View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UUM Repository

Conceptual Design Model Using Operational Data Store (CoDMODS) for Developing Business Intelligence Applications

Muhamad Shahbani † and Norshuhada Shiratuddin†

†Graduate Department of Computer Science, College of Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia.

Summary
Building a Business Intelligence (BI) application is very challenging as it is a young discipline and does not yet offer well-established strategies and techniques for the developments process when compared to the software engineering discipline. Furthermore, information requirements analysis for BI applications which integrate data from heterogeneous sources differs significantly from requirements analysis for a conventional information system. Conceptual Design Model Operational Data Store (CoDMODS) to build BI application that focuses on operational information to support business operations is proposed. In this model, combination of community interaction and data integration approach were used to identify the requirements for developing BI application. Furthermore, how the operational data store can be used for operational and tactical information and can be transferred to a data warehouse for supporting analytical information and decision making is also presented. Finally, to verify and validate the proposed model, the case study approach using web application development in selected subject areas is elaborated.

Key words:
Operational Data Store, Operational and Tactical Report, Data Warehouse, Business Intelligence, Web Application

1. Introduction

The success of system application depends on how well it fits the requirement from the users and its environment [1]. Requirements are the foundation for building system application where it is determined what the system must do and drive the system development. The requirement's development process identifies the activities needed to produce a set of complete and verifiable requirements. In Software Engineering (SE) knowledge area, a requirement is concerned with the elicitation, analysis, specification and validation of software requirements [2]. There are many researches discuss about software requirements in software engineering discipline, which is focused to develop transactional system development. However, information requirements analysis for Business Intelligence (BI) applications which integrate data from heterogeneous sources differs significantly from requirements analysis for a transactional information system [3].

BI is the process of gathering meaningful information about the subject matter being researched [4]. In information system perspective, BI is a combination of operational data with analytical tools to present complex and competitive information to planners and decision makers [5]. Furthermore, making decision based on comparing responses coming from different sources in operational data has always been difficult [6]. The structures commonly used in BI architecture are an operational data store (ODS), data warehouse (DW) and data mart. To date, there are many topics researched in DW structure (which support analytical information) but fewer studies on ODS structure. ODS is subject oriented, integrated, current valued and volatile collection of detailed data that provides a true enterprise view of information [7]. The major difference between the ODS and DW is ODS contains current and detail data while DW contains summary data to support analytical information for making decision [8]. Moreover, ODS structure can support both operational and analytical information for a decision making.

Since the success of a system application depends on how well it fits the requirements from the users and its environment [1], we propose CoDMODS to build BI application that focuses on operational information to support business operations. In this model, combination of community interaction and data integration approach were used to identify the requirements for developing BI application.

2. Theoretical Consideration

This paper is based on the concepts of the Corporate Information Factory (CIF) introduced by Inmon [8] and Business Dimensional Lifecycle for Business Requirement suggested by Kimball [9] to develop BI applications. Furthermore, the concept of Information Pyramid as shown in Figure 1 can be seen as different types of information and different view of users.

Manuscript received March 5, 2011
Manuscript revised March 20, 2011

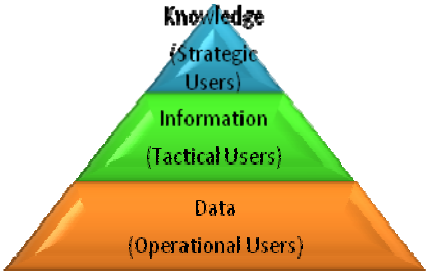

Fig.1 Information Pyramid

Inmon points out the alignment of different DW elements with different requirements as shown in Figure 2. The ODS, DW and older legacy applications are combined to create a common CIF architecture. The raw and detail data is integrated and transformed into an ODS or current detailed level of a DW. As the refined data passes out of the ODS it goes into the current level of the DW. The design process in DW must be orienting to the end users' need through the two main types of an object those are fact and dimension tables [10]. Information processing can be done throughout ODS level, at a current level of detail, or at the data mart level of detail. ODS was built by collecting and cleansing data from operational raw data [7]. Extract, transform and load (ETL) is the processes that enable an organization to move data from multiple sources, reformat and cleanse it, and load it into another database, especially ODS to support a business process in current detail information.

In contrast, the concept of Business Dimensional Life Cycle for Business Requirement proposed by [10], guides the developer in making strategic choices to prioritize subject areas and how to present required information on the users' screens. The important points to understand here are understanding business requirements, securing solid business sponsorship, defining enterprise-level business requirements and identifying detail subject areas of business requirements. The relationship between information meaning and knowledge in the subject area is important to determine business requirement [11].

