

GENERAL ARCHITECTURAL FEATURES OF THE WEB-BASED GEOINFORMATION DATA WAREHOUSE PROTOTYPE

Introduction

The end of 80th years has marked the beginning understanding of value of geographical information systems (GIS) technologies in the public and private organizations and thereof, active universal use of these technologies. About 80 percent of the information collected and used by government agencies, the service companies and many business concerns, used in daily operations, at decision-making and planning is geographically oriented. All this has led that term GIS - became one of most frequently used in modern information technologies. Now GIS is one of the most dynamically developing and profitable areas of the industry of information technologies. Segments of market GIS in 2003 (results of researches of analytical company Daratech, www.daratech.com):

- Software - 64 % from all GIS market (1,175 billion dollars);
- Services - 24 % from all GIS market (447 million dollars);
- Spatial data - 8 % from all GIS market (140 million dollars);
- Equipment - 4 % from all GIS market (70 million dollars).

The logical result on the one hand the common evolution of information technologies and geoinformation technologies and requirements of modern business, and requirements of modern business on the other hand is occurrence Corporate, or often called Enterprise, GIS (Harder 1999; Von Meyer and Oppman 1999) [1]. Considering a condition, inquiries and prospects of evolution of modern business, and also achievement and progress in the field of information technologies, it is possible to allocate the basic problems (and tasks), which decide (or will decide) Corporate or Enterprise Level GISs [0]:

- Problem of avalanche accumulation and integration of SAND - spatial data, and integrated with them problem-oriented databases (nonspatial data) from heterogeneous, sometimes inconsistent sources;
- Problem of standardization of the data;
- Problem of a uniform format of the data;
- Problem of maintenance with the necessary and timely information of separate structural divisions of the enterprise (organization);
- Problem of maintenance data integrity;
- Necessity of decision-making on the basis of the operative and qualitative analysis of the mixed data (SAND);
- Necessity in authorized and encrypted access to the data and the information at any time from any point of the World;
- Necessity implement of multiuser, Geoinformation Decision Support Systems to assist in decision-making;

Complex management of data and information storage.

One of possible ways of the complex (system) decision of these and other problems is the concept of Geoinformation Decision Support Systems Environment on the basis of GIS Data Warehouse (GIS DW), which is at the heart of informational, DSS processing. Thus an ultimate goal of development of the GIS DW should be creation of Enterprise GIS on basis of

Geographical Information System Data Warehouse, which in full measure realizes applicability of GIS DW and will bring maximal benefits [0].

Geoinformation System Data Warehouse (GIS DW) is a *subject-oriented, integrated, nonvolatile, and time-variant collection of structurally related spatial and not spatial data (SAND) in support of management's decisions, providing directly managers and analysts by trustworthy information necessary for operative analysis and decision-making* [0, 2].

Key elements GIS Data Warehouse Architecture

The six positions that generally drive the Data Warehouse Architecture are:

1. Purpose and tasks that GIS Data Warehouse solves;
2. GIS Data Warehouse Architectural Model;
3. Database Management System;
4. GIS Data Warehouse Maintenance Tools;
5. GIS Data Warehouse Access Tools and
6. Budget.

Thus the first set of elements and last factor are determining in a choice of other elements of the Data Warehouse Architecture. One of possible ways of the complex decision of a problem of definition of Geoinformation Decision Support Systems Architecture is carrying out of decomposition of **structure of invariant components of WEB-based Data Warehouse (WEB DW)** [3-5]:

- Data preparation and loading in WEB DW;
- Access, caching and load balancing of WEB/Application Servers of WEB DW;
- Storage, retrieval, update, query and distribution of loading of RDBMS Servers of WEB DW;
- Analytical data processing;
- Maintenance of information safety of WEB DW;
- Management of an infrastructure of WEB DW;
- Backup and archiving.

To questions about realization of "Multi-user GIS System Tier", "Statistical and analytical (OLAP and Data Mining) nonspatial data processing", "Tier of internal and external SAND sources", "Data storing and backup Tier" (fig. 1) it is devoted huge quantity of works [0], we shall consider more in detail "WEB-GIS" (or GIS-Internet) Tier (fig.1).

