Data transmission and representation solutions for wind power plants' management systems

ADELA BÂRA, ANDA VELICANU, IULIANA BOTHA, SIMONA VASILICA OPREA
Economic Informatics Department
Academy of Economic Studies
Transelectrica National Power Grid Company
Bucharest
ROMANIA
{Bara.Adela, Anda.Velicanu, Iuliana.Botha}@ie.ase.ro
{simona.oprea}@transelectrica.ro

Abstract: - This paper presents the data level's solutions as the base level of a decision support system that can be used in the National Power Grid. In this paper we considered spatial data, XML and object-relational databases for storing data and representing the locations and the coordinates of the wind power plants installed. We detailed the transmission and storing solutions for GIS. Also, we presented the general architecture of the DSS prototype, which can be used to analyze the current production and also the prediction of the wind energy in the wind power plants locations.

Key-Words: spatial analysis, geographic information system, spatial data, object-relational databases, XML data, prediction, forecast, wind power plants

I. Introduction

Geographic Information Systems (GIS) allow creating maps, integrating information, visualizing scenarios, solving complicated problems and developing effective solutions in a new, easy to use way. A GIS is a system used for modeling information, processes and structures that reflect the real world, including past events, in order to understand, analyze and manage resources and facilities.

To achieve a performing GIS must assure that its functionalities are independently developed and then composed as a whole within a service-oriented architecture. Web services, delivered or not as electronic services, give the possibility to implement a system in which components are weakly-coupled, composite and independent.

An example of using GIS is developing an application for wind parks. Their particularities, like produced energy, noise, altitude where are located, and the external factors that affect the functionality, like wind speed, conflict areas, weather, must be taken into consideration. Thus, such a system will be complex and will need for processing the information spatial analysis as well as data mining tools.

Considering the fact that all business decisions include components of time and space, we can say that GIS represent good solutions in order to improve decision making.

Spatial objects are a combination of non-spatial attributes, which describe characteristics of each item and geo-spatial components, which describe the relative location and geometrical shape of each object.

Nowadays, most of the data used by organizations are stored as object data because this form has the greatest resemblance to the real world and can be easily stored in relational databases with object-oriented extensions.

Organizing spatial data into spatial databases can affect the response speed of GIS systems. An important factor that may have influence on GIS usage refers to the normalization process, which tends to spread data in multiple tables. The result consists in inefficient spatial queries and negative influence for the performance of GIS systems.

The most effective way to store spatial data is given by object-relational databases (ORDB). These are a hybrid type of databases, used to combine the main advantages of relational databases and object-oriented technology [4].

Management systems databases (DBMSs) have implemented support for spatial data type, directly or by extension.

In the following sections, we'll present our solution for developing a decision support system in the national power grid company that will store and represent wind park locations in a GIS, but for advanced analyses will use the object oriented methods.

II. Problem Formulation

II.1 System requirements and constraints

The high level management, as it is in the National Power Grid companies, needs a real time analyses and forecast of the energy produced by the wind power farms. For better manage, transform, process and analyze this information, it is needed a decision support system

(DSS). Many nationally and internationally companies developed various types of frameworks, architectures, solutions and systems to provide economic assistance to decision making processes and production environments with a relatively high degree of certainty.

For environments with low predictability, such as the energy system, that integrates wind resources which depend on natural factors, the traditional forecast applications built in the Enterprise Resource Planning (ERP) solutions is inefficient, being known cases of failure to implement decision support systems [3].

First problem is that data sources provided for the decision-making process must integrate three types of information: external information gathered from the wind parks' production, information within the organization concerning the energy consumption and a combined forecast system both internally and externally. These sources need an efficient integrating framework.

The second problem is the accuracy of the existing forecast systems used to determine the wind park production. There are several studies and software tool developed nationally (studies conducted by research institutes SC ISPE SA, Tractebel Engineering SA, ICEMENERG) and European studies (EWIS study conducted in EU countries, DENNA, made in Germany). However, these studies have shown that the forecast errors propagated in the national energy system can lead to periods of maximum consumption to drop power products in certain regions. Thus, a new type of model based on data mining techniques has to be build that will be able to consider more efficiently the natural factors and also the specific characteristics of the wind power turbines and the environmental constraints.

The third problem is to gather, extract, transform and store the data into a central database from where the business intelligence application can present useful information to the high level management. Also, a data warehouse must be build to store historical data, so we need to implement a solution that will optimize the performance of queries and analyses.

