World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:9, 2015

Agile Methodology for Modeling and Design of Data Warehouses -AM4DW-

Nieto Bernal Wilson, Carmona Suarez Edgar

Open Science Index, Computer and Information Engineering Vol:9, No:9, 2015 publications.waset.org/10004413.pdf

Abstract—The organizations have structured and unstructured information in different formats, sources, and systems. Part of these come from ERP under OLTP processing that support the information system, however these organizations in OLAP processing level, presented some deficiencies, part of this problematic lies in that does not exist interesting into extract knowledge from their data sources, as also the absence of operational capabilities to tackle with these kind of projects. Data Warehouse and its applications are considered as non-proprietary tools, which are of great interest to business intelligence, since they are repositories basis for creating models or patterns (behavior of customers, suppliers, products, social networks and genomics) and facilitate corporate decision making and research. The following paper present a structured methodology, simple, inspired from the agile development models as Scrum, XP and AUP. Also the models object relational, spatial data models, and the base line of data modeling under UML and Big data, from this way sought to deliver an agile methodology for the developing of data warehouses, simple and of easy application. The methodology naturally take into account the application of process for the respectively information analysis, visualization and data mining, particularly for patterns generation and derived models from the objects facts structured.

***Keywords*—**Data warehouse, model data, big data, object fact, object relational fact, process developed data warehouse.

I. Introduction

ORGANIZATION are increasingly demanding processing large volumes of information in the most varied formats, you need to have accurate and timely information, organizations are interested in using the historical information to understand what were the behaviors of their operations (sales, customers, orders, suppliers, deliveries), in this context, business intelligence tools and big data takes courage to meet the requirements of these new requirements and especially to implement models of estimation and classification, which is why the development of data warehouse It is quite different from developing standard operating systems, as mentioned [1]; not only structures underlying source systems have to be considered, but also the objectives and strategies of the company. Data warehouse development involves integrate numerous domain specific modeling approaches. Enterprises also often model their goals in terms of formal or semiformal goal models [2].

The design and implementation of data warehouse in organizations has played a key role in the strategic development of them and especially to generate competitiveness in your environment, through knowledge extraction, estimation model generation, provides valuable elements to strategic planning and contributes to the added value from the information coming from ERP [3].

The purpose of this paper is to present a consistent methodology to develop a DW to implement the requirements of an organization. The proposed methodology takes its starting point as the requirements formulated by the stakeholders of the organization, under an agile approach and focused on the customer. It then moves to short and interactive cycles where the project team composed of Master-DW Architect-Dw, Design-DW held activities modeling, design and implementation of the DW [3].

II. Research Methodology

The research methodology was developed under an exploratory approach, partly through a review of the state of the art and standards associated with the modeling, design and implementation of DW, considering the paradigms of data modeling as MOLAP (Multidimensional), OLAP (On line), Holap (Hybrid), as well as methods for agile software development (XP, Scrum, AUP, FDD etc.). Finally, it is taking into account trends associated with Big Data technologies, Data Analytics as Hadoop, HBase, MongoDB and others.

III. Phases of the Methodology -AM4DW-

Agile methodology for modeling and designing data warehouse AM4DW, is inspired by the agile software development model in which mainly takes into account the agile manifesto in which the perspective of collaborative work, individuals and interactions highlights, focused on the client, the response to the change, interactive planning and accelerated productivity-oriented work. Fig. 1 shows that AM4DW methodology can visualize the different stages and cycles associated with the proposed development model. Following the steps of the methodology are described presented the activities, products and roles of actors in the process.

IV. Stage of the Methodology Description

The methodology is composed of 6 stages, the first focused on the management of requirements (identify, validate,

Wilson Nieto Bernal is Profesor with the Universidad Norte (Colombia). Systems Engineer and Specialist Software Engineering, Universidad Industrial de Santander (UIS) Colombia, Master /Expert in Technology Management, Master of Computer and PhD in Computer Science, ULPGC-Las Palmas GC. (Spain), 2007 (e-mail: wnieto@uninorte.edu.co).