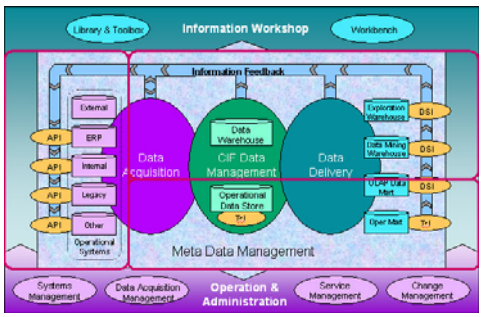

Fig.2 Corporate Information Factory [5]

Basically, existing DW development approaches can be classified within three basic groups; data driven, a goal driven and user driven [12]. The researchers such as [13] and [14] argues that a DW environment is data driven, which are requirements are understood after it is populated with data and being used by the decision support analyst. Moreover, Goal Oriented Requirement Analysis for Data Warehouse (GRAnD) suggested by [15] adopts two different perspectives for requirement analysis-organizational modeling centered on stakeholders and decisional modeling focused on a decision maker. In this approach, DW project must fit with organization business objectives.

On the other hand, user-driven or demand-driven approach adopts involvement of end users in data warehousing as suggested by [3]. Data modeling in DW requirement derives a data model directly from a user query requirement without considering the data sources and business goal. Unfortunately, there are fewer studies about ODS, which support information processing at current details level or operational data in BI applications, especially in ODS requirements compared with studies about requirement analysis in a data warehouse.

3. The Proposed CoDMODS Model

CoDMODS Model is a conceptual model to develop BI application, which is focused on operational data. In this model, a two-phase requirements process, at the organizational and subject area levels, is proposed. In between the two-phase requirement process, there is a requirement elicitation process which contains requirement gathering approach and community collaboration method.

Figure 3 illustrates the proposed model. Each phase is broken down into several sub-activities. For each phase, sub-activity begins with understanding project domain and ends with writing up requirement specification and BI implementation in a different level of details.

For the requirement elicitation process, three requirement gathering approaches (goal oriented, data oriented and user oriented) to gather requirement from an organizational level and subject area level are suggested. Besides that, community interaction and collaboration using such as interview, survey, data profiling and reports review to gather the requirements are emphasized. In addition, online collaboration for system stakeholder/users to discuss system requirements using a forum, chat, blog, email and digital article are utilized.

