

Data Warehouse Requirements Analysis Framework: Business-Object Based Approach

Anirban Sarkar

Department of Computer Applications
National Institute of Technology, Durgapur
West Bengal, India

Abstract—Detailed requirements analysis plays a key role towards the design of successful Data Warehouse (DW) system. The requirements analysis specifications are used as the prime input for the construction of conceptual level multidimensional data model. This paper has proposed a Business Object based requirements analysis framework for DW system which is supported with abstraction mechanism and reuse capability. It also facilitate the stepwise mapping of requirements descriptions into high level design components of graph semantic based conceptual level object oriented multidimensional data model. The proposed framework starts with the identification of the analytical requirements using business process driven approach and finally refine the requirements in further detail to map into the conceptual level DW design model using either Demand-driven or Mixed-driven approach for DW requirements analysis.

Keywords- Requirements analysis; Business objects; Conceptual Model; Graph Data Model; Data Warehouses.

I. INTRODUCTION

Complex, online and multidimensional analysis of data is done by fetching just-in-time information from subjective, integrated, consolidated, non-volatile, historical collection of data. Data Warehouse (DW) and On Line Analytical Processing (OLAP) in conjunction with multidimensional database are typically used for such analysis. DW facilitates data navigation, analysis, and business oriented visualization of data using multidimensional cube and OLAP query processing. DW systems are used mainly by decision makers to analyze the status and the development of an organization [1], based on large amounts of data integrated from heterogeneous sources into a multidimensional data model. As in other information systems, requirement analysis plays a key role within DW system development to reduce the risk of failure.

DW projects are long-term projects, so it is very difficult to anticipate future requirements for the decision-making process in the scenario of evolving business processes over time [2]. Further, information requirements for DW systems are difficult to specify at the early stage because decision processes are flexibly structured and shared across the different sectors of the large organization. In this scenario, requirement analysis for DW project must support reuse of domain level abstractions and step wise refinement mechanisms strictly for mapping the requirements in high level design. Hence, for the DW project, requirements analysis must start focusing with early requirements analysis and further it should move on detailed

requirements analysis. Analyzing early requirements will significantly decrease the possibility of misunderstanding the user's requests and consequently reduce the risk of failure for the DW project.

Several studies [2, 3, 4] indicate that improper analysis of user requirements or avoidance of requirements analysis phase leads to unsuccessful design of DW System. Further they recommended that, there must be a dedicated phase of requirement analysis for the purpose of DW system design. The primary deliverables of the DW requirements analysis phase is conceptual level multidimensional data model [5]. In this course, the schemata of the available operational data sources are also compared with the user driven information requirements. Indeed, the approaches to DW design are usually classified in three categories:

a) *Supply-driven / Data-driven*: In these approaches [6, 7, 8], the DW design starts from a detailed analysis of the operational data sources. In order to determine the structure of conceptual multidimensional data model the user requirements have less impact in this approach.

b) *Demand-driven / Requirements-driven*: These approaches [2, 9, 10] start from determining the information requirements of business users or stakeholders. The problem of mapping these requirements onto the available data sources is faced only at the late design phase of DW system. This approach is the only alternative whenever a deep analysis of data sources is unfeasible, or data sources reside on legacy systems whose inspection and normalization is not recommendable. However, in this case, conceptual level multidimensional data model design can be directly based on mapping of requirements.

c) *Mixed-Driven (Supply/Demand)*: In these approaches [9, 11, 12], requirements analysis and source inspections are carried out in parallel. The Supply-driven and mixed framework is recommended when source schemata are well known, and their size and complexity are substantial.

Majority of these approaches mainly have focused on requirement analysis and does not provide any definite guidelines to move from requirements model to high level conceptual design. In [9] goal oriented approaches has been described to support both demand driven and mixed framework for DW system requirements analysis. It comprehensively supports early requirement analysis phase for DW system and

provides a mechanism to map the requirements into the conceptual model. But author has not clearly explained about the detailed requirements analysis for DW system.

In this paper, a requirement analysis framework for DW system has been proposed based on concept of Common Business Objects [13]. A business object (BO) captures information about a real world (business) concept, operations on that concept, constraints on those operations, and relationships between those elements and other business concepts. The set of related BOs express an abstract view of the business's "real world". The advantage of using this concept is that, the set of BOs can be reusable in the context of business domain and it ensures the resultant system will be scalable, reliable, secure and interoperable [13].

In the perspective of requirements analysis for DW system, our proposed framework consist of three phases, namely, (i) Early Requirements Analysis Phase, (ii) Detailed Requirements Analysis Phase and (iii) Mapping Phase. The early requirements analysis phase allows for modeling and analyzing the contextual setting of the business domain, in which the DW will operate. In detailed requirements analysis phase, the early requirements specifications are refined with the structural, functional and nonfunctional features of the domain that is relevant to the participants and their role related to the analytical tasks. Moreover, with the aim of the user centric requirements analysis this phase provides a guideline to identification of materialized views [19] for the target DW system. The refinement process is largely influenced by the concepts of Feature Oriented Domain Analysis (FODA) [16]. The mapping phase used to map the DW requirement specifications to the conceptual design model and can starts just after the early requirements analysis phase. The framework supports both demand driven and mixed demand / supply driven requirements analysis approaches for DW system. Also it is supported with abstraction mechanism; reuse capability and mapping facility for requirements descriptions into high level design components.

In the context, the preliminary version of this work has been published in [17]. A BO based requirements analysis framework for generic large scale information system also has been published in [18]. Moreover, the proposed framework has used the graph semantic based conceptual level object oriented multidimensional data model proposed in [14,15] for its mapping phase. In this context, Graph Object Oriented Multidimensional Data (GOOMD) model provides a novel graph based semantic and simple but powerful algebra to conceptualize the multidimensional data visualization and operational model for OLAP, based on object oriented paradigm.

