An Improved Framework for Modelling Data Warehouse Systems Using UML Profile

Muhammad Babar[1], Akmal Khattak[2], Fahim Arif[3], and Muhammad Tariq[4]
[1]Department of Computing and Technology, Iqra University, Pakistan
[2]Department of Computer Sciences, Quaid-i-Azam University, Pakistan
[3]Signal College, National University of Sciences and Technology, Pakistan
[4]Abu Dhabi School of Management, Abu Dhabi, UAE

Abstract: *Data Warehouse (DW) applications provide past detail for judgment process for the companies. It is acknowledged that these systems depend on Multidimensional (MD) modelling different from traditional database modelling. MD modelling keeps data in the form of facts and dimensions. Some proposals have been presented to achieve the modelling of these systems, but none of them covers the MD modelling completely. There is no any approach which considers all the major components of MD systems. Some proposals provide their proprietary visual notations, which force the architects to gain knowledge of new precise model. This paper describes a framework which is in the form of an extension to Unified Modelling Language (UML). UML is worldwide known to design a variety of perspectives of software systems. Therefore, any method using the UML reduces the endeavour of designers in understanding the novel notations. Another exceptional characteristic of the UML is that it can be extended to bring in novel elements for different domains. In addition, the proposed UML profile focuses on the accurate representations of the properties of the MD systems based on domain specific information. The proposed framework is validated using a specific case study. Moreover, an evaluation and comparative analysis of the proposed framework is also provided to show the efficiency of the proposed work.*

Key Words: *UML Profile, Data Warehouse, UML, MD Modelling.*

1. Introduction

Modern civilization is characterized by a regular need to gather and store data about a variety of events. Information systems have very simply become in all the areas of the human lives, while databases that support them have enlarged to the scale of even petabytes. Billions of records in those databases do not necessarily need to be just a note that something specific happened in the past. They can be modelled and shaped in the way to represent meaningful pieces, that can be used, based on many recorded data, to infer some new knowledge, to follow the pattern of some changes and finally to help people in business to make decisions. Time when business people were stressed with huge amount of data which did not have any valuable meaning is almost passed; credit goes to Business Intelligence which has just become one of the most promising areas of IT world. Today and even more in the future, the companies would not be able to compete in the world market if they do not give an intelligent way of analyzing their data and extracting information that is critical for income growth.

Decision-making support systems represent the subclass of BI information systems. According to transactional sources used in everyday business, and additional usage of business specific logic, these systems should be able to identify incurred problems and to propose corresponding solutions. These systems are known as Data Warehousing (DW) systems, and the beginning of such systems is the appropriate design. Data Warehouse system is based on specific subject, different sources, time-variant, and non-volatility to help decision making process [5]. This actually means that the Data Warehouse represents a single storage for all domain-relevant data, from various available sources (integrated), collected during the particular period of time (time-variant). Additionally, this also states that this storage is stable in terms that data can only be inserted but never updated or deleted from data warehouse (non-volatile). Even though the area of decision making systems evolved a lot, this definition is still mostly accurate.

DW applications provide past data and information for the process of decision making of a specific organization. It is acknowledged worldwide that multidimensionality provide basis for such systems. MD modelling is based on a number of characteristics unlike those in conventional database modelling [6, 13]. The MD stores data in the shape of facts and dimensions. A fact includes motivating factor or metric of business logic and dimension characterize the viewpoint for investigating a fact.

The conventional modelling technique of database for example the Entity Relationship model and other extensions of ER, are not suitable for DW design, due to the difficulty of the corresponding models. Different authors have highlighted the issue of unsuitability of ER model for DW systems. ER models cannot be used as the foundation for DW. However, afterward data models adapted for DW, such as the famous star schema of Ralph Kimball, neither they are able to consider the main peculiarities of MD model. Different authors have highlighted the issue of unsuitability of ER model for DW systems [6, 25]. ER models cannot be used as the foundation for DW. However, afterward data models adapted for DW, such as the famous star schema of Ralph Kimball, neither they are able to consider the main nor important peculiarities of MD model [14].

The UML offers a standard way to concentrate on the design and modelling of tangible things of a system [19]. Though, it is short of modelling the essentials of a system which is not Object-Oriented. Nevertheless, UML offers the UML profiling mechanism. Profile is actually the integration and composition of an expansion which tailors UML for a specific field (such as aerospace, real-time processing, critical systems etc.,) or platform (such as the J2EE, Java Beans, NET etc.,) [1, 28]. It is preferred for systems other than object-oriented.