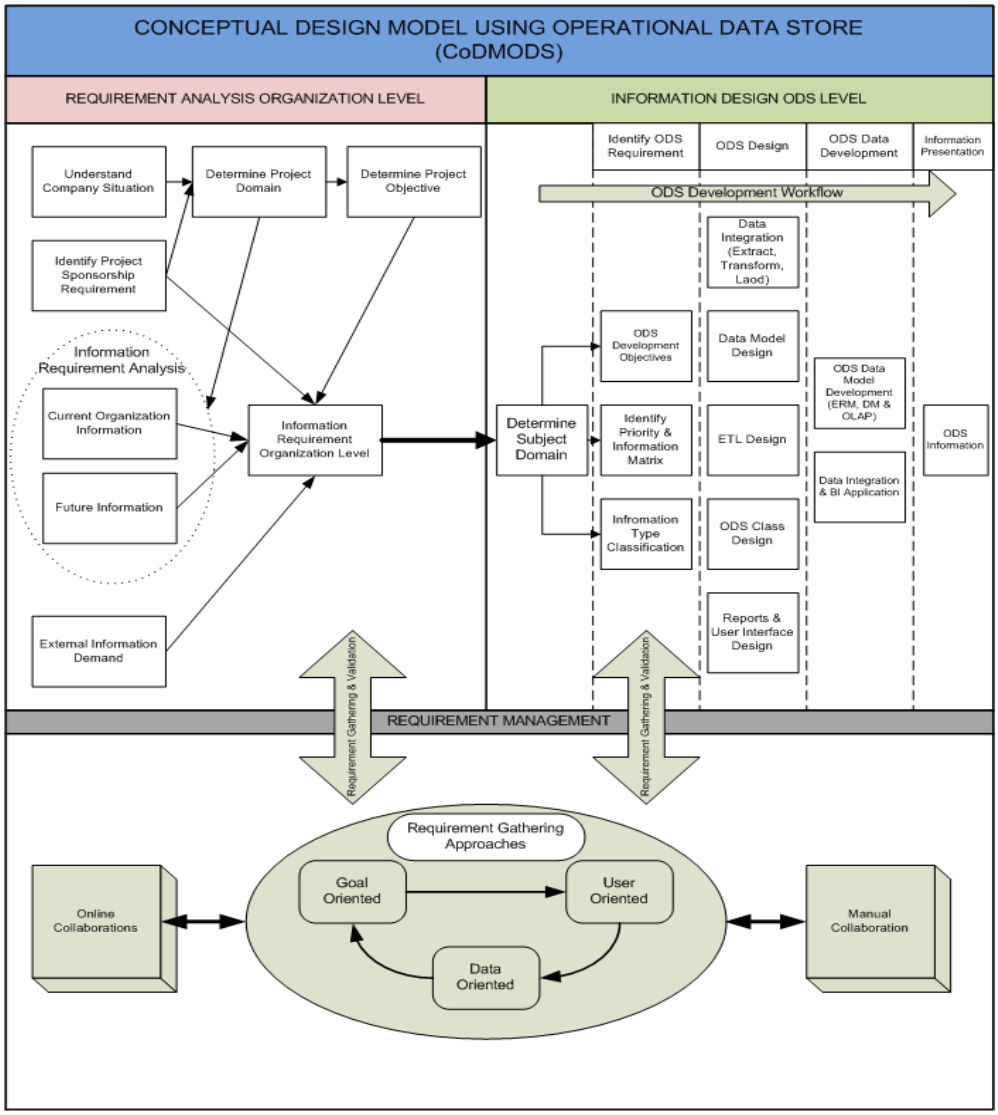

Fig.3 The Proposed CoDMODS Model

3.1Organization Level Requirements

In the Organizational Level Requirement phase, it starts with understanding project domain from an organization level to build an enterprise information infrastructure and ends with writing up the organizational level requirement specification. The organizational level requirement step gathers a broad and horizontal view of the organization from business point of view. It involves the following four steps.

Step 1: Analysis Company Situation. This process involves understanding the organization's vision and motivation, structure and business activities. The information about business organization can be retrieved from organization website, reports and sources from management staffs. By understanding business organization structure, BI developer can understand data sources and information flows in the organization. Business activities in the organization can derive a clear picture for the motivation, guideline, business functions and project scope in BI application. Then, related process in the specific business area such as insurance, retail, banking/finance, education, and telecommunication will be elaborated. Each project domain has different business functions, requirements, processes and questions depending on the type of business sector. Here domain analysis concept can be utilized to understand the domain knowledge in the specific project domain based on past business processes. The sources of domain knowledge are from technical literature, existing implementations, customer survey, expert advice and current/future

requirements.

Step 2: Identify Project Sponsorship. Project Sponsor is the person/organization that is ultimately responsible for the project within an organization. Normally, it involves a senior management post such as chief executive officer, managing director, general manager and the owner of an organization. Typically, project sponsorship is responsible for: 1) championing the project; 2) obtaining budget approval for the project; 3) accepting responsibility for problems escalated by the project manager and 4) document approval for the project. Project Sponsorship can be categorized into three types of organization that are a government agency, business organization and software developer. Each type has a different business motivation for developing BI applications. Good business sponsorship can provide the resources and support to deliver real business value.

Step 3: Organizational Requirement Analysis. In this step, the focus is on the high level requirement for gathering information process in an organization. Three approaches to gather requirements are used; 1) goal driven; 2) user driven; 3) data driven. Goal driven approach is based on business motivations set up by an organization. The user driven is based on demand from the users and the last approach is based on data profiling on existing organization data sources. In gathering the requirements, community collaboration techniques (such as survey, data profiling, a forum, chat, blog, email) are emphasized to convey requirements. Requirement validation is also ascertained in this step. A high level requirement collected from previous step will be analyzed in this step. Three main processes are proposed: 1) Build Initial Matrix; 2) Conduct Prioritization Session and 3) Write up Summary Requirement. The business processes in the initial matrix become the major inputs to the requirements for prioritization session. The prioritization process is a meeting involving the BI team and business sponsor/senior management to describe the business process. A prioritization grid is normally used to illustrate the prioritization process which consists of Y axis as business value and X axis as a level of effort. The output of this prioritization process is a list of a business process in priority order. The Business Sponsor will decide several business processes/subject area to be selected for the BI project. The requirements not only focused in the current situation but should be focused in future information. Furthermore, the information requirements from external organizations such as tax department, statistical department and partnership should be analyzed because the organization needs to prepare the information for this organization. Requirement process for organizational-level can be shown in Figure 4.