II. GOOMD MODEL WITH EXAMPLE

In this section, we will summarize the basic concepts of GOOMD model [14, 15]. The GOOMD model is the core of the comprehensive object oriented model of a DW containing all the details that are necessary to specify a data cube, a description of the dimensions, the classification hierarchies, a description fact and measures.

The GOOMD model allows the entire multidimensional database to be viewed as a Graph (V, E) in layered organization. At the lowest layer, each vertex represents an occurrence of an attribute or measure, e.g. product name, day, customer city etc. A set of vertices semantically related is grouped together to construct an Elementary Semantic Group (ESG). So an ESG is a set of all possible instances for a particular attribute or measure. On next, several related ESGs are group together to form a Contextual Semantic Group (CSG) – the constructs to represent any context of business analysis. A set of vertices of any CSG those determine the other vertices of the CSG, is called Determinant Vertices of said CSG. The most inner layer of CSG is the construct of highest level of granularity of fact in Multidimensional database formation.

This layered structure may be further organized by combination of two or more CSGs as well as ESGs to represent next upper level layers and to achieve further lower level granularity of contextual data. From the topmost layer the entire database appears to be a graph with CSGs as vertices and edges between CSGs as the association amongst them. Dimensional Semantic Group (DSG) is a type of CSG to represent a dimension member, which is an encapsulation of one or more ESGs along with extension and / or composition of one or more constituent DSGs. Fact Semantic Group (FSG) is a type of CSG to represent a fact, which is an inheritance of all related DSGs and a set of ESG defined on measures. Two types of edges has been used in GOOMD model, (i) directed edges from DSGs to FSG or constituent DSG to determinant vertex of parent DSG to represent the one – to – many associations and (ii) undirected edges between constituent ESGs and determinant ESGs to represent the association within the members of any CSG.

Since, In order to materialize the cube, one must ascribe values to various measures along all dimensions and can be created from FSG. The cube will also obey a functional constraint $f: D_1 \times D_2 \times \dots \times D_p \rightarrow MI$. Where any D_i is a member of all related top level DSGs and MI is instances of set of measures M. For schema containing multiple FSGs with shared DSGs, the DSG set $\{D_1, D_2, \dots D_p\}$ are the common set of DSGs for all FSGs of the schema.

Let consider an example, based on Sales Application with Sales Amount as measure and with four dimensions – Customer, Model, Time and Location. Model, Time and Location dimensions have upper level hierarchies as Product, QTR and Region respectively. Then in the notation of GOOMD model, there will be four DSGs $DSales = \{DCustomer, DModel, DLocation, DTime\}$ with hierarchies. Each DSG will be comprised of either a set of ESGs $EX \subseteq ESales$ or a combined set of ESGs and DSGs. As described above, the lower layer DSG will be comprised of ESGs only. The Product DSG $DProduct$ is comprised of only ESGs like EP_ID, EP_NAME and EP_DESC and will be represented as the inner layer of the graph. In the example $DModel$ DSG is an extension of $DProduct$ DSG as well as encapsulation of EM_ID and EM_NAME . The $DProduct$ and $DModel$ DSG graphically can be represented as Figure 1. The FSG for the database can be described as $FSales = \{DET(DCustomer), DET(DModel), DET(DLocation), DET(DTime), EAMOUNT\}$. Where

EAMOUNT is the ESGs defined on the measure. The schema from the topmost layer has shown in Figure 2.

GOOMD model also provides algebra of OLAP operators those will operate on different semantic groups. The dSelect (π) operator is an atomic operator and will extract vertices from any CSG depending on some predicate P. The Retrieve (σ) operator extracts vertices from any Cube using some constraint over one or more dimensions or measures. The Retrieve operator is helpful to realize slice and dice operation of OLAP. The Aggregation (α and $+\alpha$) operators perform aggregation on Cube data based on the relational aggregation function like SUM, AVG, MAX etc. on one or more dimensions. Aggregation operators are helpful to realize the roll-up and drill down operations of OLAP. GOOMD model also provides the definitions of the operators like Union (\cup), Intersection (\cap), Difference ($-$), Cartesian Product (\times) and Join (\Join), which are

operated on any CSG or Cube.

III. PROPOSED REQUIREMENTS ANALYSIS FRAMEWORK

The core of the proposed Requirements Analysis Framework for DW system is Business Object Model. A business object (BO) is a conceptual object that has been specified for the purpose of directly describing and representing a business concept with a well-defined boundary and identity. A BO encapsulates identity, domain specific features and behavioral features.

The domain specific features are comprised of structural, functional and nonfunctional features of the domain of interest. The behavior is given by the activities that the business object is capable of performing to fulfill its purpose including the collaboration with other BOs. The primary issue is capturing business semantics (for processes, events and stakeholders) having a common idea or concept that is usable by different parts of a business and by different participants of that business. In the context of DW system, Business Object Modeling is an abstraction technique that consists of identifying the set of concepts and their contexts that belong to some business domain. Such model is comprised of set of BOs to characterize the set of processes and activities of that specific business.

A. Components of Proposed Framework

The part of business object model taxonomies [13], relevant to DW system requirements analysis are as follows,

a) *Business Object (BO)*: This is an abstraction that describes a concept of interest in the business itself and capable of being specialized through inheritance mechanism. A BO is the super type of all objects that represent business concepts either entities or activities involved in such specific business domain.

b) *Entity BO*: This is a specialize form of BO that describes basic business concepts those are engaged in the conduct of business processes. For example stakeholders, products, locations etc. Two entity BOs based on some specific role can collaborate with each other in the context of some business process.

c) *Process BO*: This is a specialized form of BO that describes a business process or workflow and is comprised of a specified collection of Entity BOs, a pattern of interactions and business events. For example, order fulfillment, procurement, payment etc. Interactions represent and implement activities. The entity BO instances are the actors with specific role and subjects of action. A Process BO can be further refined as collection of related sub process BOs.

d) *Event BO*: This is a specialized form of BO that describes a business event, which may trigger and result from interactions between entity BOs in the context of a process BO. For example, inventory threshold, account overdrawn, end of fiscal year etc. Event BOs are used to capture the business constraints on the interactions.