One of possible variants of the decision of "WEB-GIS" Tier (fig. 1-2) is use of [0]:

- XHTML/XML/Scripting (JavaScript and/or ECMAScript) technologies and AJAX [6, 7] for visualization and management of spatial and nonspatial data and information on the client side;
- Multilevel architecture of application servers on the basis of Java2 Enterprise Edition [8-10] or Microsoft .NET technologies - as the base middle ware technologies;
- RDBMS for storage and management of the mixed information [11-12].

Within the framework of the specified architecture can be used a database-centric approach to J2EE (or .NET) application development. Such model allows to use with maximal benefit as opportunity of ORDBMS Oracle (first of all Oracle Spatial, provides a SQL schema and functions that facilitate the storage, retrieval, update, and query of collections of spatial features in an Oracle database and also operators and functions for performing area-of-interest queries, spatial join queries, and other spatial analysis operations [11], and also practically standard for Enterprise Level RDBMSs opportunities of storage, retrieval, update, query, statistical and analytical processing of a plenty volumes of nonspatial data [12], that finally has allowed to

realize the Kernel of GIS DW Prototype), and advantage of use of application servers that make use of J2EE architecture (multi-user and multi-application scalable framework for end-user which use various tools for access to data/information and services of the GIS DW on the basis of thin client architecture) [0].

Such approach was used by development of client applications of the GIS DW Prototype (fig. 2-3) [0].