The profiles are described using stereotypes, constraints, and tag values that are implemented to particular UML component, such as Classes etc. The UML extension mechanism is composed of the stereotypes, constraints and tagged values.

- Stereotypes A stereotype is an element that has the description for bringing novel characteristics to the standard UML components. A stereotype can also provide new notations.
- Constraints Constraints are a composition of rules articulated in Object Constraint Language or natural language. They are expressed in words to introduce new semantics to UML.
- Tagged Value Tagged value specifies a model element to indicate a novel property for current UML model ingredient.

This paper illustrates a UML profile to achieve a precise, consistent, and comprehensible framework to model the MD properties. The proposed profile is used to cater for all the main properties of MD systems. In addition, this extension focuses on the accurate representations of the properties of the MD systems. The paper is well thought-out in the mode to completely represent the research work. Section 2 represents the related work in the field of MD modelling including some early UML extension attempts and some newly approaches based on UML profiles. In the sequel, section 3 exhaustively talks about new proposed approach (UML Profile). Section 4 presents the applicability of the framework using a case study. Section 5 is about the evaluation of proposed framework through comparing the proposed framework with existing models. At the end, the conclusions are argued in section 6.

2. Related Work

Researchers of the domain of DW and MD modelling have proposed object oriented multidimensional models, and even some of them are based on UML. Some of the proposals model the basic features of MD systems, but these models lack the completeness and accuracy of the features comparing with MD properties [12, 17, 20, 26]. By lack of completeness it is meant not to model all the main properties and features of MD systems. Similarly, by lack of accuracy mean some features not modelled and covered according to the specific domain information of multidimensionality. For instance, the properties of the Fact class are modelled as Fact Attributes; however, the properties of fact are known as measures and they are categorized as Atomic and Derived measures [2, 15].

Similarly, the derived measure is not handled separately while it should be handled separately through a specific stereotype. Classic multidimensional basics are described through package diagram and 14 different stereotypes are suggested to model the different properties of the multidimensional modelling [16]. However, it is not according to the specific domain information such as the degenerate dimension. Another model is much more sophisticated in terms of completeness but still lacks to model some features [27]. Some of the models are much more relevant and sophisticated, but they lack the completeness to cover all the main properties of MD systems [3, 15, 16, 27]. These proposals miss out some of the important features and characteristics of the MD systems due to the evolving of MD area, latest version of UML.

Data mining methods are used on data warehouse to haul out data. Therefore, the data mining process ought to be kept into consideration modelling of data warehouse. With regard to mining techniques modelling at early stage, several proposals are found in literature, but did not emphasize the completeness of MD model [4, 9, 10, 29, 30, 31]. It is noticed that all the proposals of mining-aware design are focused on concept of integration of mining and data warehouse and their attention was towards the mining concepts mostly. A well-built trend subsists in Computer Science to take on the UML as a modelling standard for OO paradigm. As far as the database modelling in concerned, the UML class diagram is mostly suitable and considerable [28]. However, the class diagram substitutes the ER model for traditional

database systems and don't cover the aspects of MD systems.

The literature depicts that major perspectives of the MD model taken under consideration and innovative solutions are anticipated, but to date there are some of the features of MD model left to be modelled [7, 21]. The reason of these issues is the evolution of this field some are left because of the arguable behaviour of some of the properties. Some of the most important features of MD systems such as temporal dimension, surrogate key, reference key, degenerate key etc are not modelled previously. In addition, Object Management Group (OMG) specification suggests proposing specific icon related to the stereotypes information, but literature lacks this issue [22, 23]. However, some notations and icons are found in literature but they are not stereotype icons as suggested by OMG. Previous proposals are not according to the specific stereotype information.

Moreover, the whole-part relationship between dimension and hierarchy levels is also not discussed. In addition, some of the previously solutions are not according to the specific domain information such as the degenerate dimension. Similarly, the properties of the Fact class are modelled as Fact Attributes. However, the properties of fact are known as measures and they are categorized as Atomic and Derived measures. Similarly, the derived measure is not handled separately while it should be handled separately through a specific stereotype.

The UML profile should be preferred because the extension of the UML has the following key advantages [8, 11]:

1. To utilize the current CASE tools by profitable and open source organizations, because these tools are developed for UML and can be formulated for the UML profiles.
2. To avoid the description from scuff of a novel modelling language, which would need to detail the entire requirements and also provide tool support.

