QVT transformation by modeling
From UML Model to MD Model

I.Arrassen
Laboratory for Computer Science Research
Faculty of Sciences
Mohammed First University
524, Oujda, Morocco

R.Sbai
Laboratory of Applied Mathematics and Computer Signal Processing
Superior School of Technology
Mohammed First University
524, Oujda, Morocco

A.Meziane
Laboratory for Computer Science Research
Faculty of Sciences
Mohammed First University
524, Oujda, Morocco

M.Erramdani
Laboratory of Applied Mathematics and Computer Signal Processing
Superior School of Technology
Mohammed First University
524, Oujda, Morocco

Abstract – To provide a complete analysis of the organization, its business and its needs, it is necessary for leaders to have data that help decision making. Data warehouses are designed to meet such needs; they are an analysis and data management technology. This article describes an MDA (Model Driven Architecture) process that we have used to automatically generate the multidimensional schema of data warehouse. This process uses model transformation using several standards such as Unified Modeling Language, Meta-Object Facility, Query View Transformation, Object Constraint Language, ... From the UML model, especially the class diagram, a multidimensional model is generated as an XML file, the transformation is carried out by the QVT (Query View Transformation) language and the OCL (Object Constraint Language) Language. To validate our approach a case study is presented at the end of this work.

Key-Words: Datawarehouse; Model Driven Architecture; Multidimensional Modeling; Meta Model; Transformation rules; Query View Transformation.

I. INTRODUCTION

To support the process of making management decisions, development of data warehouses is important for organizations. According to the definition given by Bill Inmon (1996), data warehouse is a collection of data that is subject-oriented, integrated, time-varying and non-volatile. His ultimate goal is integrating data from all corners of the enterprise in a single directory, from which users can easily find answers to queries, generate reports and perform analysis.

A data warehouse is a management and data analysis technology. On this basis, the establishment of a process of building data warehouse is very important. Through this work, we use a UML class diagram summarizing the activities: requirements expression, analysis and design of information system of the organization. From this diagram, we will generate future objects decision diagram such as facts and dimensions. The decisional diagram will be in the form of a multidimensional schema, in fact, multidimensional modeling is the one that best represents the data warehouse schema.

II. RELATED WORKS

In recent years, several approaches to developing data warehouse have been proposed. In this section, we present a brief overview of some best known approaches.

In [1] several DWH case studies are presented. DWH design is based on using the star schema and its variants (snowflake schema and fact in constellation) using a relational approach: tables, columns, foreign keys and so on. However, this work is considered as a benchmark in the field of DWH, the authors are only interested in the representation of relational DWH, and they regarded any other technology.

In [2], the authors propose a model Fact-Dimension (Dimensional-Fact Model DFM), in which they define a special notation for the conceptual model of the DWH. They also show how to derive a DWH schema from the data sources described by the entity-relationship diagram.

A goal-oriented approach has been added to DFM in [3]. This approach does not consider important aspects such as the ETL (Extract Transform Load) process. In addition, the authors consider that relational schemas of data exist, which is
not always true, in addition, the use of special notations makes it difficult to apply this approach.

In [4], a model for the design of the DWH was proposed: YAM2. It is an object oriented model that uses UML notation to represent the multidimensional structure of data. However, no method that shows how to get the conceptual model is described in the paper.

In [5] the authors propose a multidimensional meta-model extended by generalizing the model heart, based on medical data, this work was done under the MAP project (Personalized Medicine Anticipation). In our opinion this type model is specific to medical data; in addition, the authors do not specify the modeling approach.

In [6] the authors created a decision support system called BIRD integrated data grid Decrypthon which is designed to process data in functional genomics. This system is a specific platform that does not meet all requirements of any DWH.

III. MULTIDIMENSIONAL MODELING

It is a logical design technique that aims to present data in a standard, intuitive way, which allows high performance access.

![Star Schema](image)

Figure 1. Dimensionnel Model Schema

IV. MDA ARCHITECTURE

In November 2000, the OMG, a consortium of over 1,000 companies, initiates the MDA approach [7]. The purpose of this standard is to separate the business logic of the enterprise, from any technical platform. It is a new way to design applications. Indeed, the technical architecture is unstable and undergoes many changes over time, unlike the business logic. It is therefore easy to separate the two to face the increasing complexity of information systems and high costs of technology migration. This separation then allows the capitalization of software knowledge and the know-how of the company.