Besides the above described taxonomies, DW context requires some new taxonomy to be introduced and are as follows,

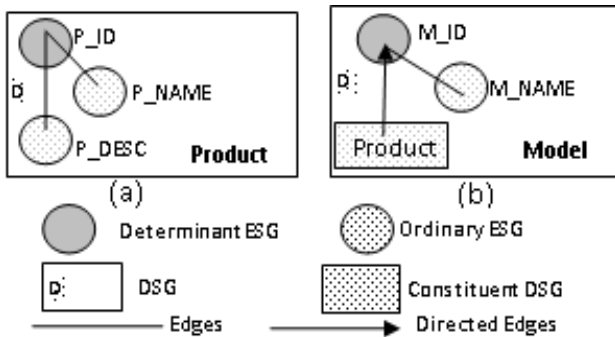


Figure 1. (a) Lowest Level DSG, (b) Higher Level DSG

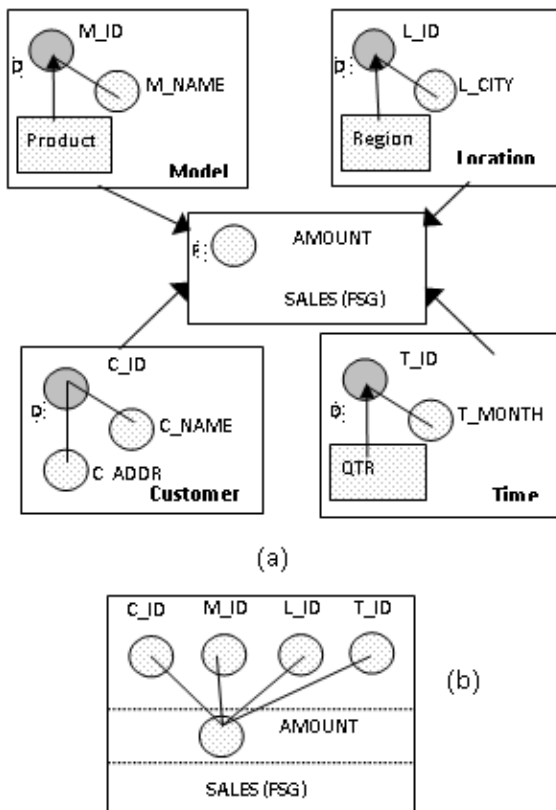


Figure 2. (a) Schema for Sales Application in GOOMD Model
(b) SALES FSG construct after inheritance

a) *Measure Attribute*: This is an attribute that describes a quantitative aspect of business process that is relevant for decision making.

b) *Fact BO*: This is a specialized form of Process BO to capture the concept of subject of analysis specific to some business process. It encapsulates the Measure Attributes along with other features.

c) *Dimension BO*: This is a specialized form of Entity BO to capture the concept of parameters over which the Fact BO will be analyzed using other Entity BO having relevant role. Any behavioral features will not be included in the Dimension BO. Further a specialized Dimension BO can be formed from generalized one in the proposed framework.

d) *Relations*: several relationship types have been used in the proposed framework like, (i) Encapsulation, (ii) Inheritance, (iii) Collaboration and (iv) Interaction.

The graphical notations for above taxonomies have been summarized in Table 1.

TABLE I. GRAPHICAL NOTATIONS FOR BO-BASED REQUIREMENTS ANALYSIS FRAMEWORK

Taxonomy	Graphical Notation
Entity BO	
Process BO	
Event BO	
Measure Attribute	
Fact BO	
Dimension BO	
Encapsulation	
Inheritance	
Collaboration	
Interaction Relation	

B. Early Requirements Analysis Phase

The focus of early requirements analysis phase is to analyze the target business domain. It includes identification along with high level of abstraction of the business processes, different stakeholders, the interactions and collaborations between them, and the events relevant for analytical task specific to the business domain. This phase is important to understand how a business is perceived by its stakeholders mainly decision makers as a set of related business processes and events. This phase is also the basis to identify the possible analytical requirements of decision maker or other stakeholders from the business process driven approach. The phase consists of five steps and are as follows,

a) *Identification of Process and Entity Level Business Objects*: In the first step the relevant business processes with their context, possible stakeholders related to those business

processes and interactions between them are represented with high level of abstraction. The business processes and stakeholders can be represented in the proposed framework using Process BOs and Entity BOs respectively. Interactions can be represented between Process BOs and Entity BOs. Interactions represent and implement the activities. In between one Process BO and one Entity BO, more than one activity may exist and this will result in more than one interaction respectively. The interactions between the set of Entity BOs and Process BOs of some specific business can be expressed using business domain level Interaction Diagram. The Interaction Diagram of each interested Process BO can be achieved from business domain level Interaction Diagram.

For example in a Retail Organization, entire business is comprised of several business processes like, (i) Procurements – for procuring the products for sale, (ii) Sales – to handle the customer orders and to sale the products as per order, and (iii) Accounting – to handle the bills, order payment, salaries etc. Several stakeholders may be involved with these business processes, e.g. Sales Manager and Customer may interact with the Sales and Accounting processes with activities like Place Order, Receive Product, Payment, Raise Bill, Bill adjustment etc. Now, the business process will be mapped into the Process BOs and the Stakeholders will be mapped into the Entity BOs. Several other Entity BOs can be involved with each Process BO without having any specific interactions, but they are use to supply important information on the interactions between the business processes and related stakeholders. For example Time, Location, Product, Transaction Type etc. The BOs and interactions relevant to the example can be captured using Interaction Diagram as shown in Figure 3. In the diagram Interactions are labeled with the activities and with direction of initiator and receiver of the interaction. The important point to note that, in this step all Process BOs and related Entity BOs are the representation of high level abstraction for the set of involved business process.