Carmona Suarez Edgar is Systems Engineer Colombia, Master Expert in Technology Management, Master of Computer and PhD in Computer Science, ULPGC-Las Palmas GC (Spain).

prioritize and developed), the following 5 stages are carried out in an interactive loop cycle in this case the modeling, design, detailed design, architecture and implementation of the DW. These are executed repeatedly getting initial, intermediate and final versions of the data warehouse. Each deliverable the project team and stakeholders (Business Architect, Data Warehouse Analyst, Data warehouse Users, Business Owner), validate the product obtained and advance to the next cycle if necessary, otherwise the cycle is closed with the final product [5]. The maximum number of iterations of the cycle is n. The cycles are characterized in that they are oriented to obtaining the product of way incremental until obtain the final DW, the development of each cycle should not take more than five weeks. The cycle ends with a respective review and takes just the necessary adjustments in the project (time, costs, risks and scope.) (time, costs, risks and scope.)

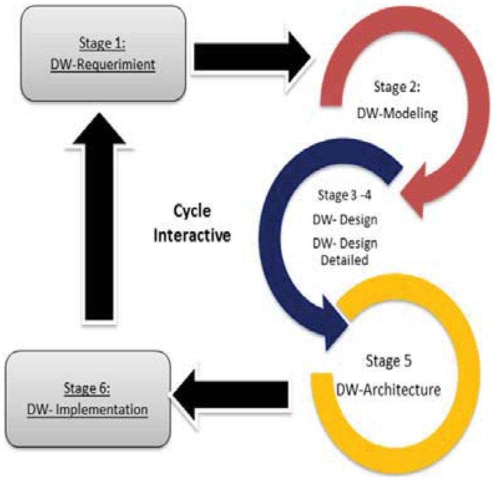

Fig. 1 Methodology -AM4DW-

A. Stage 1 Data Warehouse Requirements

This stage has as objective to identify information requirements related with the context of the data warehouse. The starting point is to understand the context of the application in which the project team can identify, analyze, understand and validate the knowledge to generate from Data Warehouse. (for example: the number of units sold in an online store units per month of the year). This data is obtained from the transactional database from the ERP, as also not structured information from sources (documents, reports, proceedings, agreements, spreadsheet, etc.) also the organization as stories expert users who provide details of transactions [6].

At this stage of the methodology, some of the following techniques can be used, such as elicitation of requirements, analysis of documents, meeting with experts from business, questionnaires, interviews structured, workshops with the stakeholders, prototype evaluation, review of standards for the developing of the data warehouse, product evaluation, user stories, business cases analysis and inverse engineering [4]. Some criteria for the evaluation and acceptance of the requirements are: check that are clearly established, complete, consistent one with other, unique identification for the respect traceability, consistent with the focus of the Architecture that is going to be developed, that facilitate the processes of: develop, verification and validation.

Activities:

- Identify stakeholders
- Obtain necessities and stories users.
- Identify interface requirements.
- Develop customer requirements
- Analyze requirements.
- Validate requirements.
- Prioritize and encapsulate requirements.

Products:

- List of requirements.
- Identify source of data. (unstructured and structured)
- Data mining Project Vision.
- Data Warehouse Development Plan
- Risk List
- Test Plan
- Glossary.
- Data warehouse Architecture Base Document

Stakeholders: Requirements Engineer, Data warehouse Analyst, Data warehouse Users, Business Owner.

V. Description of Cycle Interactive

A. Stage 2 Data Warehouse Modeling

This stage of methodology aims to obtain a first version of the DW data model, starting with the requirements established in the previous phase. Furthermore, it is important to perform a cycle planning (modeling, design and implementation) in order to obtain an initial version of the Data warehouse. [6], [7]. The requirements at this stage should be well defined and thus focus on the development of DW.

Activities:

- Integration data and modeling Data
- Integration Restrictions. (Time, Dimensions)
- Modeling Data detailed (Molap, Rolap, Holap)

Products:

- Model Data
- Model Data detailed
- Model OLAP
- Model restrictions.
- Update Data warehouse Architecture Base Document

Stakeholders: Business Architect, Data warehouse Analyst, Data warehouse Users, Business Owner.