Figure 2 below shows schematically the architecture of MDA [7]. It is divided into four layers. The OMG was based on several standards. In the center are the standard UML (Unified Modeling Language) [8], MOF (Meta-Object Facility) [9] and CWM (Common Warehouse Meta-model) [10].
D. OCL (Object Constraint Language)

OCL [11] was developed in 1997 by Jos Warmer (IBM), on the basis of language IBEL (Integrated Business Engineering Language). It was formally incorporated into UML 1.1 in 1999. It is a formal language that is based on the notion of constraint. A constraint is a boolean expression that can be attached to any UML element. It usually indicates a restriction or gives information on a model. The constraints are used in particular to describe the semantics of UML and its various extensions, participating in the definition of profiles.

E. MOF2.0 (MetaObject Facility) QVT

MOF [9] defines the structure that should have any meta-model. A meta-model is a set of meta-classes with meta-associations. MOF2.0 meta-meta-model is unique, and UML 2.0 meta-model is dedicated to the modeling of object-oriented applications. Among the objectives of MOF2.0: capitalize on existing commonalities between UML and MOF-level class diagrams and to explain the differences. The advantage of the MOF is to structure all meta-models in the same way.

QVT (Query / View / Transformation) in the MDA architecture is a standard for model transformations defined by the OMG (Object Management Group) in 2007. It is central to any proposed MDA. It defines a way to transform source models to target models. These source and target models must conform to the MOF meta-model. Specifically, this means that the abstract syntax of QVT must conform to the MOF 2.0 meta-model. QVT defines three specific languages named: Relations, Core Operational/Mapping. These languages are organized in a layered architecture. Relations and Core are declarative languages to two different levels of abstraction. QVT Core is to relational, what is the Java byte code to Java source code. The QVT Operational / Mapping is an imperative language, it provides common constructions in imperative languages (loops, conditions ...).

F. Kinds of MDA transformation.

The MDA identifies several transformations during the development cycle [12]. It is possible to make four different types of transformations:

1) CIM to CIM
2) CIM to PIM
3) PIM to PIM
4) PIM to PSM

There are three approaches in MDA to perform these transformations:

- Approach by programming: using the object-oriented programming languages such as Java, to write computer programs that are unique to manipulate models. This approach was used in [13] and [14]. The authors automatically generate a web application from a simplified class diagram.

- Approach by template: Consists of taking a "template model", canvas of configured target models, these settings will be replaced by the information contained in the source model. This approach requires a special language for defining model template.

- Approach by Modeling: The objective is to model the transformations of models, and make transformation models sustainable and productive and to express their independence of execution platforms. The standard MOF2.0 QVT [9] is used to define the meta model for the development of models for the models transformations.

In this paper we chose to transform PIM to PSM, with an approach by modeling. This type of transformation will allow us to automatically generate the multidimensional data warehouse schema from UML schema. Indeed, as shown in Figure 2, the definition of model transformation is a model structured according to the meta-model MOF2.0 QVT. Models instances of meta-model MOF2.0 QVT express structural mapping rules between meta-model source and meta-model target of transformation. This model is a sustainable and productive, it must be transformed to allow the execution of processing on an execution platform. The following figure illustrates the approach by modeling MOF2.0 QVT [12].

V. SOURCE META MODEL - TARGET META MODEL

In our MDA approach, we opted for the modeling approach to generate the data warehouse multidimensional schema. As mentioned above, this approach requires a source meta-model and a target meta-model.

A. Source Meta-Model

We present in this section the different meta-classes which form the source UML meta model used in [13], and our target meta-model MD to achieve the transformation between the source and target model. This transformation is based on that proposed by [13]. The source meta-model is a simplified UML model based on packages containing Class and Datatype. These classes contain typed properties and are characterized by multiplicities (upper and lower). The classes contain operations with typed parameters. The following figure illustrates the source meta-model:

- UmlPackage: expresses the notion of UML package. This meta-class is related to the meta-class Classifier.
• Classifier: it is a meta abstract class that represents both the concept of UML class and the concept of data type.
• Class: represents the concept of UML class.
• DataType: represents the data type of UML.
• Operation: expresses the concept of operations of a UML class.
• Parameter: parameter expresses the concept of an operation. They can be of two types or Class Datatype. This explains the link between meta-Parameter class and meta-class Classifier.
• Property: expresses the concept of Property of a UML class. These properties are represented by the multiplicities and meta-attributes upper and lower. A UML class is composed of properties, which explains the link between the meta-class Property and meta-class Class. These properties can be primitive type or class type. This explains the link between the meta-class and the Classifier meta-class Property.