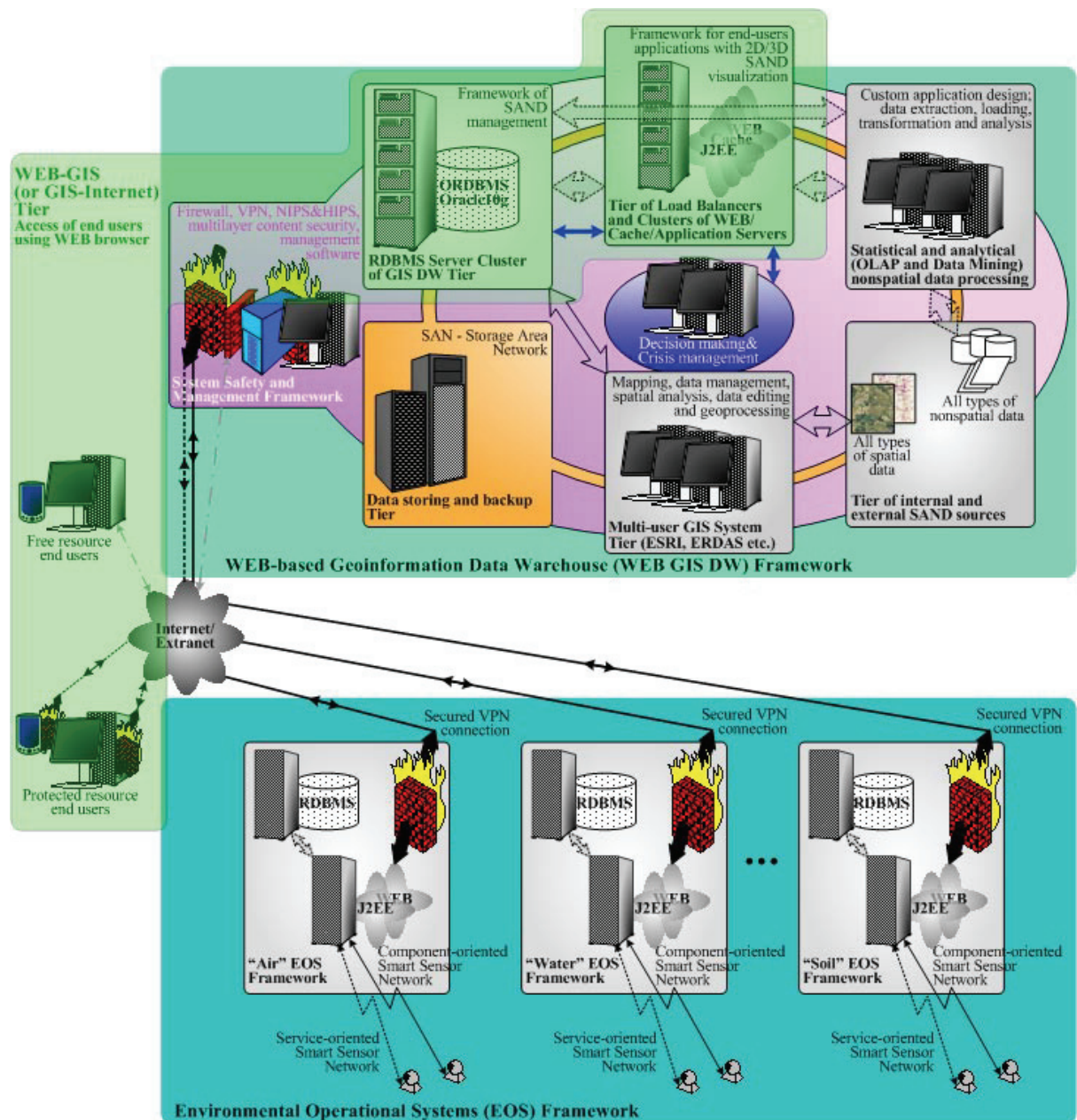


Figure 1: Concept of multi-tier Environmental Risk Management (ERM) Framework

The approach, at which WEB GIS DW is used as a core of the multi-user multi-purpose corporate geoinformation systems (Multipurpose Enterprise GIS [1]) allows successfully solve not only problems of integration of a data of various types and uses of various methods of the analysis of the information (conventional methods of analytical data processing traditional for Data Warehousing: OLAP and Data Mining, and various methods of the analysis and modeling of the spatial information: classical GIS toolkit) but also implementation of authorized,

cryptographic, access to resources and services of system [13-15] at any time from any point of the World that move out systems constructed by such principle on essentially new level [0].

Use of global networks by transfer of the information containing confidential data entails necessity of construction and use of effective system of information protection. In the first works about information protection the basic postulates which are actual and today have been stated: it is impossible to create absolute protection; the system of protection of the information should be complex; the system of protection of the information should be flexible and adapted to changing conditions [5].

Potential vulnerability of WEB-based Data Warehouses in relation to casual and premeditated negative influences has put forward problems of information safety in the category of the major, strategic, determining a basic opportunity and efficiency of utilizing of Data Warehouse. Requirements on a safety of WEB DW can differ essentially, however they are always directed on achievement of three basic properties: integrity; availability and confidentiality. Depending on character of the information contained in WEB DW, costs for maintenance of reliability and safety can make from 5-20 % up to 100-400 % from the resources used on WEB DW functional tasks, that is in special cases (critical military systems) can exceed the last in 2-4 times [5].