B. Stage 3: Data Warehouse (Object Relational) Design:

This stage has as objective to develop the design of Data Warehouse that represent functional and nonfunctional requirements from Data warehouse obtained of the previously requirements stage, this design is represented by topics that are extracted from *user's stories* or from different techniques explained before, some examples topics could be: sales, products, distribution, deliveries, transactions or services [7], [8]. The base information for elaborate Object Relational Design could come from structured or unstructured data sources, in first instance could be hosted on transactional database that commonly is supported in the organizational ERP, as second instance come from productive file repositories such as: documents, agreements, proceedings, maps, etc. Representation unity that commonly describes the data warehouse requirements is the *object fact*, which is an information entity compose of: its name, its attributes and their operations. The attributes regularly are the explicit variables that describe the dimensions, they are foreign keys to entities that describe the dimensions and the third component is the aggregation operations which come from the data consolidation. Fig. 2 shows an example of the Object Relational.

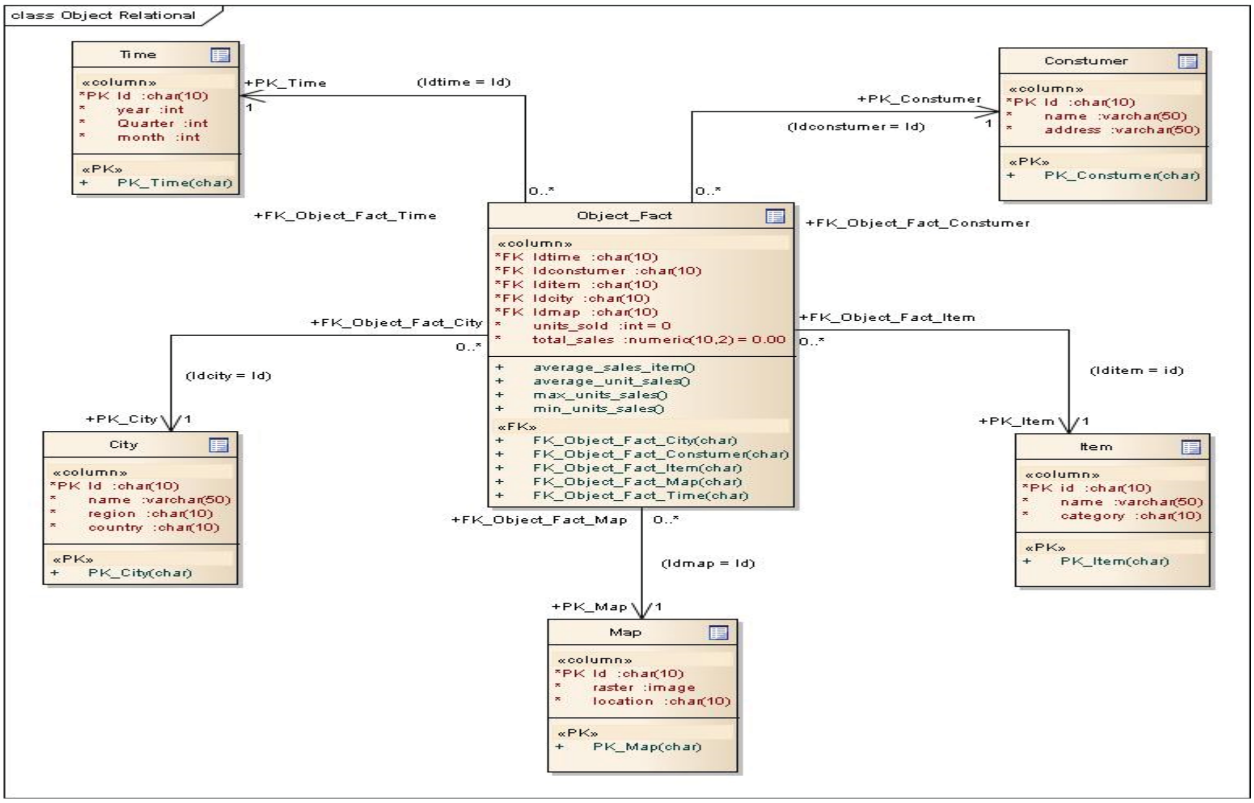

Fig. 2 Object Relational Data warehouse

Activities:

- Integrate granularity level general.
- Integrate important information unities.
- Integrate the dimensions that support the requirement.
- Integrate time unity.
- Obtain an Object Relational model.
- Validate Object Relational model.