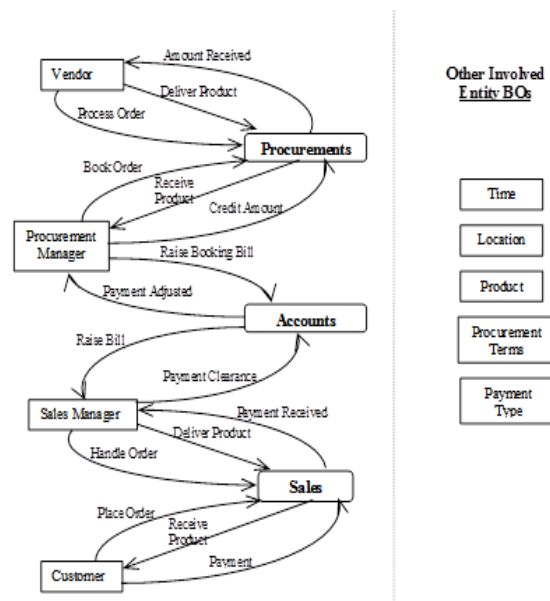


Figure 3. Business Domain Level Interaction Diagram

Now let, the decision makers' requirements include analyze of the Sales of the product. So the Sales Process BO is one of the interesting candidates of target DW system. Two stakeholders may be involved with the process and represented as Entity BOs namely, Sales Manager and Customer. The Interaction Diagram related to Sales Process BO has been shown in Figure 4.

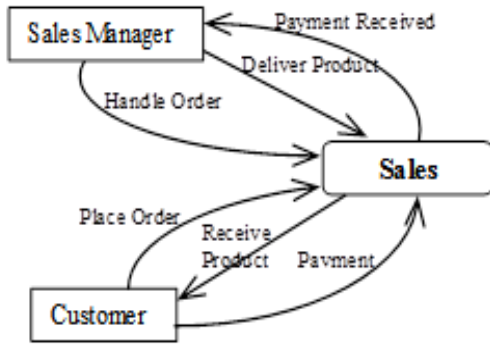


Figure 4. Interaction Diagram for Sales Process BO

b) *Identification of Collaborations:* In this step the collaborations are identified between Entity BOs in the context of some interested Process BO relevant for the purpose of analytical processing. Collaborations are occurred based on some specific roles played by participant Entity BOs. In software engineering term role can be defined as separation of concerns i.e. separation of behavioral characteristic of some Entity BO. The purpose of collaboration is to fulfill some set of activities concerned to the specific Process BO, with which the participant Entity BOs can interact. The roles and collaborations in the context of some Process BO can be identified from Interaction Diagram. This will lead towards developing the Collaboration Diagram in the context of some specific Process BO. Further, for a specific collaboration one can model the activities performed in the context of some Process BO, which will result the Collaboration Interaction Network.

In the above example Sales Manager can play the role as Supplier to fulfill the activities like Handle Order and Deliver Product and also can play the role as Payee for the purpose of activity like Payment Received. Similarly two roles like Client and Payer can be defined for Entity BO Customer. The Collaboration Diagram and Collaboration Interaction Networks are represented in Figure 5.

c) *Identification of Measure Attributes and Fact Business Objects:* Measure attributes are the quantitative aspects of some Process BO and are relevant for performing analytical tasks on that specific BO. Measure attributes related to each interested Process BO are identified in this step. This step is also the basis of Fact BO construction. Each Process BO relevant to analytical task can be specialized into a Fact BO by encapsulating the related measure attributes.

For example in the running example of Retail Organization for the Sales Process BO two possible measure attributes may

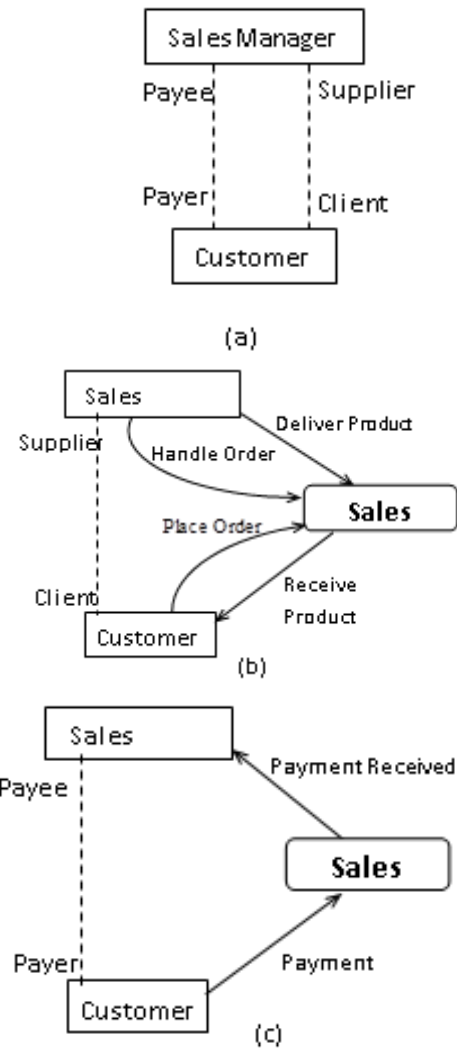


Figure 5. Collaboration Diagram. (b) Collaboration Interaction Network for Supplier / Client Collaboration, (c) Collaboration Interaction Network for Payee / Payer Collaboration

be Quantity and Amount of sales. A Fact BO for Sales can be described from the Process BO Sales by encapsulating the measure attributes Quantity and Amount.

d) *Identification of Dimension Business Objects:* In this step Entity BOs other than representing stakeholders are identified in the context of some specific Process BO. The set of Entity BOs identified in this step, in general captured the parameters over which the measure attributes related to the specific Process BO are dependent and can be analyzed. Both the set of Entity BOs that captured the stakeholders and other Entity BOs that provides the important information to the Interactions are used to describe the abstraction of Dimension BO in the context of some Fact BO. The set of related Entity BOs which are used to describe the dimension abstraction can be identified from the business domain level interaction diagram of the first step of this phase.