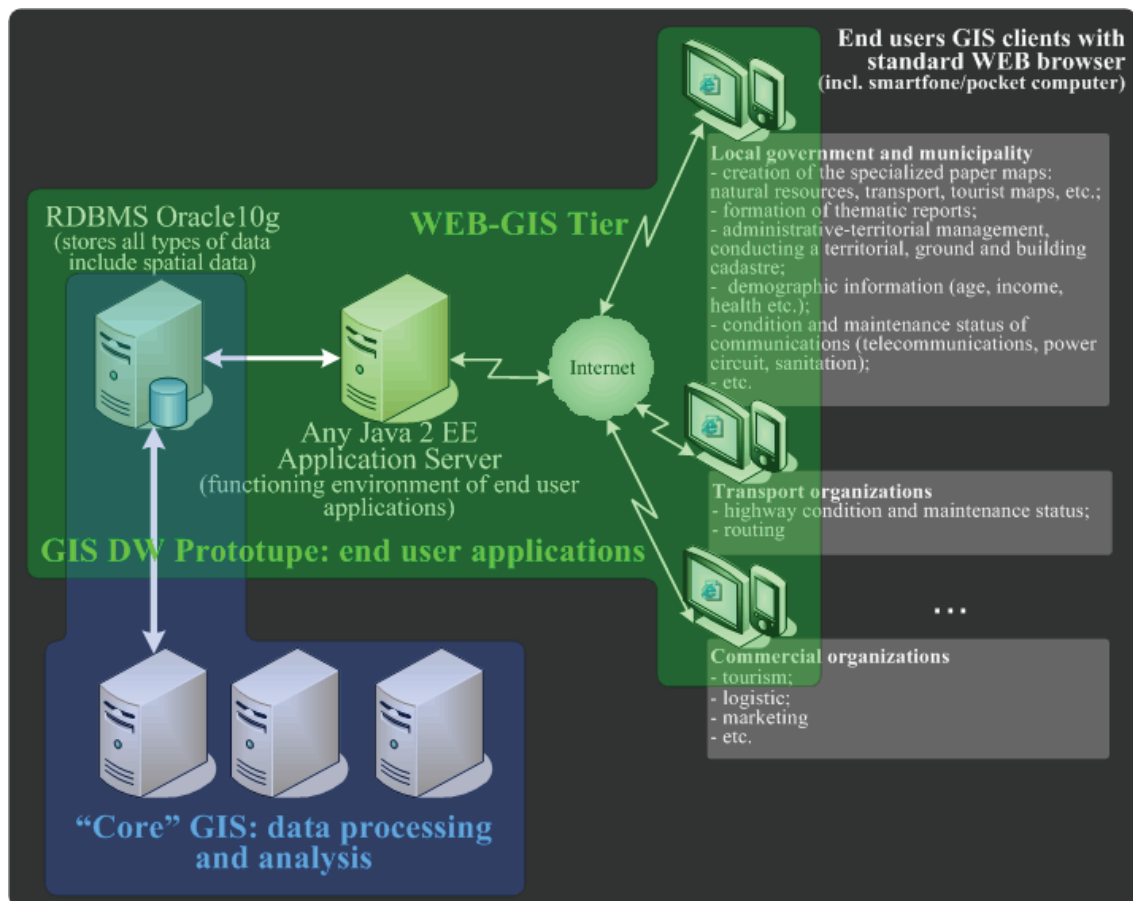


Figure 2: Concept of GIS DW Prototype

Besides now steady positive growing dynamic of a number of cybercrime incidents from inside a corporate network (about 80 % from total of cybercrime incidents [13]) is observed.

Using a principle of systemity, and also fact, that effective protection system should be complex and adapted to changing conditions [13], the **Base Concept of the Decision Support Systems Environment on the Basis of the GIS Data Warehouse** should provide presence of a of

multilevel subsystem of information protection on the basis of use of the modular scaled architecture ("System Safety and Management Framework", fig.2) based (in common case) on [13-16]:

1. The **base level** of GIS DW safety, provided with an external router.
2. The **high level of GIS DW safety** based on the concept of the multilevel demilitarized zone (DMZ) and provided by: a primary firewall (or firewall pool); additional advanced safety services in DMZ; advanced mechanisms of authorization and authentication of users and encryption of the traffic between remote Internet/Extranet users and services of GIS DW(client-to-site VPN); IDS/IPS systems of network traffic monitoring for detection and prevention of intrusions in DMZ; means and services of a safety at level of operation systems (OS), RDBMS and WEB/Application servers, and in case of especially critical GIS DW systems - the equipment on revealing technical channels of information leakage and radio monitoring and other services of safety.
3. The **maximum level** of safety of the protected segment of GIS DW. Because the protected segment of GIS DW contains the most valuable information of GIS DW, one of methods of increase efficiency of a subsystem of safety of protected segment is use essentially other products with elements of the intellectual analysis and network traffic filtering, and also advanced mechanisms of an establishment of authenticity of users and of the traffic encryption between internal users (as a part of separate structural divisions of the enterprise) and services of GIS DW (site-to-site VPN).

Basic functionalities of the WEB-based Geoinformation Data Warehouse Prototype

For acknowledgement in practice, and also for research of some positions resulted above, has been developed the WEB-based Geoinformation Data Warehouse Prototype (GIS DW Prototype) [0, 2].