3. Proposed Framework

This section provides the detailed description of proposed UML Profile. This UML profile is used to extend the UML class diagram. Class diagram is one of the most widely used diagrams of UML which is mostly used for the object-oriented environment to specify classes and objects. The reasons to extend the class diagram of UML are discussed.

- The proposed framework is used to define the structural properties of the MD model. Class diagram is most commonly used diagrams of UML to describe the structural properties.
- As far as the database modelling in concerned, the UML class diagram is mostly suitable and considerable [18, 24].

- The literature is also available which are based on class diagram to extend the UML for the MD modelling.

Considering all the aspects of the MD model, the UML extension for the mentioned properties for MD model respectively is defined. Fact and Dimension are symbolized by the stereotyped classes of Fact and Dimension respectively in proposed framework. A Fact class contains Reference Keys, Measures and Degenerate Key as Attribute stereotypes. Measures are categorized as Atomic and Derived stereotypes of Attribute. A Dimension class consists of the attribute stereotypes of Surrogate Key and Dimension Attribute and Reference Key (with specific constraints). A Dimension can be Temporal Dimension that is represented by Temporal Dimension class stereotype. The hierarchy and hierarchy levels of Dimensions are symbolized by Hierarchy class stereotypes. The description of the stereotypes and their specific icons is depicted in Figure 1.
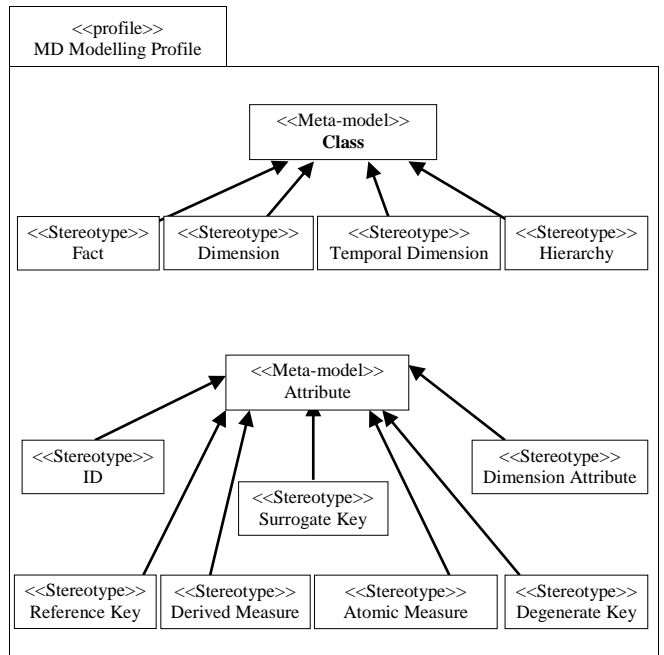

Figure 1. High level view of proposed md modeling.

3.1. Stereotype Definition

A stereotype is a modelling component that has definitions for adding new properties and constraints to the standard UML elements for a particular problem domain generating a profile tailored for a particular problem domain. The stereotypes are presented depending on the meta-class of the UML models that is specialized. This section gives the details of all the stereotypes listed in the proposed UML extension. The procedure presented in Table 1 is followed to define proposed stereotypes. A stereotype is defined with its well defined name, the model element of UML which will be extended, description about the stereotype use, the graphical representation and the list of constraints.

Table 1 shows the tabulated form of the definition of a stereotype. Each and every stereotype includes the unique name, the description of the model element of the UML from which the stereotype is derived, description of the stereotype, a specific iconic representation, and the list of constraints.

Table 1. Procedure for stereotype definition.

ACTION	EXPLANATION
Name	It represents the name of the stereotype
Model Element	The UML meta-model element which is the base for the stereotype
Description	A little description in natural language
Icon	A distinctive visual cue or a short form for representation
Constraint	A list of constraints associated with stereotype

3.2. Proposed Stereotypes of Class Model

In proposed profile class stereotypes are used to symbolize the main features of MD model. These are used to represent the focus of analysis and their corresponding different views used to analyze the main subject of analysis in MD model. In the proposed extension, four stereotypes are defined which are Fact, dimension, temporal dimension and hierarchy. All the class stereotypes and their detail including their specific icons, constraints and other details are given in the tabulated form.

The fact stereotype is defined by extending the UML class model element. It is named as Fact. A cubical icon for the fact stereotype is proposed because the fact is analyzed from different dimensions in the form of a cube. Table 2 shows the tabulated form of the fact stereotype. Fact stereotype has the following constraints:

- A Fact can only be associated with Dimension and Temporal Dimension.
- All attributes of the Fact stereotype can be Reference Key, Degenerate Key, Derived Measure, or Atomic Measure.
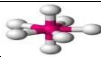
Table 2. Stereotype of fact.