In the Retail Organization example in the context of Sales Fact BO, the possible Dimension BOs can be Time, Location,

Product and Payment Type including Sales Manager and Customer.

e) *Identification of Event Business Objects:* In this step the Event BOs are identified in the context of some specific Process BO and related Entity BOs. Identification of Event BOs is important to understand the rationale for recording the process BO instances in the DW system for the analytical task. Entity BOs are also important to realize how the specific Process BO will react on some interaction. In the system level view, an Event BO can be realized as a trigger which may fire as a result of specific set of interactions between related Entity BOs and is important in the creation of the DW system.

For example, in the Sales Process BO the possible Event BOs may be Order Processed and Full Payment Received. This step provides the basis of mapping the instances in DW system.

C. Detailed Requirements Analysis Phase

The focus of detailed requirements analysis phase is to refine the Fact BOs and related Dimension BOs identified in early requirements analysis phase to satisfy the analytical requirements of decision makers or other stakeholders. The Fact BOs and Dimension BOs are refined by adding the domain specific features like structural, functional and nonfunctional feature to capture the activities, interactions, collaborations and other stakeholders' requirements. Further this phase provide the guidelines to select materialized views for the target DW system. This phase is also capable to implement the Supply-Driven part of the framework by further refining the identified Fact BOs and Dimension BOs with the comparison with existing operational schemas. The refined Fact BOs and Dimension BOs produced in this case will be more realistic and appropriate for mixed analysis approach of DW system requirements analysis. But it is important to note that the detailed analysis of source operational schemas must be available a prior for this purpose. This phase consists of *four steps* and are as follows,

a) *Refinement of Fact Business Objects:* In this step the Process BOs identified in early requirements phase are refined through two tasks. In the first task, Process BOs are refined as possible collection of related sub Process BOs. The measure attributes identified for top level Process BOs are placed with appropriate sub Process BOs. This task may be iterative in nature to achieve further details sub Process BOs. The Interaction Diagram concerned to the specific Process BO acquired from earlier phase is extended accordingly.

Further the refined Fact BOs can be realized from the refined Process BOs to satisfy stakeholders' expectations. For example, the Sales Process BO may be thought as the composition of two sub-Process BOs Order Processing and Payment Processing. Measure attributes Quantity and Amount can be associated with Order Processing and Payment Processing Process BOs respectively. Henceforth two possible Fact BOs can be realized based on the sub-Process BOs. In fact, each of them encapsulates at least one measure attribute. It will result Extended Interaction Diagram for the Sales Process BO and has been represented in Figure 6.

In the second task, each refined Process BO can be further refined by adding the domain level feature like structural,

functional and nonfunctional along with the constraint specifications. The features are the attributes of the system which directly affect the stakeholders. Structural features describe the object level properties of some BO and may also include constraint specification over some features. Functional features describe the operational capabilities of some Process BO. In the context of DW system requirements analysis, functional features of any Process BO include the set of possible analytical operations. Thus functional features are the basis of identification of the possible set of OLAP operations those are required to perform by the decision makers, over the related Fact BO. The prime focus of Nonfunctional features of some Process BO is related with the expected QoS requirements of stakeholders concerned to the specific BO. This may include the features for Security, Performance, Usability etc., specific to some Process BO. Nonfunctional features may be optional for certain Process BO. The features of some Process BO can be represented using Feature Tree Diagram.

Also, Process BO may have specialized Process BOs.

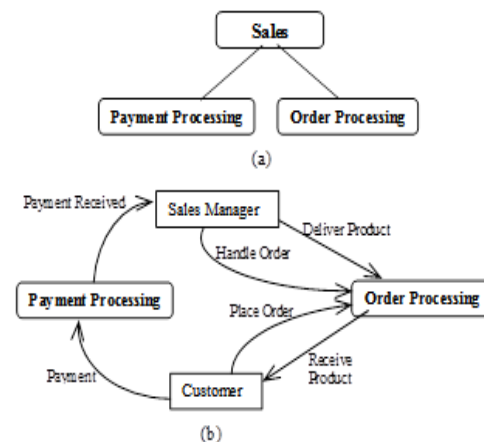


Figure 6. (a) Refined Sales Process BO, (b) Extended Interaction Diagram.

Henceforth a Process BO can contain both basic and derived features. Besides representing the features of some business process as per the stakeholders' perspective, a Feature Tree is capable to represent the logical grouping (AND, OR and EXOR grouping) of the features to satisfy the decision makers or other stakeholders need. Several features of the specific Process BO are the basis of the set of attributes that can be analyzed along with the parameter over which those can be analyzed, in the context of the related Fact BO. Henceforth, the feature tree exhibits the set of features using which the related Fact BO can be associated with concerned Dimension BOs for the formation of conceptual level multidimensional schema for the DW system. The concerned high level Dimension BOs are already identified in the first phase of the framework. Further the feature tree may be accompanied with set of composition rules to express the existing semantics between the subset of features. An example of partial Feature tree for the Process BO, Payment Processing has been shown in Figure 7.

b) *Refinement of Dimension Business Objects:* In this step, the set of Dimension BOs identified in the early