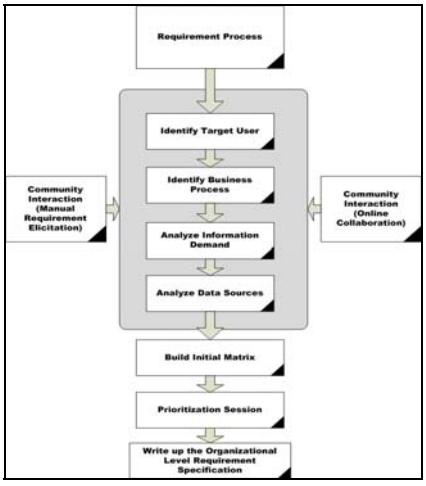

Fig.4 Organizational Requirement Analysis Process

Step 4: Information Requirement in Organizational Level. BI project manager is responsible for writing the summary of requirement specifications in the organizational level and this specification is used for the next detail requirements process for ODS data development.

3.2 ODS Design Level

ODS Design Level focuses on requirement, analysis, design and development of ODS Data Model in a specific subject area, especially to develop operational and tactical information. This phase involves the following five steps.

Step 1: Determine Subject Area Domain. In this step, a specific subject area which is identified in the previous phase is elaborated. General sub-steps as proposed by [20] are adopted: establish the subject area, collect domain expert, establish the depth and width analysis and define the specific domain objects, relations and constraints. The outputs of this step are taxonomies, standard interfaces, functional models and domain languages as a requirement to develop a software system. Furthermore, in this model we proposed subject area analysis, which is consisted of source of subject area knowledge, source of subject area references, requirement gathering methods and information requirements for specific subject area to define an information requirement in a subject area as shown in Figure 5. Subject area analysis process can deliver a detail information requirement in the specific subject area based on domain knowledge and the information can be divided by operational, tactical and strategic information.

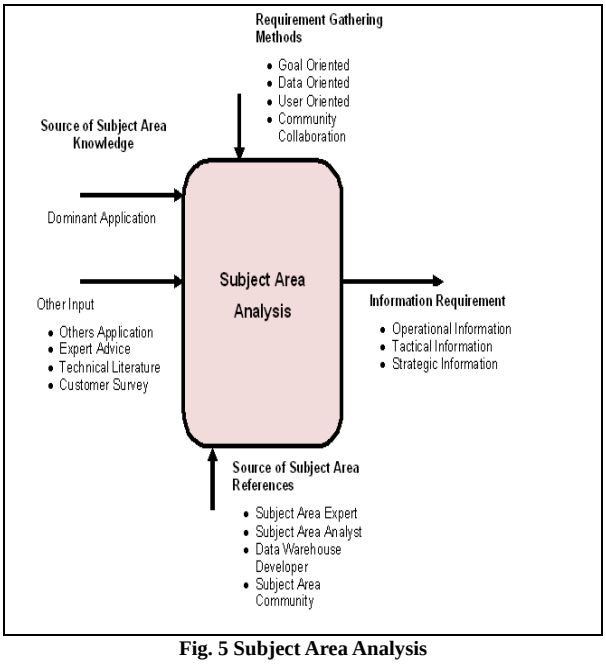

Fig. 5 Subject Area Analysis

Step 2: Identify ODS Requirement. Here attention is given to the detail requirements in specific subject areas. In this step, ODS development objective is defined to guide the information building in the organization. Then, this process follows by identified information priority and develops the information matrix. Another process in this step is information classification, which is identified information type in three different categories that is strategic, tactical and operational information. This step involves a process to analyze reports required by organization based on a previous requirement process. The suggested reports obtained from a requirement process are matched with BI users and type of information status.

Three groups of BI users are defined: strategic, managerial and operational users. The criteria to define these groups are based on discussion results from the community collaboration and type of reports such as detail level of information, frequency of data updated, summarization of data and type of information users. The reports are also classified by information status: 1) need to have; 2) nice to have and 3) not needed. An example of how the specific reports relate to information status and report users can be shown in Figure 6. The example of specific report and relation with information status and report users can be shows as follows: [RS1 : {SU1,1}, {SU2,1}, {SU3,2}, {MU1,3}, {MU2,3}, MU3,2}, {MU3,3}, {OU1,3}, {OU2,3}, OU3,3}].

Step 3: ODS Design. An ODS is an environment where data from a different operational database is integrated to provide the end user community with an integrated view of operational and tactical information. Hence, this step focuses on a detailed requirement analysis for the ODS in specific subject areas. Briefly, there are five tasks in this step: 1) Data Integration Design; 2) Data Model Design; 3) ETL Design; 4) ODS Class Design and 5) Reports and User Interface Design. Data integration design is a task to handle how to extract a quality data from data sources and store in ODS storage. Data integration consists of two primary tasks involving data integration and data transformation from operational data sources to the ODS. The first task in ODS integration is integrating multiple operational data sources from various sources and store in ODS. There are some issues to handle in ODS integration such as what data to store in ODS, type and format of data sources and details level of data to be extract. In data transformation, a data is converted from a source data format into destination data using ETL scripts. Data model design is focused on how to develop the ODS data model in a specific subject area. ODS data model can be design by using Dimensional Model (DM) and Entity Relationship Model (ERM) [9]. Figure 7 shows the example of Dimensional Model diagram, which is consisted of fact and dimension tables.

