A Multidimensional Data Model and OLAP Analysis for Soil Physical Characteristics

CONSTANTA ZOIE RADULESCU
National Institute for Research and Development in Informatics,
8-10 Averescu Avenue, 011455, Bucharest 1, ROMANIA,

MARIUS RADULESCU
Institute of Mathematical Statistics and Applied Mathematics, Casa Academiei Române,
Calea 13 Septembrie nr.13, Bucharest 5, RO-050711, ROMANIA,

VIRGIL VLAD, DUMITRU MARIAN MOTELICA
National Research and Development Institute for Soil Science, Agrochemistry and
Environment Protection, Bd. Marasti 61, Bucharest 1, 011464, ROMANIA,

Abstract: The paper describes the construction of a multidimensional data model intended for the analysis of soil physical properties. The data from the model are provided by two agricultural soil/land databases. Based on the multidimensional data model we build an OLAP cube called AGRISOL_FIZ. The "cube" accepts queries on several dimensions and hierarchies, including land uses, soil types, soil textures, counties, micro-zones, physical geographical units and relief forms. Note that OLAP is a flexible tool that is suitable for complicated analyses of multidimensional data. The analyses are done in a screen-efficient way. We give two examples with OLAP operations on the cube AGRISOL_Fiz.

Key-Words: On-Line Analytical Processing (OLAP), multidimensional data model, OLAP operations, data cube, land use, soil physical properties.

1 Introduction

The recent advances in database technology and data warehouses have proved very useful for the management of natural resources. The data warehouses, the multidimensional databases, OLAP (On Line Analytical Processing**)**, SOLAP (Spatial OLAP) and data mining techniques were successfully applied to the conservation management of the natural resources.

The multidimensional database is a new database concept dedicated to solving the demands of a decision supporting system. To understand what the data is really saying, the managers usually need to investigate data from different perspectives and change the navigation according to the previous observation. Toward this purpose, data from various operational sources are reconciled and stored in a repository database using a multidimensional data model [1]. The data warehouses and multidimensional data analysis use the multidimensional data models.

These models are based on a technique called denormalization. Denormalization use double data in one or several files that are necessary for time minimization or reduction for joints. The use of multidimensional data models allows analysts to navigate easily in data structures and to understand and exploit all the data. It improves the analysts capacity of visualizing abstract queries.

The multidimensional modeling is a conceptual modeling technique used by the OLAP applications. Statistical databases, geographical and temporal databases are strong connected to multidimensional data modeling.

On-Line Analytical Processing (OLAP) is a trend in database technology, based on the multidimensional view of data and it is an indispensable component of the so-called business intelligence technology [3], [4], [11].

Various standardization committees defined their own models [5], [10]. Most of them are data logic models and only a few of them can be considered purely conceptual. Several formal multidimensional models and corresponding query languages were proposed. However each model present a specific outlook for the requirements of multidimensional analysis, a specific terminology and formalism.

At present the land management is a very important problem for the development of a sustainable economy. Sustainable land management rise many problems of increasing complexity and diversity, that can not be solved without computerized tools. The design and realization of databases with physical and chemical indicators of land is crucial for the conservation management of the natural resources. Knowledge of land, particularly of soil physical properties, is very important for various kinds of environmental studies, including crop simulation where the intended users are agronomists, consultants, and analysts from environmental departments, ministries, etc. However, acquiring and tabulating a complete set of land analysis data for a particular region or a country is expensive and laborious. For that reason it is very important to reuse as much and complete as possible the existing land data for as much as possible different problems. For this, more advanced and powerful tools need be used to analyze and take advantage of existing data. This paper describes the construction of a multidimensional data model intended for the analysis of soil physical properties. The data from the model are provided by two agricultural soil/land databases that were realized in the National Research and Development Institute for Soil Science, Agrochemistry and Environment Protection, Bucharest. Based on the multidimensional data model we build an OLAP cube called AGRISOL_Fiz. The cube accepts queries on several dimensions and hierarchies, including land uses, soil types, soil textures, counties, micro-zones, physical geographical units and relief forms.