2.2 System's architecture

The system's architecture has to be developed to solve the three major problems described in the previous section.

Thus, for wind park locations and production and also for energy consumption areas the data collected will be stored as spatial data in order to representing them on interactive maps monitoring the activities, and other data like multimedia data and LOB will use specific operations. The data sources will be integrated into a central database. The data will be received in XML format and will be stored in object-relational databases, in order to realize further analysis, necessary for decision making. The section 3 and 4 will present detailed the solutions adopted for this level of the architecture.

For prediction and forecast component, based on our previous research as described in [5], [6] and [7], we'll build a model based on data mining algorithms with a 95% accuracy. The system will use data that can influence wind parks' performances, like meteorological data or data resulted from national energy production regular monitoring. The software component will store predicted data to the central database in a relational form.

The system will use a central data warehouse used by business intelligence tools to present and analyze wind park production and energy consumption, both for historical and forecast periods. An extract, transform and load process need to be developed. In the first phase it will have two components: one for integrate the data sources into the database and the other one for loading the data from the database into the data warehouse.

In the figure 1 is presented the system's architecture.

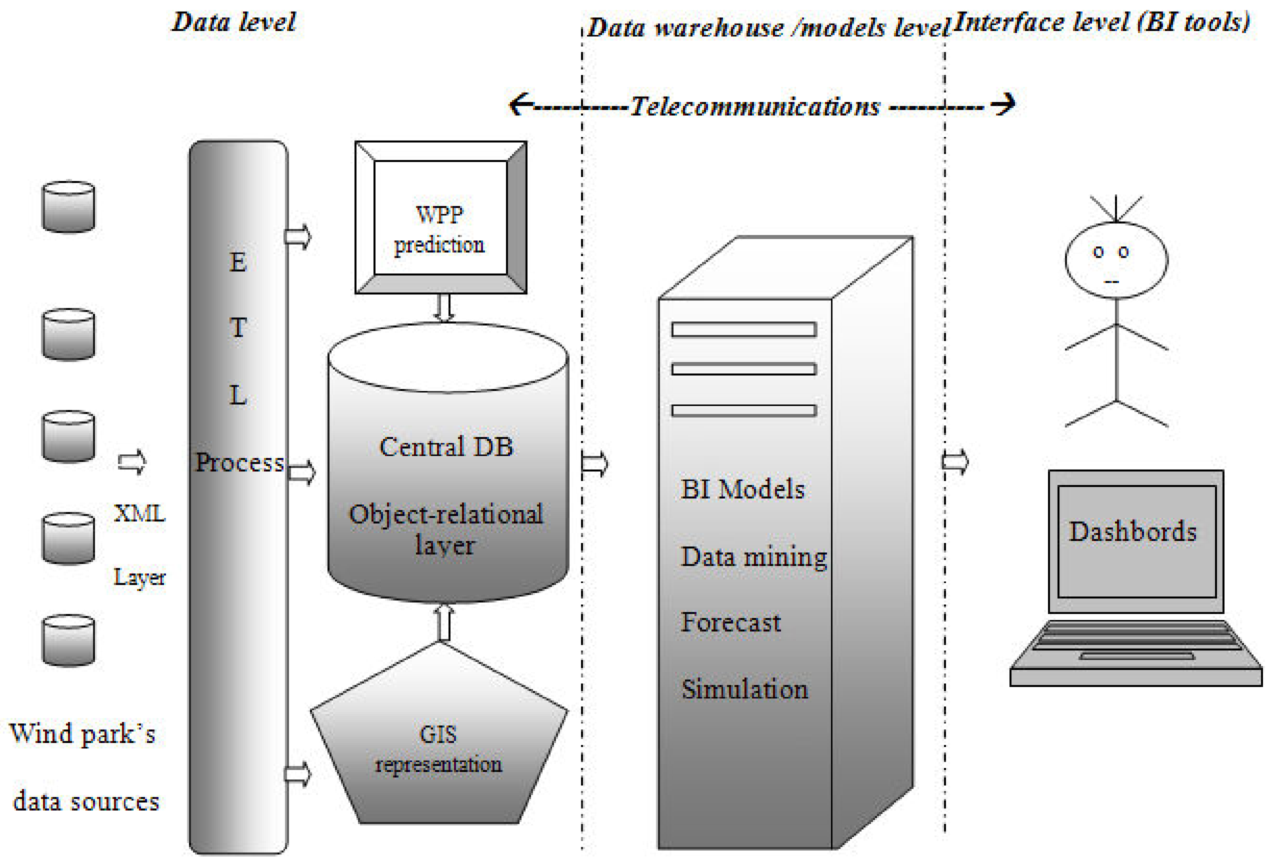

Figure 1 – The system's architecture (Source: Authors)

The components of the architecture are structured on the four traditional levels of a decision support system, namely: 1) data level; 2) system of analytical, mathematical and statistical models; 3) interface; 4) component to ensure communication.