Products:

- Data mart under Object Relational focus.
- Object Relational fact.
- Constellations and their dimensions.
- Glossary enlargement.
- Supplementary Specifications.
- Update Data warehouse Architecture Document

Stakeholders: Data warehouse Analyst, Data warehouse Users, Business Owner

C. Stage 4: Data Warehouse Design Detailed

This stage seeks to develop a detailed design of the DW; it takes as its starting point the Relational Object Design. At this stage, detailed design contains levels of granularity, the information entities, dimensions, and time units comprehensively describing the DW requirement. This detailed design has a main objective to model and design the object fact, which describes the dimension data warehouse. The object fact can be developed as a relational or multidimensional constellation.

Activities:

- Design detailed granularity level.
- Design Detailed and integrate entity.
- Design Detailed and integrate the dimensions.
- Design Detailed and integrate time unity.
- Design and Validation Object Relational model.

Products:

- Data mart under Object Relational focus.
- Object fact.
- Constellations and their dimensions.
- Data model specifications.
- Glossary enlargement.
- Supplementary Specifications.
- Update Data warehouse Architecture Document

Stakeholders: Business Architect, Data Analyst (Data Warehouse), Data warehouse Users, Business Owner.

D.Stage 5: Data Warehouse Architecture

This stage has as an objective to obtain a data warehouse architecture, which is going to be developed into an incremental manner from the object relational model; the architecture is conforming by components or technology artifacts [8], [9]. Its description is represented by a of layers superimposed model which is shown in Fig. 3. This architecture could be represented by components diagram under UML. The layers that compose the Object Relational Architecture are: *the Data layer, the Object Relational Layer, the Server Object Relational Layer, the customer and user interfaces layer*.

- **The Data Layer:** Contains the structured and unstructured information sources which could be used to build the repository that in future is going to be used in the data warehouse construction.
- **The Object Relational Layer:** Contains the data warehouse object relational model, which is a data warehouse conceptual model.
- **The Server Object Relational Layer:** Contains the Object Relational server and all the tools that are going to facilitate the information visualization tasks.
- **The Customer and User Interfaces Layer:** Contains the user interfaces that allowed visualize the models, patterns, data warehouse views related with established requirements.

Activities:

- Integrate and adapt the data warehouse architecture implementation.
- Describe, specify and established the data warehouse architecture implementation.
- Validate the data warehouse architecture implementation.

Products:

- Data warehouse architecture.
- Update Data warehouse Document.

Stakeholders: Data warehouse Architect, Data warehouse Users, Business Owner.