We give two examples with OLAP operations on the cube AGRISOL_Fiz.

2 The multidimensional data model AGRISOL_Fiz

AGRISOL_Fiz is a multidimensional data model of soil physical properties organized according to land uses, soil types, soil textures, counties, micro-zones, physical geographical units and relief forms.

The data source for the AGRISOL_Fiz multidimensional data model is BDTADTA – the database for advanced decision techniques on agricultural lands. The data of BDTADTA is provided by two agricultural soil/land databases that were realized in the National Research and Development Institute for Soil Science, Agrochemistry and Environment Protection, Bucharest. These land databases are:

- The Romanian soil profiles database (PROFISOL) [12]. This database contains data for characterization Romanian representative land units and its corresponding soil profiles: general data of land units (soil profiles), morphologique data of soil profiles, physical analytical data of soil profiles and chemical analytical data of soil profiles.
- The database of the Romanian soil quality integrated monitoring [8], [2]. This database includes the main categories of land-soil data/indicators as in the soil profiles database PROFISOL and, also, a data set that characterize the soil pollution: the content of heavy metals (Cu, Pb, Zn, Cd, Co, Ni, Mn, Cr), soluble sulphur, soluble flour and residues of organochlorurated insecticides (DDT and HCH).

The model dimensions are:

1. *Land uses* that refers to the land uses. In the frame of this dimension the considered hierarchy is : «Land use category - CFol» -> «land use type - TFol » -> «land use - Fol ». « Land use category » contains two elements {agricultural use, forestry use}.
2. *Counties* contain the set of counties from Romania - (Jud).
3. *Physical-geographical units* contain the physical-geographical units from Romania (UFG). Example: Carpathian Mountains, Transylvania Tableland, Romanian Plain, etc.
4. *Relief forms* contain the main relief forms and relief forms connected in the hierarchy". Main relief forms (FoPRe)" -> Relief forms (FoRe)".
5. *Soil types* contains the set of classes, types and sub-types of soil. The considered hierarchy is " Soil class (CS)" -> "Soil type (TS)" -> "Soil Sub-Type (STS)".
6. *Soil textures* contain the set of soil textures in the soil upper zone. The considered hierarchy is "The soil texture class (CTx1)" -> "Soil texture (Tx1)".
7. *Pedoclimatic Micro-Zones* contains the pedoclimatic micro-zones (MzPC).

Facts are represented by the soil physical characteristics of the agricultural land units (UT). These are:

- NsG1 – Coarse sand in the genetic horizon (layer) Ap or in sect. 0 - 20 cm
- NsF1- Fine sand in Ap or in sect. 0 - 20 cm
- Praf1 - Silt in Ap or in sect. 0 - 20 cm
- Arg1- Clay in Ap or in sect. 0 - 20 cm
- Sch1- Skeleton in Ap or in sect. 0 - 20 cm
- VEU – Edaphical Useful Volume
- DA1 – Bulk density in Ap or in sect. 0 - 20 cm
- PT1 – Total porosity in Ap or in sect. 0 - 20 cm
- KS150 – Saturated Permeability (K Sat.) in sect. 20-150 cm

For the above facts the considered measures are: min, max, avg, dev.

The corresponding schema for the multidimensional data model is a Snowflake schema. It is illustrated in Fig.1.