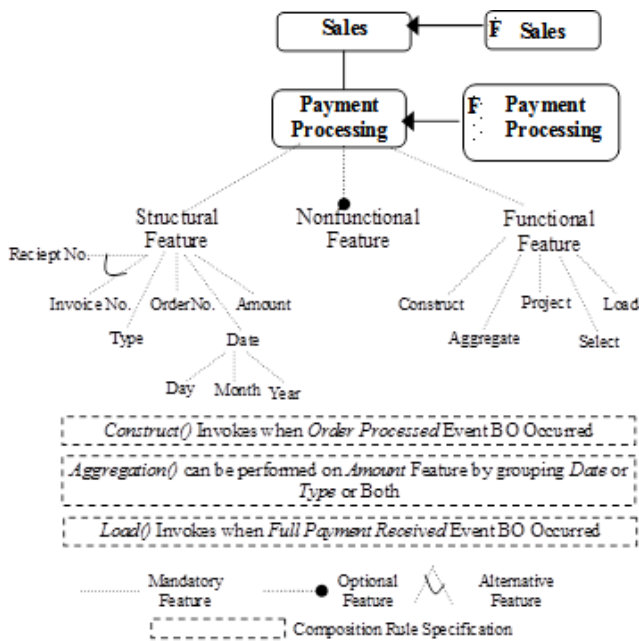


Figure 7. Partial Feature Tree for the Process BO Payment Processing

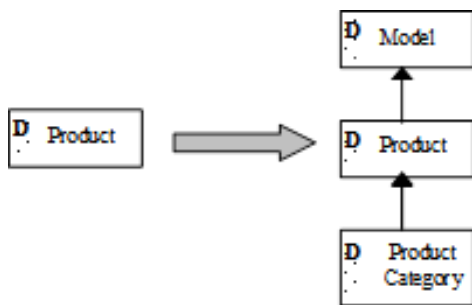


Figure 8. Vertical Refinement of Dimension BO Product

requirements analysis phase are refined. The refinement process involves two tasks. Firstly, the different levels of granularity required for the multidimensional information are identified for some specific business processes. Basis on that, the set of related Dimension BOs need to refine starting from the lowest level granularity to the highest one. The Dimension BO with highest level granularity will exhibit lowest level abstraction in Dimension BO instances. The Dimension BOs with highest level granularity in the context of some Fact BO already have been identified in early requirements analysis phase. This task facilitates the vertical refinement of Dimension BOs and focus on the creation of dimension hierarchy in the context of some fact for the DW system. As an example, the possible vertical refinement of Product Dimension BO in the context of Sales Fact BO has been represented in Figure 8.

In the second task, the refined set of Dimension BOs (each of all hierarchies) are further refined by adding the domain level features in the same way as it has been done in case of Process BOs in the previous step of this phase. It will result the Feature Tree diagram for each Dimension BO. This task facilitates the horizontal refinement of Dimension BOs. The

only difference here is, the accompanied composition rules will express the semantics between the subset of features from the specific Dimension BO or in the context of the concerned Fact BO. This will facilitate to realize the different constraint requirements for the DW system.

c) *Compare with Operational Schemas:* This step is to implement the Supply-Driven part to the framework. The refined Fact BOs and Dimension BOs resulted from the last two steps can further be modified by comparing with the existing operational schemas of the organization. To perform this step a prior knowledge of detailed analysis of operational schemas are required. The set of Fact BOs and Dimension BOs can be refined or modified or also can be filtered by navigating the source operational schemas. Several literatures suggest algorithmic approaches mostly based on the stakeholders' requirements using navigation through the path of many-to-one associations from the attributes of facts.

In fact the steps can be performed by drawing the dependencies between each of Fact BOs and related Dimension BOs, and related operational source schemas. The feature tree diagram of each BO on next can be modified as per the related operational schemas features. The steps are as followed,

1) Draw the dependencies between the Measure Attributes of concerned Fact BO and the Appropriate Attribute of existing source Operational Schemas. It is important to note that the Measure Attribute names may be decided by the decision makers or designers and henceforth attribute names may not match.

2) From each of such Operational Schemas identified in previous step, start navigation using the available foreign keys to other associated Operational Schemas. The each of associated set of Operational Schemas is required to map as Dimensional BO. Dependency can be drawn between the Dimension BOs achieved from the previous steps of the framework to the Operational Schemas achieved on navigation.

3) If dependencies have not been drawn from some Dimension BOs then those BOs are purely demanded by the decision makers and can be labeled as "Demanded Dimension BO". On other hand, where dependencies have drawn from some Dimension BOs, those are labeled as "Supplied Dimension BO".

4) For each Supplied Dimension BO, the attributes of the appropriate operational schemas are compared with Feature Tree of the specific Dimension BO. On comparison structural features in the feature tree can be refined or modified as per the attribute details of the related operational schemas.

5) For each operational schema, identified in step (ii) but related Dimension BO does not exist, are required to map as New Dimension BO. The feature tree of new Dimension BO can be drawn from the attributes descriptions of the related operational schemas.

6) The feature tree of each Fact BO can be modified further from the available Dimension BOs and newly achieved Dimension BOs from the step (v).

d) *Identification of Materialized Views:* For user oriented DW requirements engineering, it is also important to analyze that how user will efficiently interact with the DW system to perform the necessary analysis activities. Materialized views are the central issue for the usability of the DW system. DW data are organized multi dimensionally to support OLAP. A DW can be seen as a set of materialized views defined over the source relations. Those views are frequently evaluated by the user queries. The materialized views need to be updated when the source relations change. During the DW analysis and design, the initial materialized view need to be selected to make the user's interactions simple and efficient in terms of accomplishing user analysis objectives. In the proposed requirements engineering framework, the domain boundary has been drawn through identifications of Fact BOs, Dimension BOs, Actor BOs, and interactions between them in the first phase and which have been further refined in this phase. The list of analysis activities may be performed by Actor BOs based on their roles and also Event BOs have been identified in the same phase. Moreover, the feature tree concept explores the constraint requirements for the interest of domain. Based on those identifications, the different materialized views can be identified in this step. In this step the materialized views are used to represent semantically in the context of some Fact BO and in terms of actor along with their roles, analysis activities those may be performed, events those may be occurred, related Dimension BOs involved and the related constraints. Related to one Fact BO, there may exist several materialized views to minimize the views level dependency and to meet the analytical evaluation requirements of the stake holders. Semantically, a materialized view will be represented using View Template. The Interface Template will contain the information of View name, identification, analysis objectives, target Fact BO, Actor BO, roles, related activities, related Dimension BOs to realize the source relations, related Event BOs and related constraints. Any view template is reusable and modifiable through iterative process to accommodate the updatable materialized view.

In case of *Retail Organization*, to interact with the *Sales* Fact BO, one example view template related to *Customer* has been shown in Figure 9.