Now GIS DW Prototype possesses the following capabilities (examples of GIS DW Prototype end-user interfaces are represented on fig. 3) [0, 2]:

1. Use of a standard WEB-browser for user applications.
2. Accumulation, storage and search of the mixed information - spatial and nonspatial data (SAND) in a uniform database (recommended and current ORDBMS: Oracle10g, but it is possible adopt for use on the base of noncommercial RDBMS, for example MySQL).
3. Storage of the spatial data in ORDBMS Oracle10g: object-based, two-dimensional vector model (supported types of the spatial data: polygon, multipolygon, line, multiline, point, etc.), geodetic coordinate system; current (loaded in ORDBMS) format of spatial data: ESRI Shapefile; opportunity transition to storage / representation / manipulations / processing of the 3D measured spatial data.
4. High-quality multi-color 2D vector visualization of SAND with capabilities of zoom and pan and management of SAND layers on client and server levels.
5. Realization of basic operations of the spatial analysis (calculation of the area, centroid, perimeter, length; association, subtraction, interaction; topological operations: "within a distance"; "overlap", " ", "touched" etc.); Spatial Clustering.
6. Dynamic search of the shortest routes ("routing") on the basis of network model.
7. Real Time (more exact quasi-real time) SAND Subsystem (interval of SAND auto updating: from 5sec).

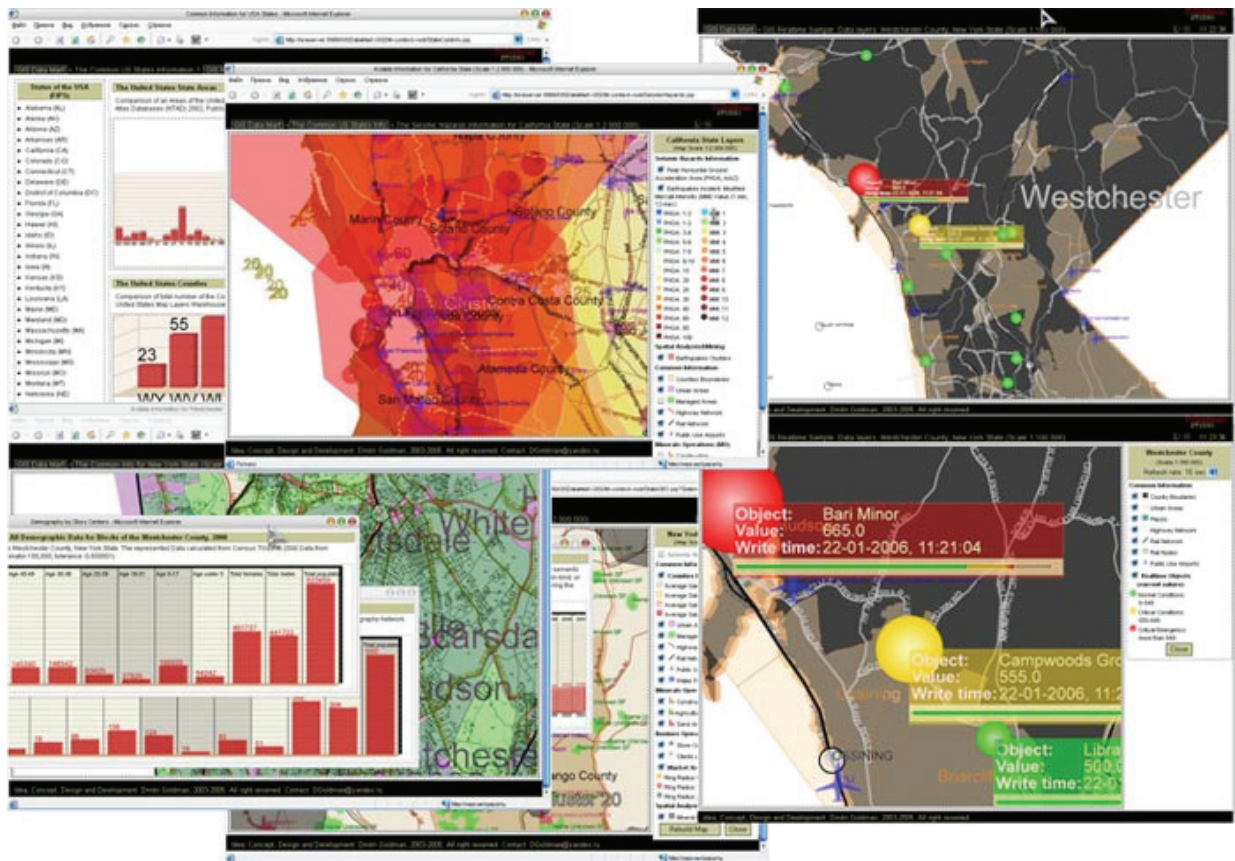


Figure 1: Concept of multi-tier Environmental Risk Management (ERM) Framework

GIS DW Prototype includes more than 40 thematic layers of SAND - information about USA Counties:

- Counties and States boundaries (scale 1:100 000 and 1:2 000 000);
- Census Block Group Data (scale 1:100 000);
- Demography;
- Estimates of total personal income, by county (in 1969-2001);
- Water features (scale 1:100 000 and 1:2 000 000);
- Airports (scale 1:100 000 and 1:2 000 000);
- Significant, historic earthquakes (scale 1:2 000 000);
- Ecosystems of regional extent;
- Seismic hazard in the United States. The data represent a model showing the probability that ground motion will reach a certain level (scale 1:2 000 000);
- Major roads (scale 1:100 000 and 1:2 000 000);
- All roads - streets (scale 1:100 000);
- All roads (1:2 000 000);
- Railroads (scale 1:100 000 and 1:2 000 000);
- Rail node (scale 1:100 000);
- Mineral operations (agricultural, construction, crushed stone operations, ferrous metals processing plants, ferrous metal mines, refractory, abrasive, and other Industrial operations: scale 1:2 000 000);
- Urban boundaries (scale 1:100 000 and 1:2 000 000);
- Cities (scale 1:2 000 000) and other

obtained from **data sources**:

- USDA Forest Service;
- National Weather Service;
- U.S. Geological Survey;

- USDOT/BTS;
- Environmental Systems Research Institute (ESRI) and other.

Demos of WEB-based Geoinformation Data Warehouse Prototype may be obtained from Downloads section.

Conclusion

The offered decision takes into account modern tendencies both in GIS and Data Warehouse areas, and also made high demands of qualifying standards of a hardware-software complex of information systems, based on the concept of WEB DW (scalability, mobility, portability, reliability, safety, modularity, functional completeness) [0, 3-5, 13-21].

Developed GIS DW Prototype (in view of necessity revision and expansion of functionality) can be used as a kernel for construction of the multi-user multi-purpose information systems (for example GIS-Internet systems, WEB-based quasi-real time control and management systems), and as additions to existing Enterprise GIS decisions and products.

Taking into account an opportunity of use of GIS DW Prototype for creation of real systems on the basis of noncommercial platforms of RDBMS and J2EE Application Servers, the practical importance of GIS DW Prototype concludes in an opportunity of creation of information systems of a various level of complexity for SME business [0].

References

- [0] **Goldman, D.** (2006). *WEB-based Geoinformation Data Warehouse Prototype in the Concept of Multi-purpose, Multiuser GIS Enterprise*. Vestnik of YANKA KUPALA STATE UNIVERSITY OF GRODNO (YKSYG) (accepted, in Russian).
- [1] **Harmon, John E. & Anderson, Steven J.** (2003). *The Design and Implementation of Geographic Information Systems*. John Wiley & Sons, Inc.
- [2] **Goldman, D.** (2005). *WEB-based Geoinformation Data Warehouse*. In Collection "FKS - XIII". Grodno, 289-291 (In Russian, ISBN 985-417-478-6).
- [3] **Goldman, D.** (2004). *The System Approach in Decomposition of Invariant Components of WEB-based Data Warehouse*. In Collection "New mathematical methods and computer technologies in designing, manufacture and scientific researches". Gomel, 259-260 (In Russian, ISBN 985-439-103-5).
- [4] **Voltchok, V. & Goldman, D.** (2004). *The Structure of Invariant Components in Integrated Model of WEB-based Data Warehouse*. In Collection "Soft Computing and Measurements" (SCM'2004). Russian Academy of Sciences. St. Petersburg (In Russian).
- [5] **Voltchok, V. & Goldman, D.** (2004). *Modeling of Invariant Components of WEB-based Data Warehouse*. State Organization "Belarusian Institute of System Analysis and Information Support for Scientific and Technical Sphere", 65 pages, part n.: D200497 (In Russian).
- [6] **Asleson, Ryan & T. Schutta, Nathaniel.** (2006). *Foundations of Ajax*. Apress.
- [7] **Crane, Dave & Pascarello, Eric with James, Darren.** (2006). *Ajax in Action*. Manning Publications Co.
- [8] **Johnson, Rod.** (2003). *Expert One-on-One J2EE Design and Development*. Wrox Press. Wiley Publishing, Inc.
- [9] **Ashmore, Derek.** (2004). *The J2EE Architect's Handbook: How to be a Successful Technical Architect for J2EE Applications*. DVT Press.
- [10] **Kuchana, Partha.** (2004). *Software Architecture Design Patterns in Java*. CRC Press LLC.