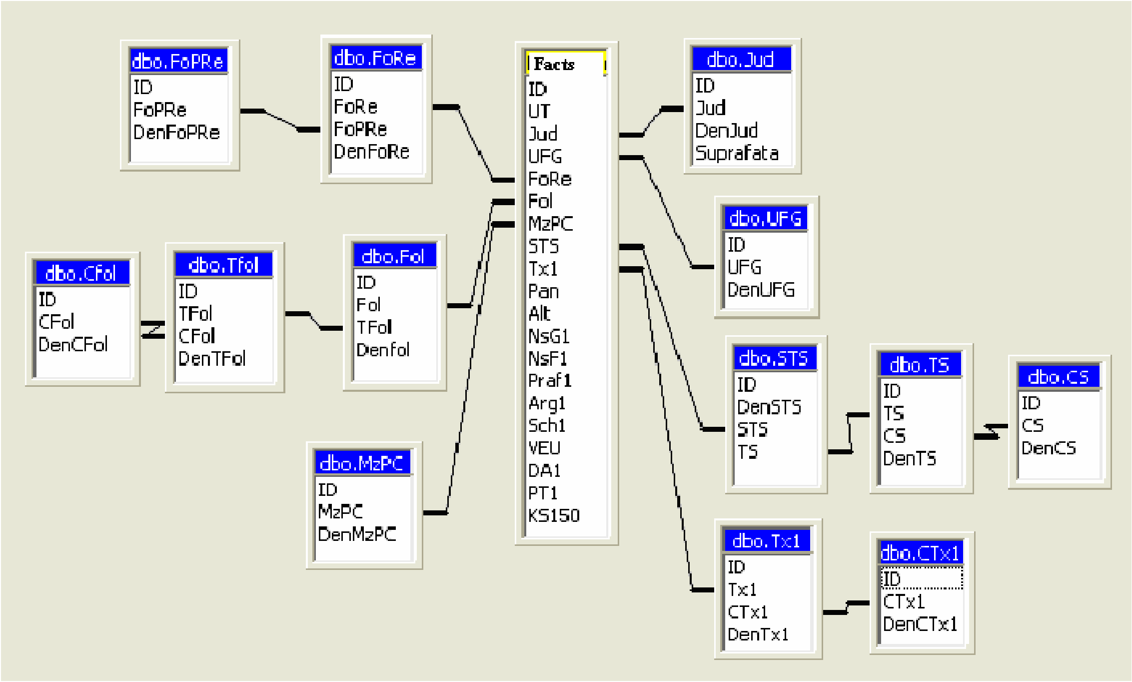


Fig. 1. "Snowflake" structure for the multidimensional data model

The link between dimensions and facts is realized by link codes between the corresponding tables from the "Snowflake" schema. The chosen granulation corresponds to the last hierarchy level.

The cube that implements the multidimensional model was called AGRISOL_Fiz. Data from the data cube could be analyzed by the use of OLAP operations.

Typical OLAP operations for a "cube" are roll-up, drill-down, slice and dice and pivoting.

A typical operation is data aggregation over one or more dimensions. For example: finding the maximum value quantity measured on the Transylvania Tableland, for all type of land use, all type of soil for the Argyle and Skeleton.

The roll-up operation determines data synthetization. This synthetization is realized going from a lower level to a higher level in a hierarchy of a dimension or by the reduction of the number of dimensions.

The drill-down operation is inverse to the roll –up operation. It supposes the decreasing of the level of aggregation or increasing the level of detail. The drill – down operation can be realized by the addition of new dimensions. The roll-up and drill-down operations may be executed over the components of a hierarchy of a dimension. For example if we apply the drill-down operation over a "Land use type" - {Arable} we shall have direct descendents {Common Arable, Pasture, Vegetable Gardens, Rice Plantations}. The slice and dice operations suppose:.