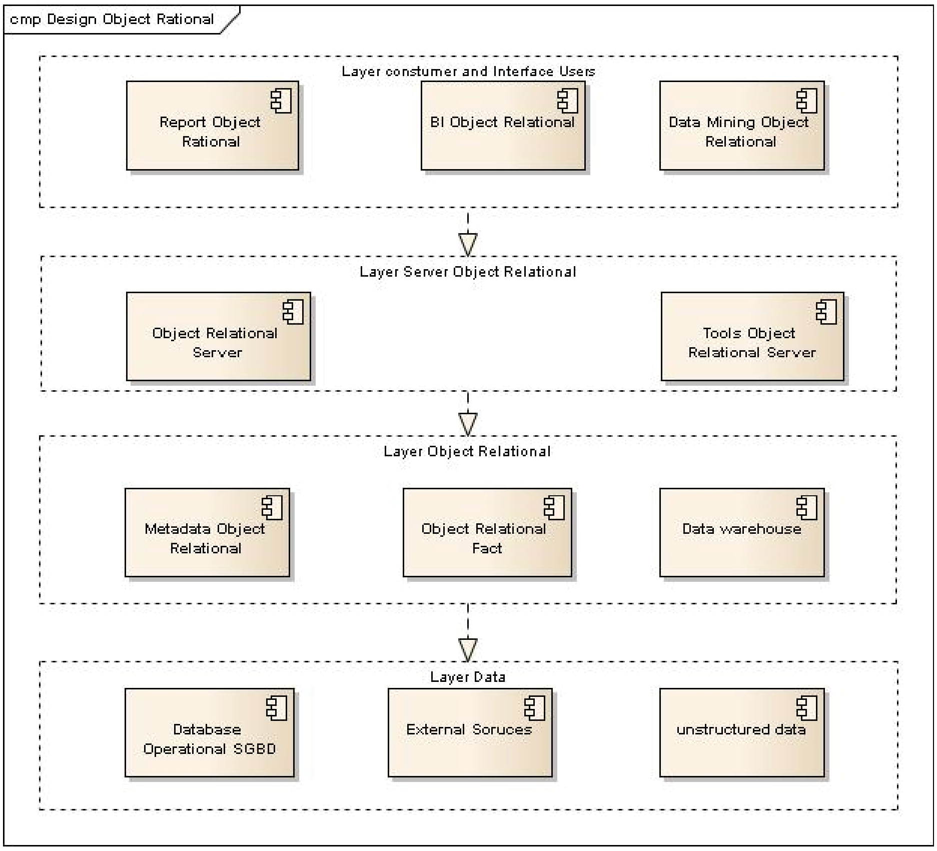
cmp Design Object Rational
Layer constumer and Interface Users
Report Object Rational
BI Object Relational
Data Mining Object Relational
Layer Server Object Relational
Object Relational Server
Tools Object Relational Server
Layer Object Relational
Metadata Object Relational
Object Relational Fact
Data warehouse
Layer Data
Database Operational SGBD
External Soruces
unstructured data

Fig. 3 Architecture Data Warehouse

E. Stage 6: Data Warehouse Implementation

The data systems that contain the data warehouse demand a high technological performance in storage and processing, related requirements with performance in terms of interoperability are special key if they need to be integrated from different data sources. The implementation of this type of system requires a high level support processing and modeling, computational some valid are especially algorithmic techniques that allow to obtain models or patterns, in this case, Bayesian models, tree based models, cluster methods, linear and nonlinear regression, as well as neural networks and the more recent techniques of spatial data mining [10], [11]. This implementation could be represented with an UML components model. Fig. 4 shows a view of the Relational

Open Science Index, Computer and Information Engineering Vol:9, No:9, 2015 publications.waset.org/10004413.pdf

Object Implementation Framework with could be deployed through an UML deployment diagram. The objective is obtain into an incremental manner a stable data warehouse object relational architecture, which articulated with the Object Relational model, allows visualize each technological components from data warehouse implementation.