- [11] **Murray, Chuck.** (2003). *Oracle® Spatial User's Guide and Reference 10g Release 1 (10.1)*. Oracle Database Documentation Library 10g Release 1 (10.1). Part Number B10826-01.
- [12] **Kyte, Thomas.** (2001). *Expert One on One*: Oracle. Wrox Press Inc.
- [13] **Voltchok, V.; Goldman, D.; Kuzmich, M.** (2005). *The Model of Commercial Information Protection and Intrusions Monitoring of the Enterprise-Scale Data Warehouse*. The collection of the proceedings "Nowoczesne technologie informacyjne i ich wpływ na funkcjonowanie podmiotów gospodarczych". Białystok: The University of Finance and Management in Białystok, 171-187.
- [14] **Voltchok, V.; Goldman, D.; Kuzmich, M.** (2005). *Model of Information Protection and Monitoring of Intrusions into WEB-based Data Warehouse*. Vestnik of YANKA KUPALA STATE UNIVERSITY OF GRODNO (YKSYG), 1(35), 113-127 (In Russian).
- [15] **Voltchok, V.; Goldman, D.; Kuzmich, M.** (2004). *Decomposition of Invariant Components in the Model of Information Protection and Monitoring of Intrusions into WEB-based Data Warehouse*. State Organization "Belarusian Institute of System Analysis and Information Support for Scientific and Technical Sphere", 22 pages, part n.: D200497 (In Russian).
- [16] **Voltchok, V.; Goldman, D.; Kuzmich, M.** (2003). *Simulation of Cluster Architecture Subsystem of Caching and Load Balancing of WEB/Application Servers of Data Warehouse*. Vestnik of YANKA KUPALA STATE UNIVERSITY OF GRODNO (YKSYG), 2(22), 107-117 (In Russian).
- [17] **Goldman, D. & Kuzmich, M.** (2004). *Cluster Architecture Dynamic Model of Caching and Load Balancing of WEB/Application Servers*. In Collection "FKS - XII". Grodno, 20-22 (In Russian).
- [18] **Kostka, M. S.; Voltchok, V.; Goldman, D.** (2005). *The Elements of the Conceptual Approach to Creation of the Multidisciplinary Educational GIS Center (MEGIC) in the University of Finance and Management in Białystok*. The collection of the proceedings "Nowoczesne technologie informacyjne i ich wpływ na funkcjonowanie podmiotów gospodarczych". Białystok: The University of Finance and Management in Białystok, 33-49.
- [19] **Voltchok, V.; Kostka, M. S., Goldman, D.** (2003). *The Integrated Concept of Creation a Business-Ecological Analytical-Information System of Transboundary Regions of Poland and Western Belarus on the Basis of GIS Data Warehouse*. In Collection "Soft Computing and Measurements"(SCM'2002). Russian Academy of Sciences. St. Petersburg, V.2, 105-108.
- [20] **Voltchok, V. & Goldman, D.** (2004). *Elements of Integrated Methodology of Development of Data Warehouse*. State Organization "Belarusian Institute of System Analysis and Information Support for Scientific and Technical Sphere", part n.: D200429 (In Russian).
- [21] **Voltchok, V. & Goldman, D.** (2004). *The Integrated Concept of Creation of Business - Ecological Information-analytical System on the Basis of GIS Data Warehouse*. Vestnik of YANKA KUPALA STATE UNIVERSITY OF GRODNO (YKSYG), 1(25), 89-99 (In Russian).