- the partition choice for each dimension of a multidimensional model. This can be performed by queries using "group by" clauses
- cutting from a specific partition along one or several dimensions (corresponding the „where" clause)

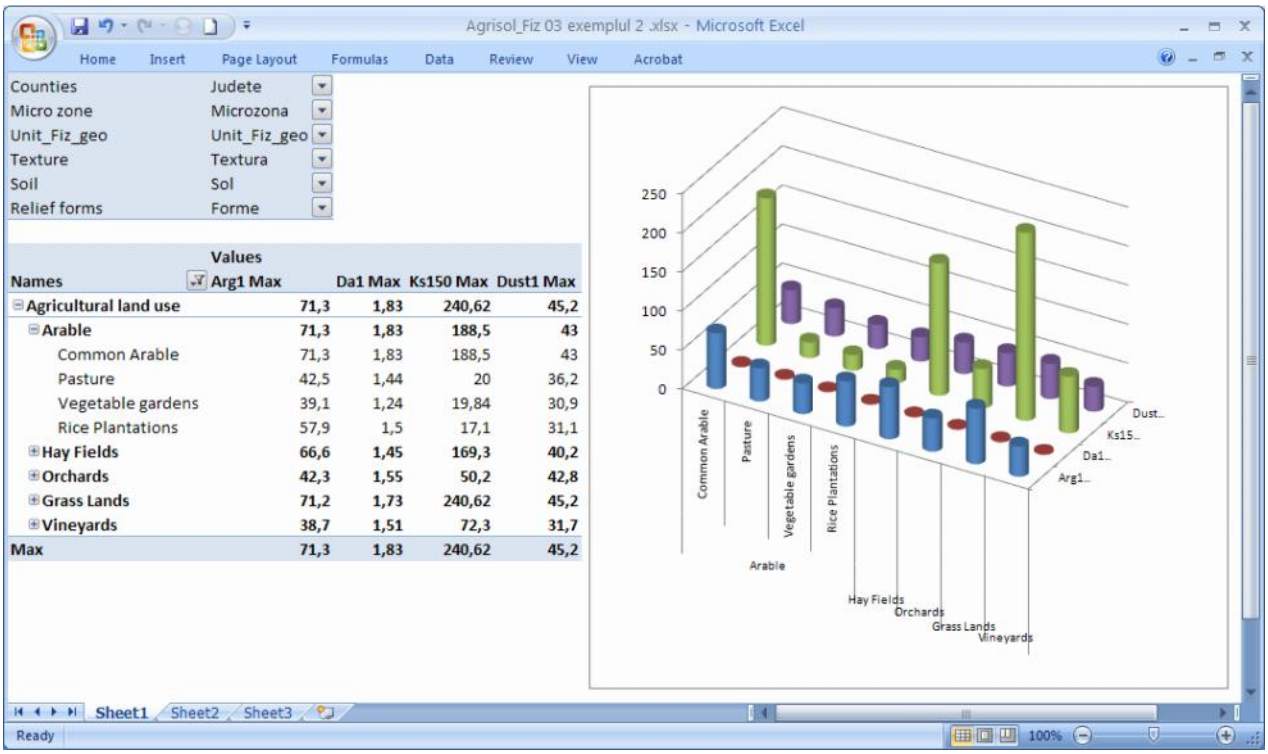
The pivoting operation supposes the reorientation of data cube (3D) for visualization in 2D planes.

In order to analyze the data cube one can use an existing commercial software product or some specific applications designed for this purpose. Examples of commercial software products for cube analysis are the following: Alea Decisionware (MIS AG), BI2M, Microsoft Analysis services, ContourCube, Crystal Analysis, Cube Explorer, Databeacon, Essbase, IntelliBrowser, Microsoft Data Analyzer, MicroStrategy, PowerOLAP, SoftPro Manager 4.0, Microsoft Excel.

OLE DB for OLAP has defined a query language for querying OLAP cubes, called MDX – MultiDimensional Expressions. MDX is similar to SQL in the sense that it follows the "Select... From ... Where ..." framework. But MDX is not an extension of SQL and its syntax is much more complicated than the standard SQL syntax. SQL only deals with two-dimensional data, while MDX allows for querying data with almost any number of dimensions. MDX has defined hundred of functions, which help users specify dimension navigations and calculations [9].