The following sections describe some of the specific components of the data level: spatial analysis with GIS and data representation in an object-relational method.

III. Spatial Analysis with GIS

The main purpose of Geographic Information Systems (GIS) is to perform Spatial Analysis (SA) for geographically referenced data. Simply, we can say that GIS is a digital environment for spatial analysis. The information environment, in which the spatial analysis is made, is the map composed of layers and data. For spatial analysis, the data must be geographically referenced and other GIS related functions must be used: acquisition, editing, validation, storage, primary processing, visualization, posting. Spatial modeling is seen, in this case, as a special spatial analysis, which has spatial scenarios results. Spatial analysis must simultaneously fulfill the following goals: review and interpretation of data, getting extra seemingly hidden information, quantitative and qualitative assessment of the entities, processes and phenomena in the analyzed space, providing practical support for decision making situations. Making spatial analysis involves the use of analytical procedures, combined with: database management, statistical and geostatistical data analysis, image processing and computer mapping elements. Particularly important to the use of GIS and AS there are the knowledge of computer graphics and the specific of information technology in the GIS. Thus, geographic information system is seen as a spatial information science rather than as a technology [1].

According to [2] the basic concepts of spatial analysis are spatial dependence, spatial autocorrelation, and statistical inference for spatial data, stationary and isotropy.

Spatial dependence is a key concept for understanding and analysis of spatial phenomena. This means that certain properties of spatial objects may depend on the same or other properties of spatial objects. In [2] is cited the shadow that a tree can produce around it and how it can affect the growth of other trees nearby. Another example is the discovery of oil spots in a lake and the probability to also discover oil around that stain.

Spatial autocorrelation is a mathematical expression of spatial dependencies. The term comes from the statistical correlation, which is used to measure the dependence between two variables. Autocorrelation implies that the variables compared are the same, but the measurements are made in different places. Spatial autocorrelation verifies how the spatial dependence varies by comparing a sample and its neighbors. In [2] is presented three such formulas:

$$\Gamma(d) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)\xi_{ij} \quad (1)$$

Where: $\Gamma(d)$ is an expression of the link between two random variables as the product of two matrices. Posing a distance d, the w_{ij} matrix measures spatial continuity between z_i and z_j random variables. This link indicates whether or not the two variables are at a distance smaller than d. □ij is a matrix that measures the correlation of these two random variables and it can be their product, as expressed in (2).

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \quad (2)$$

Where: w_{ij} is 1 if the geographical areas associated with z_i and z_j spatially touch, and is 0 otherwise.

Another example is the variogram, where values are formed as squared differences:

$$\hat{\gamma}(d) = \frac{1}{2N(d)}\sum_{i=1}^{N(d)}[z(x_i) - z(x_i + d)]^2 \quad (3)$$

Where: N(d) is the number of variables separated by the d distance.

An important consequence of spatial dependencies is that statistical inference on such data is not as effective as if the independent variables have the same size. In general, this is reflected in higher variances for estimators, lower levels of significance in testing hypothesis and less beneficial adjustment of estimated models. In other words, such a pattern of assumptions usually does not work as properly as when the data compared is the same size and appears independent.

Key statistical concepts that define a spatial data structure are based on primary and secondary order effects. Therefore, the primary order effect is the expected value, which is the average of process in space. Secondary order effect is the covariance between s_i and s_j surfaces. In this context, a process is considered stationary, if the primary and secondary order effects are constant throughout the region studied. This implies that there isn't a trend. A process is isotropic if it's stationary and if the covariance depends only on the distance between the points and not on their direction.

A Z stochastic process is stationary of order 2, if the expected value of Z(u) is constant in all areas studied, meaning it does not depend on its position.

E{Z(u)} = m (4)

Also, the spatial covariance's structure depends only on the relative vector of points difference

h = u-u'

C(h)=E{Z(u) · Z(u+h)}-E{Z(u)}E{Z(u+h)} (5)

Standard GIS systems are very slow in the process of analyzing large volumes of data. It is therefore necessary to explore new methods of analysis that take into account both non-spatial and spatial objects stored in large databases.