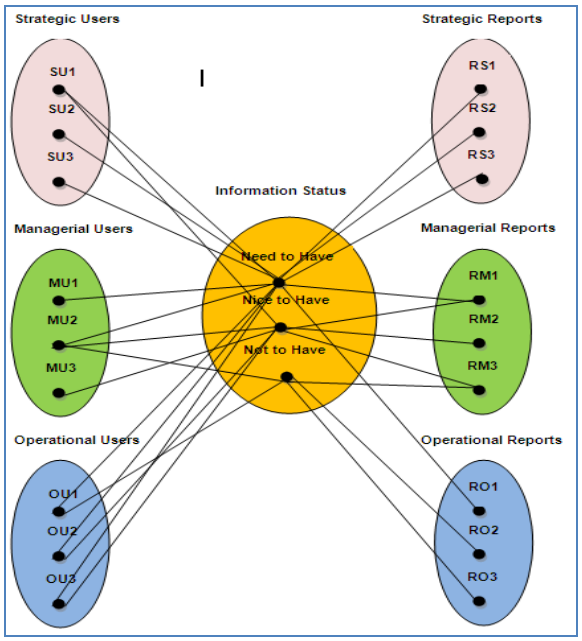

Fig.6 Relationship between reports, report users and information status.

ETL design is focuses on importing of files, summarization, standardization, filtration and condensation of data in requirement format. Transformations include cleansing, summarization, conversion, reformatting, conversions and encoding process. ETL is used to migrated data from an operational database to another format of database. ETL has three database functions that are 1) Extract – the process of

reading data from an operational database; 2) Transform – the process of converting the extracted data into the form it needs to be and 3) Load- the process of writing the data into the target database. ETL Specification, which is referred as a guideline for software/script process to extracting data from multiple sources in various formats, transform it to conform with business need, and loading into the target ODS database.

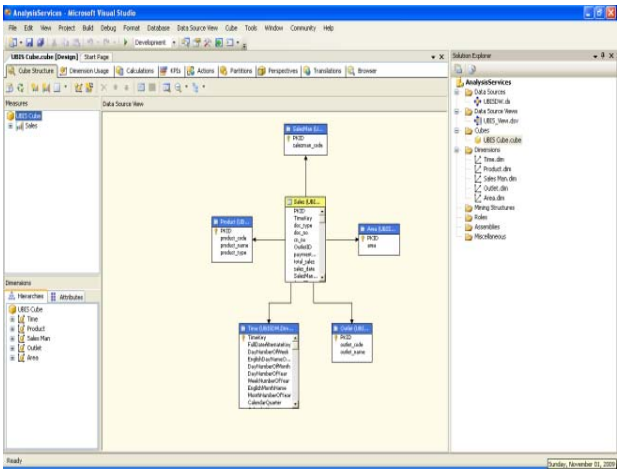

Fig. 7 Dimensional Model

ETL process can be described as sequence activities transforming heterogeneous data into an integrated target ODS database. This process takes time to get the source data, understand the necessary columns, understand the business rules, and understand the logical and physical data models. The requirement for data transformation is important to understand for designing data model and ETL specification. The example of a data transformation process from data sources (Oracle 9i) to target ODS (Microsoft SQL) using an ETL aggregation script can be shown in Figure 8 below.

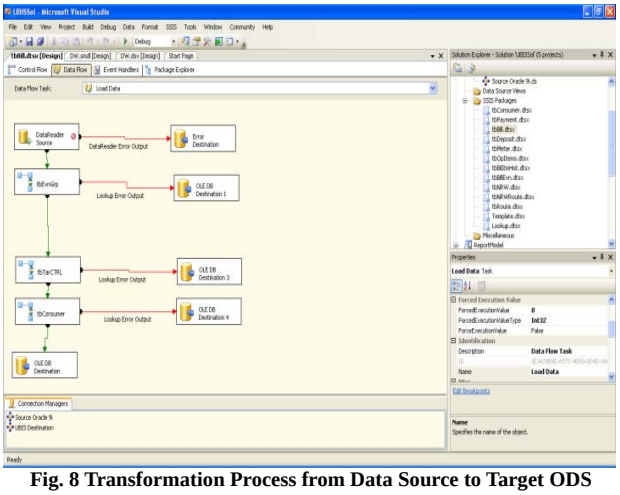

Fig. 8 Transformation Process from Data Source to Target ODS

One of the challenges in the early phases of developing BI application is difficult to identify the appropriate data sources and to specify the operations and transformation needed by a target database. Although there is a tool to facilitate this procedure, we need a detail requirement specification to identify appropriate sources, right transformation and load to the proper target database. Figure 9 shows how data from heterogeneous sources is aggregated and store the formulated value in ODS database. This task needs a detail ETL specification to formulate data from a different environment and position in the required format.