View Name and ID: Customer View 1					
Target Fact BO: Sales					
Requirements Objective: Analysis of Order Placing of Location XXX					
Target Measures: Quantity					
Actor BO: Customer	Role: Client	Activities: Place Order; Receive Products;	Dimension BOs: Time; Product; Procurements Term ;	Event BOs: Order Status	Constraint: Construct() invoke.

Figure 9. View Template related to Sales Fact BO

IV. REQUIREMENTS ANALYSIS TO CONCEPTUAL MODEL

This section will describe the Mapping Phase of our framework. In this phase the DW requirements specifications achieved from the early requirements analysis phase and

consequently from the detailed requirements analysis phase are mapped into conceptual level multidimensional design model. In this phase, we will map the requirements specifications in the constructs and concepts of GOOMD model described in Section 2. But there is no binding on mapping the proposed requirements framework to any multidimensional conceptual model. This phase can starts just after the early requirements analysis phase and consists of two levels, namely, Early Mapping and Detailed Mapping.

a) *Early Mapping:* Early mapping can be done just after the early requirements analysis phase for DW system. The early requirements analysis phase is business process driven approach and is used to identify the set of Fact BOs and related Dimension BOs relevant to analytical requirements of decision makers and other stakeholders. Each identified Fact BO and encapsulated Measure Attributes will be mapped as Fact Semantic Group (FSG) and Elementary Semantic Group (ESG) for measures respectively as described in GOOMD model concept. Each Dimension BO related to the specific Fact BO can be mapped as Dimension Semantic Group (DSG). On next, each DSG need to be connected with the FSG using Link. The early mapping will yield the topmost layer of the GOOMD model Schema which exhibit high level abstraction.

b) *Detailed Mapping:* In detailed mapping steps, the multidimensional schema achieved from the early mapping can be further refined to deliver the full-fledged conceptual schema for DW system. The topmost layer GOOMD model schema will be further modified in this step, from the specifications available from detailed requirements analysis phase. Firstly, the Dimension BOs of different granularity related to each top level Dimension BO are mapped as separate DSGs and are connected using Link to form the dimension hierarchies. On next, the basic structural features and identity from the Feature Tree of each Dimension BO are mapped as ESG and Determinant ESG respectively. The related ESGs and Determinant ESGs are connected using Association and need to be encapsulated in the specific DSG. Finally, each FSG can be modified from the feature tree of related FO by extracting the relevant features into set of ESGs and by encapsulating those ESGs. This step will exhibit all the inner layers of GOOMD model schema.

The descriptions of possible OLAP operations for the refined GOOMD model schema can be guided from the Functional features of related set of Fact BO and Dimension BOs. The DW System QoS requirements can be realized from nonfunctional features of Fact BO.

V. FEATURES OF PROPOSED REQUIREMENTS FRAMEWORK

The proposed requirements analysis framework for DW system has been drawn from concepts of business object model. Besides the advantages of using the business object concept, the proposed framework facilitates several other features and are as follows,

a) *Process Driven Approach:* The proposed requirements analysis framework for DW system starts with

business process driven approach. The early requirements analysis phase in the framework identify the concerned business processes, events, stakeholders and interactions between them possibly relevant to analytical task. The phase is not biased from perspective of stakeholders' understanding about the system. Rather it ensures the modeling of the business concept in high level of abstraction, how it is exist in reality. So it is more realistic approach.

b) *Object Orientation*: The core of the proposed framework is Business Object model, which supports the general concepts and characteristics of object oriented paradigm. Further, the requirements specification resulted from the proposed framework can be mapped into any object oriented conceptual level multidimensional data model (GOOMD model as an example).

c) *Abstraction*: The proposed requirements analysis framework is capable to represent the different business concepts and stakeholders' requirements in different level of abstraction. Most abstract description of the analytical requirements in the context of DW system is available from the early requirements analysis phase of the framework. Moreover the early mapping, which can be performed just after the early requirements analysis phase, will produce the conceptual multidimensional data model schemas with high level of abstraction. In the detailed requirements framework the early requirements descriptions are refined stepwise to lower the abstraction level (see Table 2).

TABLE II. ABSTRACTION AND REUSE POTENTIAL IN EACH PHASE OF PROPOSED REQUIREMENTS ANALYSIS FRAMEWORK

Base Concepts and Phases for proposed Requirement analysis framework	Level of		
	Abstraction	Reuse Potential	Productivity
Business Object (Base Concept)	High	High	Low
Early Requirements Analysis Phase	↓	↓	↓
Early Mapping Phase			
Detailed Requirements Analysis Phase	↓	↓	↓
Detailed Mapping Phase	Low	Low	High

d) *Reusability*: One of the major advantages of the proposed framework is that, it supports reuse of domain level abstractions and step wise refinement mechanisms for mapping the DW requirements in high level design. This is important because the anticipation of future requirements for the decision-making process is very difficult in large system like DW. The support of abstraction mechanism and feature oriented stepwise refinement of the BO based requirements descriptions in detailed requirements analysis phase enable the capability of reuse (see Table 2) of different types of BOs specific to some business. A new BO can be formed from the existing BO either at high level of abstraction in early requirements analysis phase or by adding new features in

detailed analysis phase. The refinement processes of BOs are iterative in nature.

e) *Support for Multiple Analysis Approaches*: The proposed framework supports both Demand-driven and Mixed analysis approaches for the requirements analysis of DW system. Also it is capable to map the requirements to object oriented multidimensional conceptual model by transforming different detailed BOs into the relevant high level components of the design model.

In the proposed framework, the early requirements analysis phase and subsequent early mapping step are basically business process driven and independent of any specific DW requirements analysis approaches. But in detailed requirements analysis, the step for comparison of refined Fact BO and Dimension BO with the operational schemas [subsection 3.C] is to support supply-driven part in the proposed framework. By omitting this step from the detailed requirements analysis phase, the proposed framework is fully compatible with *Demand-Driven* DW requirements approach. But with the presence of that step, the resultant GOOMD model schema from the detailed mapping phase from refined Process BOs and related Dimension BOs, are largely influenced by the source operational schemas specifications. In that case the proposed framework will support *Mixed* Analysis approach towards DW requirements analysis.