B. Target Meta-Model

The target is a simplified dimensional meta-model based on packages containing Facts and Dimensions. A Fact contains attributes related to dimensions. Figure 5 shows the Target Meta Model:
• FactPackage: expresses the concept of package and we consider that each class contains a FactPackage.
• Fact: expresses the concept of Fact in a multidimensional schema. Each fact is contained in FactPackage, which explains the link between FactPackage and Fact.
• FactAttribute: expresses the concept of attributes for the element Fact and Fact contains FactAttribute.
• Dimension: expresses the concept of dimension in a multidimensional schema.
• DimensionAttribute: expresses the concept of dimensions attributes. A dimension contains attributes, which explains the link between a DimensionAttribute and Dimension.
• Datatype: represents the type of data in a multidimensional schema.

VI. QVT TRANSFORMATION OF THE UML META MODEL TO THE MULTIDIMENSIONAL META MODEL.

A. SmartQvt

To achieve the transformation cited above, we opted for the tool SmartQvt. This tool appeared in 2007 [17] which is also an implementation of the QVT-Operational language standard that allows the transformation of models. This tool compiles model transformations specified in QVT to produce Java code used to performs these transformations. This tool is provided as Eclipse plug-ins based on the meta-modeling framework, EMF, and is composed of three elements:
QVT Editor: helps end users to write QVT specifications.

QVT Parser: converts the textual concrete syntax on its corresponding representation in terms of meta-model QVT.

QVT Compiler: produces, from the QVT model, a Java API on top of EMF generated for the implementation of transformations. The input format is a QVT specification provides in XMI 2.0 format in accordance with the QVT meta-model.

B. The QVT transformation

The QVT transformation is a function of two arguments: the first is the source model, the second is the target model as shown in the figure below:

```
transformation uml2md(in srcModel:UML,out dest:MD);
main(){
  srcModel.objects()[UmlPackage] ->
  map UmlPackage2MDPackage(); }
```

Figure 6. Transformation function uml2md()

In this figure, we see the transformation function called uml2md() that takes as input source model, the simplified UML model shown in Figure 4, and as output the target model, the simplified multidimensional model shown in Figure 5.

As explained in the previous section, we consider a simplified UML model that consists of a package called UmlPackage. This package will be transformed into a multidimensional package called MDPackage, using the transformation rule sets in Figure 7. The MDPackage generated will be named MD followed by with the name of UmlPackage. The Association class will be transformed into a Fact by using the function associationclass2fait ().

```
mapping UmlPackage::UmlPackage2MDPackage () : MDPackage { name := 'MD' + self.name;
  fact:=srcModel.objects()[AssociationClass]->map
  associationclass2factattribute();}
```

Figure 7. Transformation rule of UMLPackage to MDPackage

```
mapping AssociationClass::associationclass2fact () : Fact {
  name := 'Fait'+self.name;
  factattribute:=self.assoproperty->map
  property2factattribute();
  dimension= self.assoproperty->select (upper=-1)->type->
  map type2dimension(); }
```

Figure 8. Transformation rule of Associationclass to a Fact

```
Figure 8 shows the transformation rule of an association class in a UML class diagram to a Fact.
```

```
mapping Classifier::type2dimension () : Dimension { srcModel.objects()()[Class]->forEach(c){
  if(self=c){ name:='Dimension'+self.name;
    base:= c._property->select (upper=-1)->type->
    map type2base(); } } }
```

Figure 9. Transformation rule of a Classifier to a Dimension

```
mapping Property::property2factattribute: FactAttribute {
  name:=self.name; }
```

Figure 10. Transformation rule of Property to FactAttribute

```
Figure 10 shows the transformation rule of UML class property to a Fact entity attribute, this rule uses the function property2factattribute ().
```

```
mapping Parameter::property2factattribute () : FactAttribute {
  name:=self.name; }
```

Figure 11. Transformation rule of Parameter to FactAttribute

```
mapping Property::property2dimensionAttribute () : DimensionAttribute {
  name:=self.name; }
```

Figure 12. Transformation rule of Property to DimensionAttribute

VII. CASE STUDY

In our case study, we use the UML model representing a business, among its activities, the sale of products to its customers, these products are classified under categories, and several sub categories form a category. The UML model representing part of the Information System of the company is represented in a format file XMI2.0. This file is composed of the element UmlPackage, which contains the following elements:
Figure 13. Overview of the instance of the UML model of our case study

An element named Customer, type of class, which has the following properties:
- **CustomerID**: Customer Identifier.
- **CustomerName**: Customer Name.
- **AccountNumber**: Number of the Customer Account.
- **Customer Type**: type of Customer, I=Individual, S=Store.
- **AddressID**: the Address Identifier.