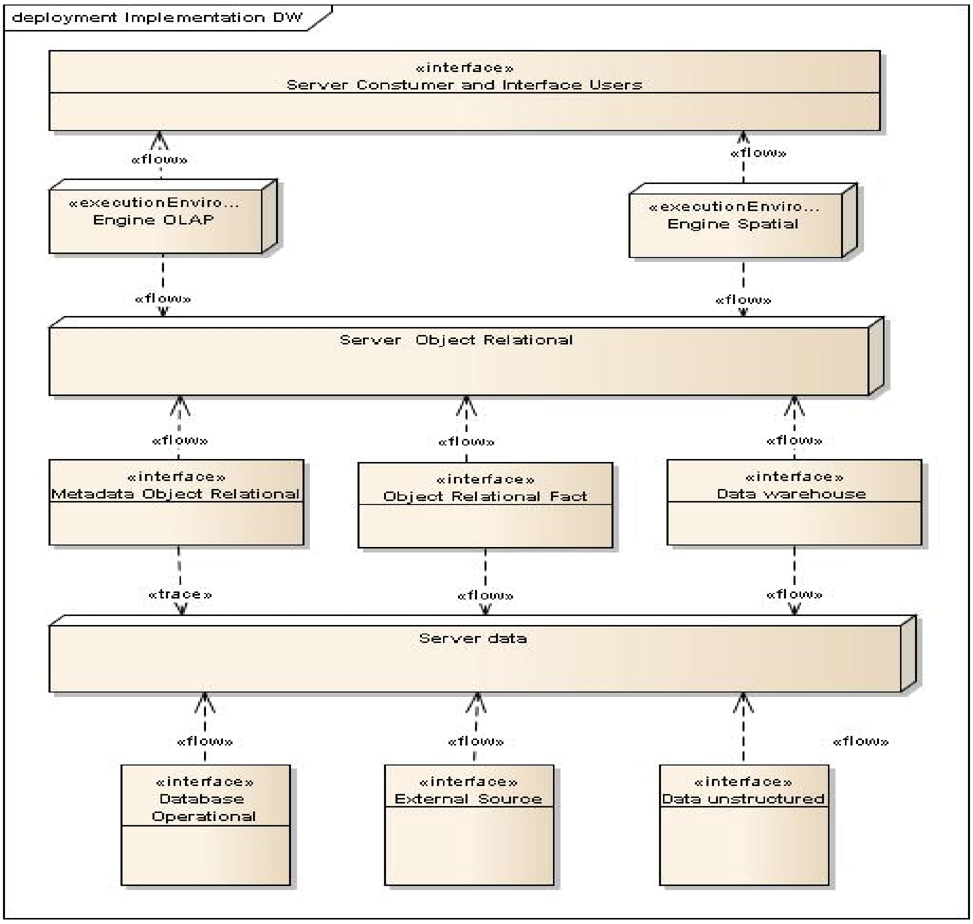

Fig. 4 Object Relational Implementation Framework

The Relational Object Implementation Framework is a computational structure (hardware and software) that support the relation object model and the relational object architecture, the implementation framework describe the technological characteristics that have to be taken care to deploy the data warehouse system [3]-[6], [12]-[14]. It's composed by the following layers which are described below and that graphically could be seen at Fig. 4:

- ***Data Source:*** It's composed by the operational database, external sources unstructured data.
- ***Data Server***: It's the physical server that contains the data sources.
- ***Interface Apps***: Contains the applications which could be accessed to Relational Object Metadata, the Relational Object Metadata, the Relational Object Fact and the data warehouse.
- ***Relational Object Server:*** Contains OLAP engines and the Spatial functions to log in to the different data warehouse operations.
- ***User Interfaces***: Contains the applications that visualize the end user.

Activities:

- Instances each component from the data warehouse architecture implementation.
- Implements relational object data model using data warehouse management systems.
- Describe, specify and established the data warehouse implementation.
- Validate the data warehouse implementation.

Products:

- Data warehouse implementation framework (Fig. 4).
- List of data warehouse implementation components.
- Data warehouse implementation specifications.
- Data warehouse glossary enlargement.
- Data warehouse supplementary Specifications.

Stakeholders: Application Developer, Information Architect, Infrastructure Architect, Data Warehouse Users, Business Owner.

VI. Roles Description Methodology -AM4DW-

Requirements Engineer: He is Responsible to identify, organize, validate, and prioritize and requirements grouping from stakeholder's interaction and product owners, and different techniques could be used in these activities, such as: structured interviews, information analysis, business expert's discussions, IT products benchmarking and the analysis of trends associated with the technological product that is going to be develop.

Business Architecture: The role of Business Architect is associated to who is responsible of the modeling, design and modeling of the seeing organization through business process modeling (BPM), workflows (WK), time line, service identification components (SOA) and the map strategy modeling.

Information Architect is responsible of creating the DW architecture. Which consists in creating a multi-layer model where the different components of DW (classes, objects, tables, etc.). Addition he is responsible for the description of the information that forms the Data Object Relational, design Database from information objects (Physical Database), Enterprise Information Integration (EII), Database, Data Repository, Data Warehouse design and the Information Repository using the Database Management System.