IV. Data transmission

Collecting data has long been one of the most expensive procedures in GIS. Now, this data gathering process has become faster, easier and cheaper thanks to the development of data acquisition technologies such as Global Positioning System (GPS) on aerial photos, automatic or semi-automatic scanners and barcode readers. Because of this, geographic information systems have evolved from technologies that use less data and computing power, to technologies that use data-rich environments and high computing power.

In general, a service-oriented architecture can be used for data transmission process and, in the most simplified form, this architecture can be represented as follows:

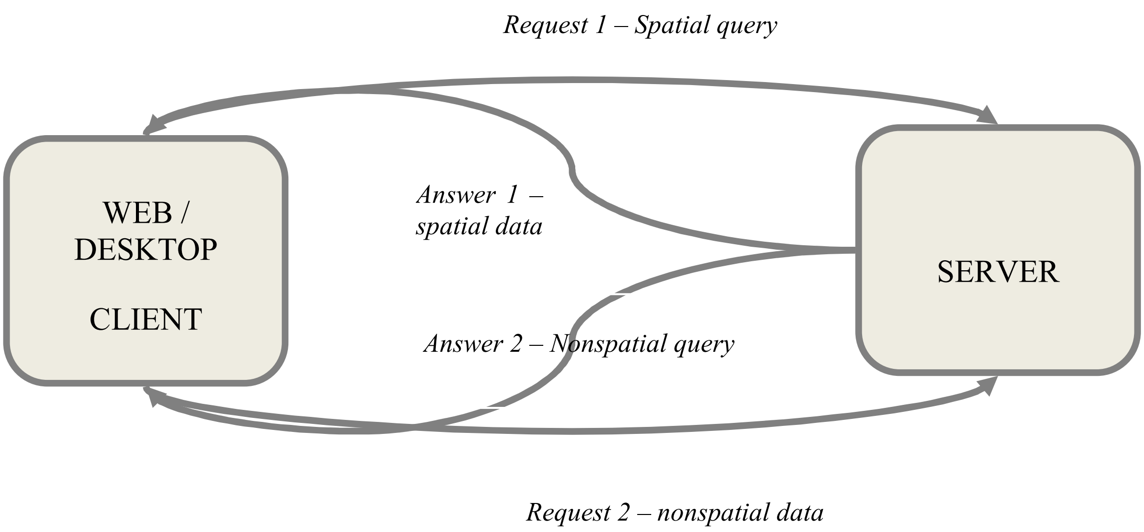

Figure 2 – Simplifed spatial service-oriented architecture (Source: Authors)

Where:

- Request 1 is represented by a spatial query and Answer 1 returns spatial data;
- Request 2 is represented by a common query and Answer 2 returns non-spatial data.

It is noted that requests of the client application may be spatial or not and thus can be encountered two situations described below:

- The client launches a spatial request through a spatial query addressed to the Server. It executes the query and sends the Answer 1;
- The client launches a non-spatial request through a common query addressed to the Server. It performs a function or procedure, working with spatial data, but returns regular data and sends the Answer 2.

If the client requests a regular service, with spatial data already processed as numeric or character data, then the processing of spatial data will be made on the server by using spatial operations in a PL/ SQL block (for Oracle DBMS) or special functions of spatial object classes in programming languages such as Java or C#. For example, in Java there is a JGeometry class which handles spatial data and can be used under *oracle.spatial.geometry* package.

Another spatial service-oriented architecture, more detailed and with examples in Oracle technology, appears in Figure 3. Oracle PL / SQL code is stored as a function or procedure and then transformed into services.

