	186532	2005147	1361248537	6167753
		Thursday	204507	204507	2410712	1610836126	7093152
		Friday	127838	127838	1300578	858748077	4015304
		Saturday	3979	3979	40924	39449318	153030
		Total	847705	847705	9269928	6295678920	28152214
	May	Sunday	189	189	879	432160	3158
		Monday	69557	69557	629560	437255439	2110935
		Tuesday	103068	103068	1163567	840516689	3800005
		Wednesday	106128	106128	1154334	904592462	3875078
		Thursday	108118	108118	1226517	954837979	4011881
		Friday	48739	48739	568007	416443224	1818190
		Saturday	487	487	3378	2270382	10575
		Total	436286	436286	4746242	3556348335	15629822
	June		379424	379424	4203845	3138770762	13967883
	July		374830	374830	4262306	3188698387	14138444
	August		377607	377607	4378935	3273107288	14513250
	September		376392	376392	4175715	3238628613	14953541
	October		386973	386973	4291684	3324027794	15107697
	November		394333	394333	4473006	3392673971	15678239

Source: own elaboration

Figure 6. Exemplification of slice operation by decreasing cube dimensionality to one slice (Cluj County)



Source: own elaboration

Figure 7. Exemplification of a cube rotate operation

The Year	The Month	The Day	Denumire
2006	All	Thursday	All

Drop Column Fields Here

Denumire	Denprefix	Com25	Numar firme	Numar declaratii	Numar salariati	Fond de salarii	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

Source: own elaboration

3. Conclusions

This paper focuses on the ways and the working tools to define OLAP cubes. These are suggested as a powerful tool which may be helpful at any organizational level, as the possibility of scalable, multi-dimension cubes is considered. Thus, and even when department Data Marts can be created, OLAP may very well operate directly on Data Warehouse, as working interface.

Therefore, from the results of a simplified model, 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 results presented in this paper were obtained as part of the 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"), managed by Prof. Dr. P. Bresfelean.

References

Humprhries, M. W., Hawkins, M. W., & Dy, M. C. (1998). *Data warehousing architecture and implementation*. Upper Saddle River NJ: Prentice Hall.

Hyperion Solutions (2000). The Role of the OLAP server in a data warehousing solution. Hyperion Solutions. http://www.oracle.com/technetwork/middleware/bi-foundation/olap-in-a-data-warehousing-solution-128690.pdf/. Accessed 31 May 2012.

Inmon, W. H. (2005). *Building the data warehouse*. 4th ed. New York: John Wiley.

Kimball, R. (2002). *The datawarehouse toolkit OLAP solutions*. 2nd ed. New York: John Wiley.

MicroStrategy (1995). OLAP. MicroStrategy. http://www.microstrategy.es/OLAP/. Accessed 31 May 2012.

Thomsen, E. (2002). *OLAP solutions: building multidimensional information systems*. New York: John Wiley.