VI. CONCLUSION

Requirements analysis plays a key role within DW system development with the aim to reduce the risk of failure. A good DW design method should be preceded by the requirements elicitation and their analysis methodologies by considering both user requirements and operational data sources for data warehouse development. For the purpose, the Business Object based requirements analysis framework for DW system has been devised. The framework is comprised of three phases namely, Early Requirements Analysis phase, Detailed Requirements Analysis phase and Mapping phase. It starts from understanding the set of business processes, events and stakeholders in terms of set of well-defined BOs concerned to some business domain in which DW system will operate. In the framework, several requirements modeling elements like Fact BO, Dimension BO, Measure attributes, Interaction Diagram, Collaboration Diagram, Interaction Collaboration Network, Feature Tree etc. have been described to express different business concepts of the domain, relevant to DW system and in the real business scenario. Finally the framework results the mapping of DW requirements descriptions in high level design components of conceptual level object oriented multidimensional data model. The proposed framework supports abstraction mechanism and reuse of different well defined elements those have been used to realize the different business concepts of the domain and useful for analytical task. These features enable the framework to be used efficiently in the evolving business processes over time. Further the framework supports both demand-driven and mixed analysis approach of DW requirements analysis.

Future work will include developing of a prototype tool in the support of the proposed DW requirements analysis framework. Moreover, quality evaluation of the proposed

requirements analysis framework also is a prime objective of the future work.

REFERENCES

- [1] R. Kimball, M. Ross, "The Data Warehouse Toolkit", Book Wiley & Sons (2002).
- [2] R. Winter and B. Strauch, "A method for demand-driven information requirements analysis in data warehousing projects", In Proc. HICSS, PP 1359–1365, 2003.
- [3] F. R. S. Paim and J. B. Castro, "DWARF: An approach for requirements definition and management of data warehouse systems", Proc. of Int. Conf. on Requirements Engineering, 2003.
- [4] Robert Winter, Bernhard Strauch, "Information requirements engineering for data warehouse systems", Proceedings of the ACM Symposium on Applied Computing, 2004.
- [5] Rizzi, S., Abell'ò, A., Lechtenb'orger, J., Trujillo, J., "Research in data warehouse modeling and design: dead or alive?" Proc. of the 9th ACM Int. Workshop on Data warehousing and OLAP, PP 3–10, 2006.
- [6] M. Golfarelli, D. Maio, S. Rizzi, "The dimensional fact model: A conceptual model for data warehouses", Intl' Journal of Cooperative Information Systems, Vol. 7(2-3), PP 215–247, 1998.
- [7] B. H'usemann, J. Lechtenb'orger, and G. Vossen. "Conceptual data warehouse design", Proc. 2nd Int. Workshop on Design and Management of Data Warehouse, PP 3–9, 2000.
- [8] D. Moody and M. Kortink, "From enterprise models to dimensional models: A methodology for data warehouse and data mart design", Proc. 2nd Int. Workshop on Design and Management of Data Warehouse, 2000.
- [9] Paolo Giorgini, Stefano Rizzi, Maddalena Garzetti, "GRAnD: A goal-oriented approach to requirement analysis in data warehouses", Decision Support Systems, Vol. 45(1), PP 4-21, 2008.
- [10] N. Prakash and A. Gosain, "Requirements Driven Data Warehouse Development", Proc. of CAiSE Short Paper, 2003.
- [11] Jose-Norberto Mazón, Juan Trujillo, Jens Lechtenb'orger, "Reconciling Requirement-driven Data warehouses with Data Sources via Multidimensional Normal Forms", Data & Knowledge Engineering, Vol. 63(3), PP 725-751, 2007.
- [12] Jose-Norberto Mazón and Juan Trujillo, "An MDA approach for the development of data warehouses", Decision Support Systems, Vol. 45(1), PP 41-58, 2008.
- [13] OMG, Business Object DTF – Common Business Objects, OMG Document bom/97-11-11, ftp://ftp.omg.org/pub/docs/bom/97-11-11.pdf, 1997.
- [14] A. Sarkar, S. Bhattacharya, "The Graph Object Oriented Multidimensional Data Model: A Conceptual Perspective", 16th Int. Conf. on Software Engineering and Data Engineering (SEDE 2007), PP 165 – 170, 2007.
- [15] A. Sarkar, S. Choudhury, N. Chaki, S. Bhattacharya, "Conceptual Level Design of Object Oriented Data Warehouse: Graph Semantic Based Model", International Journal of Computer Science (INFOCOMP), Vol. 8(4), PP 60 – 70, December 2009.
- [16] K. C. Kang, S. G. Cohen, J. A. Hess, W. E. Novak, A. Spencer Peterson, "Feature-Oriented Domain Analysis (FODA) Feasibility Study", Technical Report, Software Engineering Institute, Carnegie Mellon University (USA), 1990.
- [17] A. Sarkar, S. Choudhury, N. Chaki, S. Bhattacharya, "Business-Object Oriented Requirements Analysis Framework for Data Warehouses", 22nd International Conference on Software Engineering and Knowledge Engineering (SEKE 2010), PP 34 – 37, July 2010.
- [18] A. Sarkar, N. C. Debnath, "Business Object Oriented Requirements Analysis for Large Scale Information System", 20th International Conference on Software Engineering and Data Engineering (SEDE 2011), PP 103 - 108, June 2011.
- [19] A. C. Dhote, M. S. Ali, "Materialized View Selection in Data Warehousing", 4th International Conference on Information Technology (ITNG '07), PP 843 – 847, 2007.

AUTHORS PROFILE



Anirban Sarkar is presently a faculty member in the Department of Computer Applications, National Institute of Technology, Durgapur, India. He received his PhD degree from National Institute of Technology, Durgapur, India in 2010. His areas of research interests are Database Systems and Software Engineering. His total numbers of publications in various international platforms are about 30.