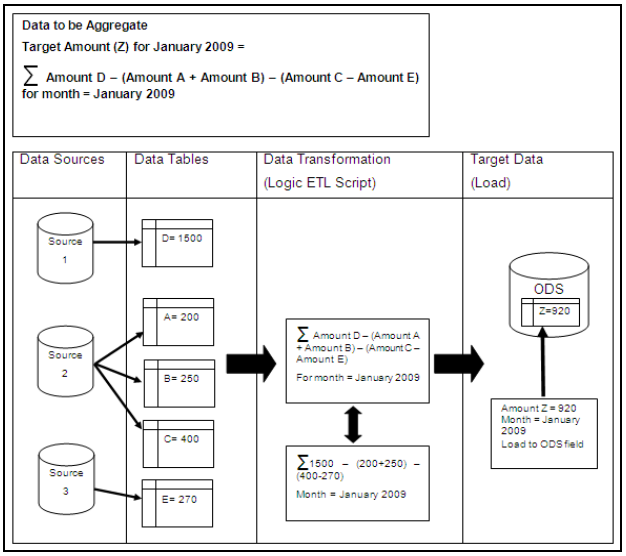

Fig. 9 The Example of Data Source Aggregation

Next, ODS classification design is needed to identify ODS type for every transaction, data model development and a requirement for ETL specification. In this model, there are three types of ODS that is a) ODS type A – Synchronously Update; b) ODS type B – Data Store and Forward and c) ODS type C – Asynchronously Update. They are differentiated by the level of integration between the operational system and ODS. Finally, in ODS design step involves report and user interface design for information presentation.

Step 4: ODS Data Development. ODS data development is a task to develop a data and BI application based on previous design. There are two tasks involves in this step that is 1) ODS Data Model Development and 2) Data integration and BI application development.

Step 5: ODS Informational Presentation. The main purpose of this task is to classify an information type required by organization based on the information pyramid which is, consist of operational information, tactical information and strategic information. This model

emphasizes the role of ODS functions to classify an information for tactical and operational information in business intelligence architecture. However, an information in ODS database can be transformed to a data warehouse for strategic information. ODS Data Modeling focused on preparing a data to support two different roles of ODS that is data operational process and data preparation for decision making.

4. Case Studies - BI Web Applications Using CoDMODS Model

In validating the proposed CoDMODS model, three BI web applications for three organizations were developed as case studies. The organizations involved are utility company, a telecommunication company and an entrepreneur department. Figure 10 depicts a screen shot of one of the applications.

Fig.10 BI Applications Developed Using CoDMODS Model

Figure 11 shows a system flow process started with extracting data from data sources and end with the data presentation layer. ETL scripts can be found in every task for extracting, transforming and loading data to the target database. This diagram also shows the ODS can provide data for operational information while a data warehouse is used to produce strategic information. The sources of ODS database can be constructed from data sources and predefined process.

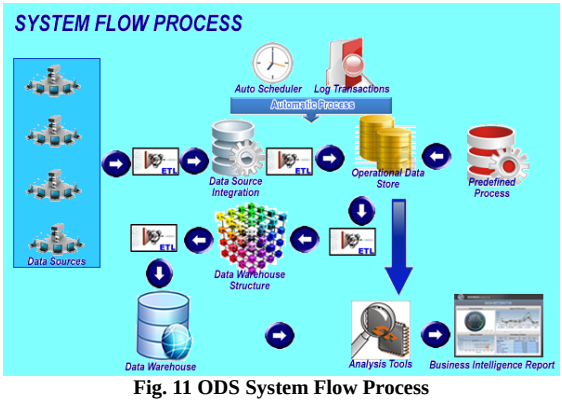

Fig. 11 ODS System Flow Process

Formative evaluations of the applications indicated that these BI applications are usefulness, easy to be used and produced reliable information. After the development of a model, it is necessary to evaluate a model to determine how well the model works. In this case, the proposed CoDMODS model is necessary to validate and verify to ensure this model can work for analyzing a requirement in BI application especially using ODS component. The comment and suggestion from expert review are an input for model modification and enhancement. Three main point issue by experts after validate CoDMODS Model are 1) requirement must relate with an organization vision, mission and main objective; 2) ODS data model must be flexible with a new requirement (based on Kano Model) and 3) strongly agreed for requirement gathering using traditional approach and online collaboration.