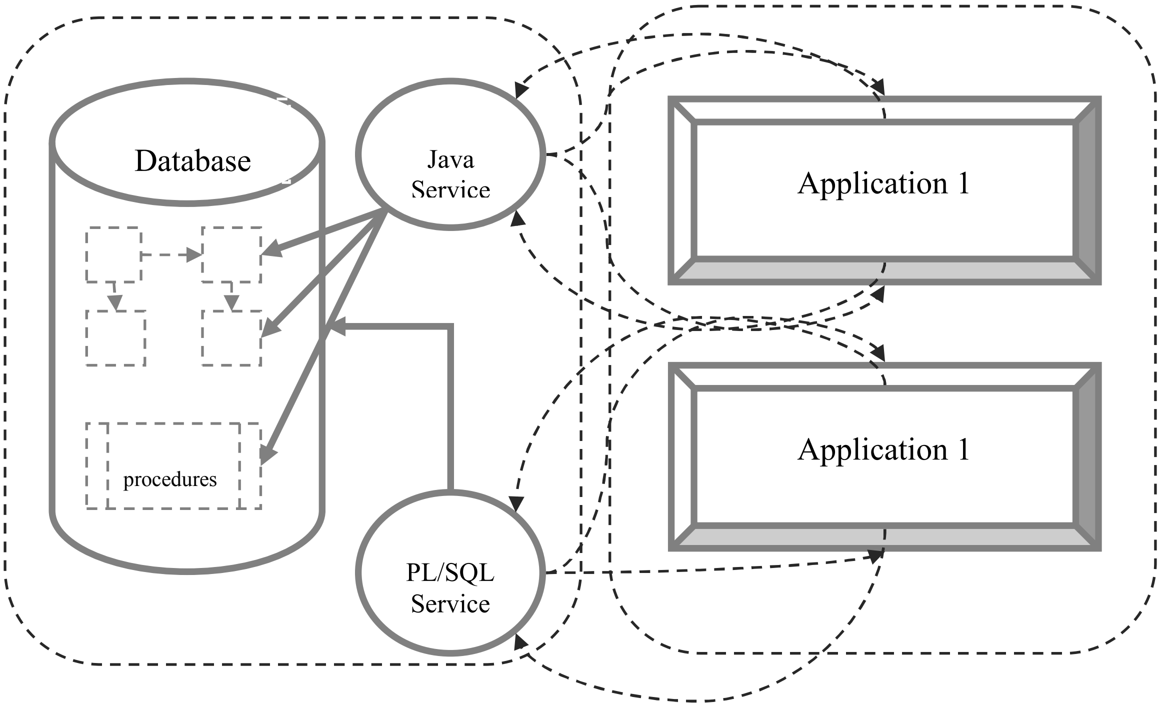

Figure 3: Spatial Service Oriented Architecture exemplified in Oracle

Both services shown in the figure can return spatial and non-spatial data, the Java service may work either directly with the database tables, either with functions or stored procedures, while the PL/SQL service is issued following the conversion of a stored procedure in a Web service.

The data from spatial database tables can be transmitted through web services as it is or processed.

Such processing can be performed either in high level programming languages (e.g. Java, C#), either directly via database stored functions or procedures transmitted through web services as it is or processed.

Example: you want to calculate the minimum distance between a given city and the nearest other city. It can be done as follows:

Method 1:
- In the database located on the server a function/procedure is created;
- The stored procedure returns the name of the found city;
- Based on this procedure a PL/SQL web service is created;
- The Web Service is called by the client.

Method 2:
- On the server a Java function is created to query the database;
- The function returns the name of the found city;
- Based on this procedure a Java web service is created;
- The Web Service is called by the client.

Method 3:
- On the server a Java function is created to call a stored procedure;
- The function returns the name of the found city;
- Based on this procedure a Java web service is created;
- The Web Service is called by the client.

Method 4:
- In the database located on the server a function/procedure is created;
- The stored procedure returns the found city as a geometry;
- Based on this procedure a PL/SQL web service is created;
- The Web Service is called by the client.
- There is a geometry processing at the client application.

Method 5:
- On the server a Java function is created to query the database;
- The function returns the found city as geometry;
- Based on this function a Java web service is created;
- The Web Service is called by the client;
- There is a geometry processing at the client application.

Method 6:
- On the server a Java function is created to call a stored procedure;
- The function returns the found city as geometry;

- Based on this function a Java web service is created;
- The Web Service is called by the client;
- There is a geometry processing at the client application.

Another way to manipulate the spatial data (thorough web services and other ways) is to create virtual tables, especially if you want these tables to store non-spatial data resulting from a spatial query.

For example, you can create a virtual table called *WPP areas* to store the areas of the wind power plants (WPP) that are already stored in a spatial table. Oracle SQL code to create such virtual tables is:

CREATE VIEW WPP_areas AS SELECT SDO GEOM. SDO AREA (geometry, 0.05) FROM
WPP;

Where *geometry* is the SDO GEOMETRY type of WPP, and 0.05 is a parameter that specifies the precision with which approximations are made. The query must have the same value as given in the USER SDO GEOM METADATA metadata table.

The advantage of using virtual tables is that the updates made in the main table (in the above example, the main table is WPP) are reacted in the virtual table (in the above example, WPP areas) so that applications (web services or not), which later on use this table, dont have to work with spatial data, but directly with character or numeric data (as in the above example). Through this way, the processing of spatial data is made by the DBMS that stores the data, client applications only being preoccupied with data representations, which in most cases are in the form of interactive maps.