An element named Address, type of class, which has the following properties:
- **AddressID**: Identifier of customer address.
- **AddressLine1**: Line 1 of the address.
- **AddressLine2**: Line 2 of the address.
- **City**: defines the city of customer.
- **State**: defines the state of customer.
- **PostalCode**: defines the postal code.

An element named Product, type of class, which has the following properties:
- **ProductID**: Product Identifier.
- **ProductName**: Product Name.
- **StandardCost**: Product Standard Cost.
- **ListPrice**: sale price
- **Weight**: Product Weight
- **Style**: W=Women, M=Male, U=Both
- **Color**: Product Color.
- **DayToManufacture**: Number of days to manufacture the product.
- **ProductSubcategoryID**: the product belongs to this sub category.

An element named ProductSubcategory, type of class, which has the following properties:
- **ProductSubcategoryID**: Product Sub category Identifier.
- **ProductSubcategoryName**: Product Sub category Name.
- **ProductCategoryID**: the sub category belongs to this category.

An element named ProductCategory, type of class, which has the following properties:
- **ProductCategoryID**: Product Category Identifier.
- **ProductCategoryName**: Product Category Name.

An element named SalesOrderDetail, type of class, which has the following properties:
- **SalesOrderDetailID**: Identifier of sales order.
- **OrderQty**: Quantity of Sales.
- **UnitPrice**: the selling price of a single product.
- **OrderDate**: Creation date of the sales order.
- **CustomerID**: Customer Identifier.
- **ProductID**: Product Identifier.

This file in XMI format is provided as input of QVT transformation, achieved under Eclipse Europa. As output of the QVT transformation, file in XMI format is generated.
automatically. It represents the multidimensional schema of company Datawarehouse in our case study. This file is composed of a Fact named FactSalesOrderDetail surrounded by Dimensions: Customer, Address, Product, Category, SubCategory.

The generated XMI file contains the element MDPackage composed of the following:

```xml
<?xml version="1.0" encoding="ASCII"?>
  <xsi:Type name="MDFactory">
    <Fact name="FactSalesOrder" factattribute=""/1/@fact.0"/>
    <Fact name="FactCustomer" factattribute=""/1/@fact.0"/>
    <Fact name="FactAddress" factattribute=""/1/@fact.0"/>
    <Fact name="FactProduct" factattribute=""/1/@fact.0"/>
    <Fact name="FactSubcategory" factattribute=""/1/@fact.0"/>
  </xsi:Type>
</xs:XML>
```

Figure 14. Overview of the instance of Multidimensional Model generated after execution of the transformation

VIII. CONCLUSION AND FUTURE WORKS

We applied the MDA approach for the engineering of SID. The objective is to define a model transformation. It is based on a UML source model and a Multidimensional target model. This transformation takes as input the source model and as output the target model. Once done, we can generate from a simplified UML model instance, a simplified multidimensional model.

Looking ahead, we plan to develop a graphical plug-in integrated to Eclipse, which automatically generates from a UML class diagram, a multidimensional diagram. In addition, we can extend this work to more generalized meta-models.

REFERENCES


MDA en action Ingénierie logicielle guidée par les modèles, Xavier Blanc, édition Eyrolles 2005.


SmartQVT, http://smartqvt.elibel.tm.fr/


AUTHORS PROFILE

L. Arrassen Graduate as Computer Science Engineer from the INPT(National Institut of Poste and Telecommunication) and Ph-D-Student at Faculty of Sciences, Laboratory for Computer Science Research, Mohammed First University, Oujda, Morocco.

A. Meziane is a Professor of Computer Sciences, Laboratory for Computer Science Research, Mohammed First University, Oujda, Morocco.
R. Esbai Ph-D-Student at Faculty of Sciences, Laboratory AMCSP: Applied Mathematics and Computer Signal Processing, Mohammed First University, Oujda, Morocco.

M. Erramdani is a Professor of Computer Sciences at Superior School of Technology Laboratory AMCSP: Applied Mathematics and Computer Signal Processing Mohammed First University, Oujda, Morocco.