The prototypes have been tested throughout the users feedbacks by using Computer System Usability Questionnaires (CSUQ) which is measured the user's satisfaction and usability [16]. CSUQ questionnaire contains 19 questions, and applies the 7 degrees of Likert Scale (1-strongly disagree–7-strongly agree). There are 59 respondents from three organizations that participated to respond to the questionnaires. The results presented were based on these questionnaires (mean between 5-7) shows the prototypes are satisfied for all users. The descriptive statistics Mean and Standard Deviation for prototype evaluation results were shown in Figure 12. The result shows the prototypes developed by using CoDMODS model is satisfied by users in terms of system usability.

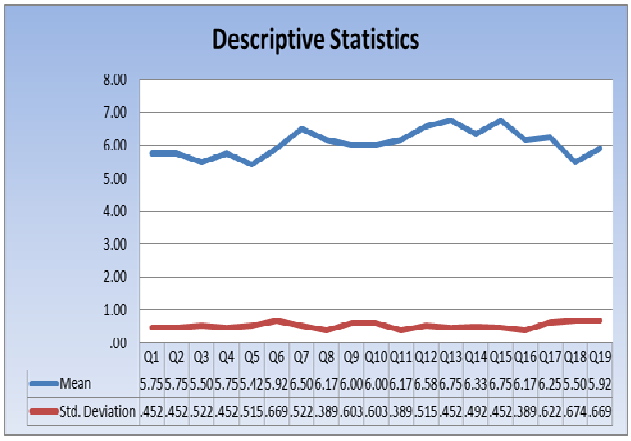

Fig. 12 Descriptive Statistic Mean and Std. Deviation (n=59)

Currently, there are many researchers proposed the methodology and approaches how to develop a data warehouse project. Anyway, the researchers are focused more in developing analytical information for BI applications. Figure 13 below shows the comparison of DW development methodologies for different information type building. Based on the figure, many researchers emphasize to develop DW for analytical information but less focus on operational and tactical information.

Researcher	Research Title	Type of Information
Paim & Castro (2003)	DWARF: An Approach for Requirement Definition and Management of Data Warehouse System [17].	Analytical
Dale (2004)	Defining User Requirement for a Large Corporate Data Warehouse : An Experiental Case Study [18].	Analytical
Prakash & Gosain (2003)	Requirement Driven Data Warehouse Development [19].	Analytical
Giorgni *et al.* (2005)	Goal-Oriented Requirement Analysis for Data Warehouse Design [15].	Analytical
Winter & Strauch (2004)	Information Requirement Engineering for Data Warehouse Systems [3].	Analytical
Guo *et al.* (2006)	Triple-Driven Data Modeling Methodology in Data Warehousing: A Case Study [20].	Analytical
Boenlein & Ende (1999)	Deriving Initial Data warehouse Structures from the Conceptual Data Model of the Underlying Operational Information Systems [13].	Analytical
Inmon W.H. (1999)	Building the Operational Data Store [8]	Operational and Tactical
Shahbani & Norshuhada (2011)	**Proposed Model : Conceptual Design Model Using Operational Data Store (CoDMODS) for Business Intelligence Applications**	**Operational and Tactical (Can Support Analytical Information)**

Fig. 13 Comparison of Data Warehouse Development Methodologies for Information Type Building

5. Conclusion

The CoDMODS model proposed in this paper represents a requirement process and ODS design for developing BI system that is focused on ODS function, which support operational and tactical information. The model is divided into organizational requirement and ODS design level requirement. Organizational requirement level focuses on a broad and higher level requirement in an organization while operational data store requirement emphasizes in a more specific subject area. This model also verified and validated using expert review and formative evaluations.

References

[1] B. A. Nuseibeh and S. M. Easterbrook, "Requirements Engineering: A Roadmap", In A. C. W. Finkelstein (ed) "The Future of Software Engineering", IEEE Computer Society Press, 2000.

[2] Software Engineering Body of Knowledge, SWEBOK, 2004.

[3] Winter, R., & Strauch, B. A Method for Demand-driven Information Requirements

[4] Analysis in Data Warehousing Projects, Paper presented at the 36th Hawaii International Conference on Systems Sciences, 2003.

[5] Wu, J., What is Business Intelligence? Retrieved 21 April, 2009, from http://www.dmreview.com/article_sub.cfm?articleId=1924, 2000.