The services working with spatial data are commonly known as geo-service or geographical services.

Paper [12] shows how geographical services can be coordinated using LCC (Lightweight Coordination Calculus) language. This article also states that the best way to build a geographic system is by using SOA.

The specific characteristics of spatial data sets were analyzed in [13], [14], [15] and they refer to the following aspects:

- Spatial data can be differently represented using different geometric figures in the same or another database;
- Semantic problems may occur by naming di_erent classes or tables with the same name;
- Connections (implicit or otherwise) may arise between spatial objects. For example, one river may cross a lake;
- Spatial data appears in large data sets so storage or transfer issues may occur (e.g. through web services);
- The same set of data may differ significantly depending on the data model, the scale, the geographical contexts generalization or the semantic used by the developers.

All these issues can inuence the way spatial data is transmitted to client applications, mainly due to the large memory space they occupy, their diversity in various aspects and specific tools necessary for their processing.

When using Web services data must not be made public, the client applications working only with data services or operations.

The structure of services or database should not be known by the client applications, which also ensures their independence from the server. On the server, service applications can run on different platforms, different programming languages and data can be stored in different DBMS.

Integrating services into applications can be made using WSDL (Web Service Description Language), which is an XML format for describing Web services as a set of parameters of the messages transmitted via services and containing document or procedure oriented information (usually web pages).

Web services are defined by the following elements: data types, messages, operations, type of ports, connection protocol, the port and service used as a collection of interconnected parameters.

An effective solution for spatial data representation in SOA is to meet the spatial data structure of a specialized DBMS such as Oracle Spatial.

SDO GEOMETRY structure is:

- SDO_GEOMETRY (polygon dimension, latitude longitude, SDO_POINT_TYPE, SDO_ELEM_INFO_ARRAY (SDO_STARTING_OFFSET, SDO_ETYPE, SDO_INTERPRETATION) SDO_ORDINATE_ARRAY (variable number of parameters)).
- SDO_STARTING_OFFSET is the o set value (limit) for storage in the SDO_ORDINATE array and starts from 1and not 0.

The chosen coordinate system will depend on the representation possibilities of the GDBMS, and in Oracle it may be a x0y or latitude - longitude system (of which the most common are ETRS89, WGS84, NAD83). Spatial data type must be uniform in each table, spatial operations working better in a DBMS and depending on its capabilities. Given this solution, it is preferable that the mode of transmission of spatial data through web services to be made through Method 1 or 2, Method 3 being an extension of Method 1 and Methods 4-6, by working with spatial data through services, doesn't provide the major features of a SOA (like weak coupling or the autonomy of services).

V. Object-relational representation

The fast development of the IT&C domain was accompanied constantly by significant changes in economic informatics. The current informatics technologies have created the need for the use and storage of complex object types within databases, while keeping the power of the relational structured query language (SQL). Therefore, the new generation of database systems supports a unified relational and object-oriented data model. Thus, the relational model is extended with the object-oriented concepts such as scalability, complex data types (large objects, multimedia data, XML data, spatial data, user defined object types, etc.) and the fundamental object characteristics like encapsulation, inheritance and polymorphism [10].

An advantage of using applications with object-relational databases, despite the relational ones, leads in the use of a unique data model, recognized in both database and programming language (which is most certainly an object-oriented programming language) [9].