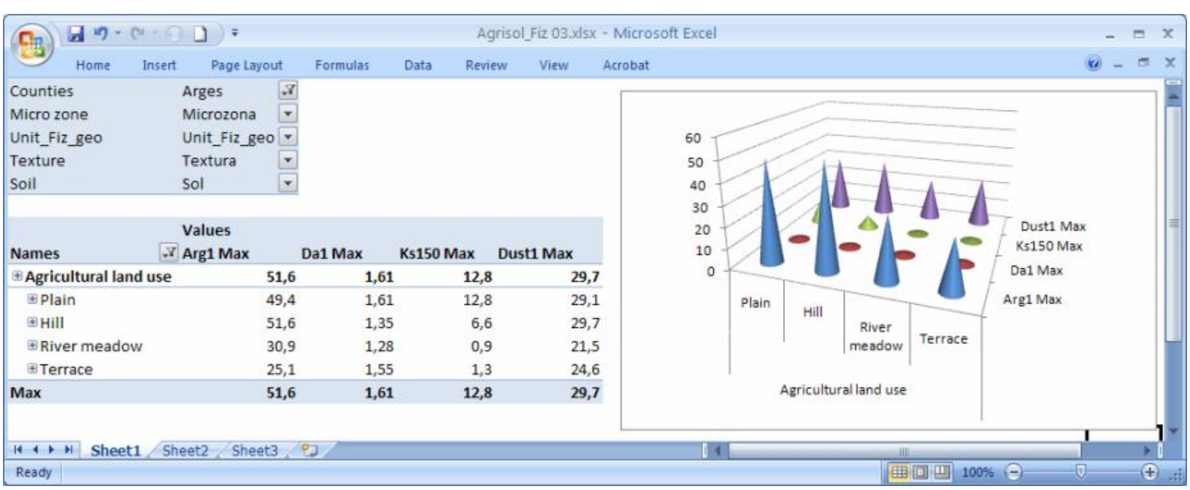
Let D be the set of model dimensions, that is $D = \{d_1, d_2, ..., d_7\}$. The set of all nonempty subsets of

D is: $\mathscr{P}^*(D) =$
$= \{\{d_1\}, \{d_2\}, \ldots \{d_7\}, \{d_1, d_2\}, \ldots, \{d_1, d_2, d_3\}, \ldots, \{d_1, \ldots, d_7\}\}$
Note that $\mathscr{P}^*(D)$ has $2^7 - 1 = 127$ elements. To each element of $\mathscr{P}^*(D)$ corresponds a unique visualization of data at a given level of granularity. Each view can be considered a type of OLAP cube analysis. By selecting various subsets of D one can perform a data analysis with respect to the chosen measures.
To the data cube one can apply the OLAP operations: roll-up, drill-down, slice and dice and pivoting. Any kind of query can be applied to the data cube.
An example of the OLAP drill-down operation applied to the AGRISOL_Fiz cube for the dimension „Land use" is presented in Fig. 2. An example of OLAP „slice and dice" operation for „Arges" county, „Agricultural land use" type and main relief forms is presented in a table form and in a graphical form in Fig. 3. In order to perform this analysis we used Microsoft EXCEL.



Names	Arg1 Max	Da1 Max	Ks150 Max	Dust1 Max
Agricultural land use	71,3	1,83	240,62	45,2
Arable	71,3	1,83	188,5	43
Common Arable	71,3	1,83	188,5	43
Pasture	42,5	1,44	20	36,2
Vegetable gardens	39,1	1,24	19,84	30,9
Rice Plantations	57,9	1,5	17,1	31,1
Hay Fields	66,6	1,45	169,3	40,2
Orchards	42,3	1,55	50,2	42,8
Grass Lands	71,2	1,73	240,62	45,2
Vineyards	38,7	1,51	72,3	31,7
Max	71,3	1,83	240,62	45,2

Fig. 2. OLAP drill-down operation for the dimension „Land use"



Names	Arg1 Max	Da1 Max	Ks150 Max	Dust1 Max
Agricultural land use	51,6	1,61	12,8	29,7
Plain	49,4	1,61	12,8	29,1
Hill	51,6	1,35	6,6	29,7
River meadow	30,9	1,28	0,9	21,5
Terrace	25,1	1,55	1,3	24,6
Max	51,6	1,61	12,8	29,7

Figure 3. OLAP „slice and dice" operation for „Arges" county and main relief forms

More information regarding this OLAP application can be found in [6], [7].
The AGRISOL_Fiz model could be enriched with spatial dimensions, that could contain maps and other additional geographical data that may require SOLAP (spatial OLAP) analyses.