[6] Negash, S., & Gray, P. Business Intelligence. Paper presented at the Ninth Americas Conference on Information System (AMCIS 2003), Tampa, Florida, (2003, 4-6 August).

[7] Sylvia B. Encheva & Sharil Tumin. Automated Decision Support System Based on Ordered Sources. Journal of WSEAS Transactions on Information Science and Applications, Vol 5, Issue 4, April 2008, pp-356-364.

[8] Imhoff, C. Crystal Clear Customer: The Role of the Operational Data Store. Retrieved 21 April, 2009, from http://www.intelsols.com/documents/Imhoff_10-02.pdf, 2002.

[9] Inmon, W. H. Building the Operational Data Store (2nd ed.): John Wiley & Sons, Inc., 1999.

[10] Kimball, R., and Ross, M. The Data Warehouse Toolkit:The Complete Guide to Dimensional Modeling (2nd ed.): John Wiley & Sons, Inc., 2002.

[11] Mundy J., Thornthwaite W. and Kimball, R. The Microsoft Data Warehouse Toolkit:With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset (2nd ed.): John Wiley & Sons, Inc., 2011.

[12] Carlo. D.A, Ezio L. and Filippo.T., Design and Implementation of National Data Warehouse. Proceeding of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, Feb 15-17, 2006, pp. 342-347.

[13] List B., Bruckner R., Machaczek K., and Schiefer, J. A Comparison of data warehouse development methodologies: Case study of the process warehouse. In Proc. DEXA, 2002.
[14] Boehnlein, M., & Ende, A. U.-v. Deriving Initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems, 1999.
[15] Moody, D. L., & Kortink, M. A. R. From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. Paper presented at the International Workshop on Design and Management of Data Warehouses (DMDW'2000), Stockholm, Sweden, 2000, 5-6 June.
[16] Giorgini, P., Rizzi, S., & Garzetti, M. Goal-Oriented Requirement Analysis for Data Warehouse Design. Paper presented at the ACM Eighth International Workshop on Data Warehousing and OLAP (DOLAP'05), Bremen, Germany, 2005.
[17] Lewis, J. R. (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 7(1), 57-78.
[18] Paim, F. R. S., & Castro, J. F. B. (2003). DWARF: An Approach for Requirements Definition and Management of Data Warehouse Systems. Paper presented at the 11th IEEE International Requirements Engineering Conference (RE'03).
[19] Dale, M. (2004, 6-7 December). Defining User Requirements for a Large Corporate Data Warehouse: An Experiential Case Study. Paper presented at the 9th Australian Workshop on Requirements Engineering (AWRE'04), Adelaide, Australia.
[20] Prakash, N., & Gosain, A. (2003, 16 - 20 June). Requirements Driven Data Warehouse Development. Paper presented at the The 15th Conference on Advanced Information Systems Engineering (CAiSE '03), Klagenfurt Velden, Austria.
[21] Guo, Y., Tang, S., Tong, Y., & Yang, D. (2006, 10 November). Triple-driven data modeling methodology in data warehousing: a case study. Paper presented at the DOLAP '06: 9th ACM international workshop on Data warehousing and OLAP, Arlington, Virginia, USA.

Muhamad Shahbani received the B.Sc. Computer and M.Sc (IT) from, Universiti Putra Malaysia and Universiti Utara Malaysia in 1990 and 1999, respectively. Currently, he is Ph.D postgraduate student in Universiti Utara Malaysia. After working as an analyst programmer and system analyst (from 1990-2000) in private and government sector and a lecturer (from 2000 -2009), he has been a senior lecturer in IT Department, Universiti Utara Malaysia since 2009. His research interest includes software engineering, requirement engineering, web applications, information requirement, data warehouse and business intelligence.

Norshuhada Shiratuddin educated in the United Kingdom: PhD Computer & Information Sciences (1999-2002) from University of Strathclyde, Glasgow, Scotland, MSc. Information Technology (1994-1995), University of Nottingham, UK. and BSc. Mathematics, Statistics & OR (1990-1992) from University of Manchester Inst. of Science and Technology, UMIST, UK. Currently, she is a Professor at College of Arts and Sciences, Universiti Utara Malaysia. She has been the Applied Science Chair (College of Arts & Sciences, UUM, January, 2008 – now), Deputy Dean (Research & Graduate Studies, Jan 2006 – Dec 2007), Faculty of IT, and Program Coordinator (Graduate Studies, 2003 – 2005). Her research interests include software engineering, mobile applications, e-books, e-commerce, Internet and mobile applications and multimedia development.