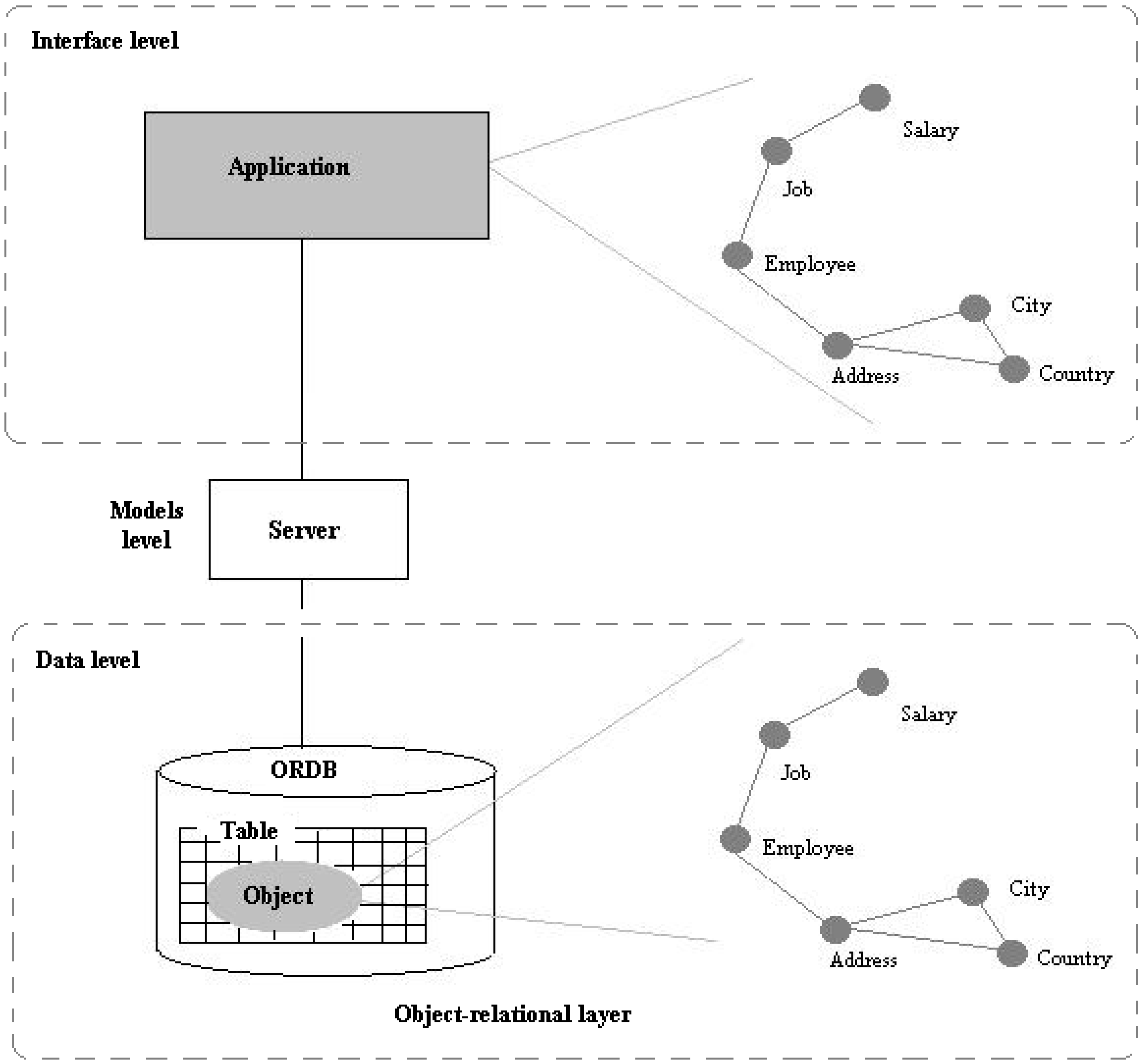

Figure 4 – Using a unique data model in object-relational databases (Source: authors)

As shown in the figure 4, object-relational databases can by-pass the necessity of using mapping techniques between relational database and application programs. However, it is needed for specific techniques in order to map the objects stored in relational tables into instances of classes used in application programs.

The proposed DSS will integrate data from various systems. There are source modules from which are received XML documents with wind park's details and destination modules represented by object-relational databases. XML documents are processed and the data are stored into database.

The advantage of using an object-relational database is that we can get the benefits of both relational and object-oriented technologies, while the disadvantage translates into lower performance due to XML data mapping to the relational data, which can produce a database schema with many relations [11].

When used in object-relational databases, XML data must be mapped into relations. In order to transfer the data between XML documents and object-relational structures are used specific mapping methods. The study

[8] makes a basic classification of these mapping methods, as follows:

Generic methods – that are not using any schema of stored XML documents;

Schema-driven methods - that are based on existing schema of stored XML documents;

User-defined methods - that are based on user-defined mapping.

In order to integrate the XML data about wind parks into the database and realize the mapping, we will use the schema-driven methods. The mapping process will consist in three phases. We will start with the conceptual model, then we will create the XML schema and finally we will map the specific elements on object-relational database schema.

VI. Conclusions and future work

The paper presents the necessities for having real time analyses and forecast of the energy produced by the wind power farms. The high level management of the National Power Grid companies needs solutions like decision support systems (DSS) in order to better manage, transform, process and analyze the information.

The integration of the specific data (like multimedia, XML and spatial data) into object-relational databases is treated as a necessity for today's enterprises that maintain informatics systems in the unpredictable environments.