3. Conclusions

The great majority of multidimensional data models are business oriented. The paper presents a multidimensional data model that is oriented towards environment protection and conservation of natural resources. The model has seven dimensions and is intended for the analysis of land physical properties. For the analysis of the model was chosen OLAP since it is appropriate for analyses of multidimensional data. We have performed an analysis with OLAP operations on the cube AGRISOL_FIZ.

Acknowledgements

The work described in this paper was supported by the CEEX – National Research and Development Program of the Ministry of Education and Research – Contract 28/2005.

References:

[1] S. Chaudhuri, U. Dayal, An overview of data warehouse and OLAP technology, *ACM SIGMOD Record* 26(1), 1997, pp. 65–74.

[2] M. Dumitru, C. Ciobanu, D. M. Motelică, E. Dumitru, G. Cojocaru, R. Enache, E. Gamenţ, D. Plaxienco, C. Radnea, S. Cârstea, A. Manea, N. Vrânceanu, I. Calciu, A.M. Mashali, Soil quality monitoring in Romania. ICPA-FAO, Ed. GNP, Bucharest, 2000, 53p (A3) + 25 maps A3, bilingv.

[3] F.G. Filip, *Computer Aided Decision Making: Decisions, Decision-makers and Basic Methods and Tools*, Editura Tehnica and Editura Expert, Bucuresti. (2002). (in Romanian)

[4] J. Han, M. Kamber, *Data mining, Concepts and Techniques*, Elsevier, 2006.

[5] Metadata Coalition, Meta Data Interchange Specification (MDIS version 1.1), http://www.tdwi.org/research/display.aspx?ID=5003, 1997.

[6] C. Z. Rădulescu, D. Enachescu, M. Rădulescu, V. Vlad, D. Veverca, I. Antohe, Set of models, techniques and methods for decision making in sustainable agriculture, Research report, CEEX project 28/2005, ICI, Bucharest, 2006. (in Romanian).

[7] C-Z. Rădulescu, Decision making for sustainable agriculture, *Revista Română de Informatică şi Automatică*, vol 16, nr. 4, 2006, pag. 129- 138; (in Romanian).

[8] C. Răuţă, M. Dumitru, C. Ciobanu, V. Blănaru, St. Cârstea, L. Latiş, D. M. Motelică, R.Lăcătuşu, E.Dumitru, R. Enache. Monitoring of the Romanian soil quality. National Research and Development Institute for Soil Science, Agrochemistry and Environment Protection, Publistar SRL, Bucharest, vol. I and II, 1998, 414 pg. (in Romanian).

[9] Z. Tang and J. MacLennan, *Data Mining with SQL Server 2005*, Wiley, 2005

[10] Transaction Processing Council, TPC: TPC Benchmark, 2007 http://www.tpc.org/default.asp,

[11] E. Turban, J. E. Aronson, T.P.Liang, R. Sharda, *Decision Support and Bussiness Inteligence Systems*, Prentice Hall, 2007

[12] V. Vlad, E. Târhoacă, D. Popa, V. Albu, R. Iancu, M. Băluţă, M. Tapalagă, A. Canarache, I. Munteanu, N. Florea, A. Rîşnoveanu, L. Vlad, M. Nache. Database of soil profiles (PROFISOL) - Structure and functions, Stiinta Solului / *Soil Science*, Bucharest, XXXII, nr.2, 1997, pp. 93-118. (in Romanian).