Application Architecture: is the one responsible of elaborate the Application Architecture, High Level Software Description, Software Components Design and service details, Software Components Integration and services and the extended system design or distributed based on software components.

Computing Architecture or Infrastructure Architect: is responsible of doing the modeling, design of the Infrastructure Architecture, describe the physical business distribution (networking), design the physical business distribution (network), design the physical business distribution (Lan, Man, Wan, Wifi, Pan) and the design of extended physical business distribution (Networking Extend).

VII. Conclusion

The agile methodology for the development of Data Warehouse proposal -AM4DW- seeks to respond to projects related to data modeling large scale. That attempts to break the technological myth, from there, it presents a methodology composed of a set of phases, activities and techniques relational data modeling and data modeling object-oriented architecture that create adequate information for organizational data warehouse projects.

The proposed methodology seeks to carry out projects of under time constraints, requirements, specifically oriented productivity and obtaining partial results or deliverables of the solution, which can develop technologies based on SQL or NoSQL.

References

[1] N. Schahovska, Data warehouse and Dataspace – information base of Decision Support System, CADSM'2011, 23-25, pp 170-173.

[2] V. Stefanov, B. List, Business Metadata for the Data warehouse Weaving Enterprise Goals and Multidimensional Models, 10th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW'06), pp 0-7695-2743-4/06.

[3] Boutkhoum, Hanine, Tiknioiune, Agouti, Integration approach of multicriteria analysis to OLAP systems: Multidimensional model, ISBN 978-1-4799-0792-2, 20013, pp 1-4.

[4] H. Kuchibhotla, D. Dunn, D. Brown, Data Integration Issues in IT Organizations and a need to map different data formats to store them in relational databases, 41st Southeastern Symposium on System Theory University of Tennessee Space Institute Tullahoma, TN, USA, March 15-17, 2009, pp: 1-6.

[5] M. Mohajir, A Latrache, Unifying and incorporating functional and non functional requirements in Data warehouse conceptual design, 978-1-4673-2725-1/12 ©2012 IEEE, pp 49-57.

[6] A. Singh, Implementation Model for Access Control using Log Based Security, 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA). IMS Engineering College, Ghaziabad, India, pp 290-293.

[7] A. Januszewski, T. Pankowski, Modeling Analytical Indicators Using Data warehouse Meta model, Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA'06), pp 0-7695-2641.

[8] M. Mior Nasir, Enriching Hierarchies in Multidimensional Model of Data Warehouse using WORDNET, 3rd International Conference on Research and Innovation in Information Systems – 2013 (ICRIIS'13), pp 296-301.

[9] O. Boutkhoum, M. Hanine, A. Tiknioiune, T. Agouti, Integration approach of multicriteria analysis to OLAP systems: Multidimensional model, 978-1-4799-0792-2/13/$31.00 ©2013 IEEE. pp without number.

[10] C. S. Jensen, A. Kligys, T. B. Pedersen, I. Timko, "Multidimensional data modeling for location-based services", VLDB Journal, vol. 13(1), pp. 1–21, 2004.

[11] E. Malinowski and E. Zimányi, Advanced data warehouse design: from conventional to spatial and temporal applications, Springer, first edition, 2008.

[12] Z. Kouba, K. Matousek, and P. Miksovský, "On data warehouse and GIS integration, 11th International Conference on Database and Expert Systems Applications", DEXA, pp. 604–613.

[13] A. C. Ferreira, M. L. Campos, and A. K. Tanaka, "An architecture for spatial and dimensional analysis integration", 6th World Multiconference on Systemics, Cibernetics and Informatics (SCI), pp. 392–395.

[14] S. Shekhar, C. T. Lu, X. Tan, and S. Chawla, Map cube: a visualization tool for spatial data warehouses, chapter in: H. J. Miller, J. Han (eds.), Geographic data mining and knowledge discovery, Taylor and Francis, pp. 74 – 109.