In the future, our research will continue with the development and implementation of the conceptual model proposed in this paper. For this purpose, will be used specific technologies for each level of the model architecture [11]. Thus, the data level will use solutions for organizing and integrating data, the models level will use solutions of multidimensional analysis, forecasting models, simulation and extrapolation, and for the interface level will be used solutions for data analysis and dynamic presentation of data.

Acknowledgments

This paper presents some results of the research project PN II, TE Program, Code 332: "Informatics Solutions for decision making support in the uncertain and unpredictable environments in order to integrate them within a grid network", financed within the framework of People research program.

This paper is also a result of the project POSDRU/6/1.5/S/11 „Doctoral Program and PhD Students in the education research and innovation triangle". This project is co funded by European Social Fund through The Sectorial Operational Program for Human Resources Development 2007-2013, coordinated by The Bucharest Academy of Economic Studies.

References:

[1] Branişte N.F. - Sisteme informatice geografice. Folosirea eficienta a informaţiilor geografice. In: IF?! Magazine, no 11/ 2005, pg. 9.

[2] Câmara G., Monteiro A. M., Druck Fucks S., Sá Carvalho M. - Spatial Analysis and GIS: A Primer, 2008, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.4115

[3] Bâra A., Lungu I., Velicanu M., Oprea S.V. - Intelligent Systems for Predicting and Analyzing Data in Power Grid Companies, TheProceedings of the IEEE International Conf. on Information Society (i-Society 2010) London, july 2010.

[4] Botha I., Velicanu A., Bâra A. - Integrating Spatial Data with Object-Relational Databases, Journal of Database Systems, nr.1/2011, pp. 33-42, ISSN: 2069 – 3230

[5] Bâra A., Lungu I., Oprea S. V. - Public Institutions' Investments with Data Mining Techniques, Journal WSEAS Transactions on Computers, Volume 8, 2009, ISSN: 1109-2750, http://www.worldses.org/journals/computers/computers-2009.htm

[6] Bâra A., Velicanu A., Lungu I., Botha I. - Natural Factors that Can Affect Wind Parks and Possible Implementation Solutions in a Geographic Information System, International Conference on Development, Energy, Environment, Economics, Puerto de la Cruz, Tenerife, November 30 - December 2, 2010, pg. 50-54, Publisher WSEAS Press, ISBN 978-960-474-253-0.

[7] Bâra A., Velicanu A., Lungu I., Botha I. - Using Geographic Information System for Wind Parks' Software Solutions, WSEAS International Journal of Computers, Issue 2, Volume 5, 2011, ISSN 1998-4308.

[8] Mlynkova I., Pokorny J. - XML in the World of (Object-)Relational Database Systems, Information Systems Development: Advances in Theory, Practice, and Education, Vilnius, Lithuania, 2004, pp. 63 - 76, ISBN 978-0-387-25026-7.

[9] Luthra S. - Architecture In Object Oriented databases, Proceedings of National Conference on Challenges & Opportunities in Information Technology, 2007

[10] Sabau Gh. - Comparison of RDBMS, OODBMS and ORDBMS, The Proceedings of the 8th International Conference on Informatics in Economy, Bucharest, 2007, pp. 792-796, ISBN 978-973-594-921-1

[11] Botha I., Bâra A., Oprea S. - Integrating XML Technology with Object-Relational Databases into Decision Support Systems, Informatica Economica Journal, no. 2/2011, ISSN 1453-1305, accepted for publication.

[12] Vaccari L., Shavaiko P.,Marchese M. - A geo-service semantic integration in spatial data infrastructures, International Journal of Spatial Data Infrastructures Research, vol. 4, pp. 24-51, 2009, ISSN 1725-0463.

[13] Lemmens R., Wytzisk A., Granell R., Gould C., P. van Oosterom - Integrating Semantic and Syntactic Descriptions to Chain Geographic Services, IEEE Internet Computing, 10(5), pp. 42-52, 2006, ISSN 1089-7801.

[14] Friis-Christensen A., Schade S., Peedell S. -Approaches to solve schema heterogeneity at European Level, Proceeding 11th EC-GI & GIS Workshop, ESDI: Setting the Framework, 2005, Alghero, Sardinia, Italia.

[15] Komarkova J, Jedlicka M, Hub M - Usability User Testing of Selected Web-based GIS Applications, WSEAS TRANSACTIONS on COMPUTERS, Volume 9, 2010, ISSN: 1109-2750