Name	Fact
Model Element	Class
Description	Classes of this stereotype represents facts in MD model
Icon	
Constraints	• A fact can only be associated with Dimension and Temporal Dimension classes. • All attributes of facts must be Reference Key, Degenerate Key, Derived Measure or Atomic Measure.

Table 3 elaborates the dimension stereotype and shows that dimension is also extended from class element of UML and all the dimension stereotypes will represent the dimension in the MD model except temporal dimension. The temporal dimension is handled separately in the proposed model due to its importance. Table 3 also shows the proposed icon of dimension stereotype which reveals the multidimensionality. The corresponding constraints are also listed.

Table 3. Stereotype of dimension.

Name	Dimension
Model Element	Class
Description	Classes of this stereotype represents dimensions in MD model
Icon	
Constraints	• A Dimension cannot be associated with another Dimension or Temporal Dimension. • All attributes of a Dimension must be Surrogate Key, Dimension Attribute or Reference Key (if the dimension has the hierarchies)

A dimension is called Temporal Dimension if it is used to represent the time related measurement. In a DW, time is the governing dimension. Temporal Dimension class is included that indicates time dimension. In addition, this property of MD model is not highlighted in the past. The temporal dimension stereotype is also defined using class model element. This stereotype represents temporal dimension in the MD model. The icon of this stereotype is derived from dimension stereotype icon. Table 4 shows the tabulated form of the temporal dimension stereotype. A separate graphical icon is proposed for Temporal Dimension.
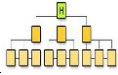
Table 4. Stereotype of temporal dimension.

Name	Temporal Dimension
Model Element	Class
Description	Classes of this stereotype represents temporal dimensions in MD model
Icon	
Constraints	• A Temporal Dimension cannot be associated with another Temporal Dimension or Dimension • All attributes of a Temporal Dimension must be Surrogate Key, Dimension Attribute or Reference Key (if the temporal dimension has the hierarchies)

A hierarchy is a mean to classify data at special level or a number of levels. To view the data and information, the forecasters employ the hierarchies of a dimension to identify tendency at different levels. Table 5 elaborates the hierarchy stereotype and shows that hierarchy is also extended form class element of UML and all the levels of dimension stereotypes will be represented by the Hierarchy class stereotype. Table 5 also shows the proposed icon of Hierarchy stereotype which reveals the concept of hierarchy. It has the following constraints:

- A Hierarchy can only be associated with a Dimension, Temporal Dimension, or another Hierarchy.
- All attributes of Hierarchy must be Identifying Attribute, Dimension Attribute or Reference Key (if it has further levels).

Table 5. Stereotype of hierarchy.

Name	Hierarchy
Model Element	Class
Description	Classes of this stereotype represents different hierarchy levels of a dimension and temporal dimension in MD model
Icon	
Constraints	• A Hierarchy can only be associated with a Dimension, Temporal Dimension, or another Hierarchy. • All attributes of Hierarchy must be Identifying Attribute, Dimension Attribute or Reference Key (if it has further levels)

3.3. Proposed Stereotypes of Attribute

Attributes stereotypes are used to represent the properties of the main elements. In the proposed UML extension, seven stereotypes have been defined from the Attribute model element of the UML, they are: Reference Key, Degenerate Key, Derived Measure, Atomic Measure, Surrogate Key, Dimension Attribute, and Identifying Attribute. The definition of these stereotypes and their details are listed in the tabulated form.

A Reference Key is used to represent the Surrogate key of Dimension classes in the Fact class. In addition, it can also be used to represent the connection between Dimension and Hierarchy classes. The Reference Key stereotype is defined to extend Attribute model element of UML and is used to represent the reference keys in the Fact class. The icon of this stereotype is derived from its name. Table 6 shows the tabulated form of the Reference Key stereotype and all the constraints related to this stereotype. A Reference Key must belong to a Fact class and may also belong to Dimension (if Dimension class has the hierarchies) and Hierarchy (if that Hierarchy class has further levels of hierarchies), where the lowest level of Hierarchy must not have this key.

Table 6. Stereotype of reference key.

Name	Reference Key
Model Element	Attribute
Description	Attribute of this stereotype represents Reference Key in MD model
Icon	**RK**
Constraints	• A Reference Key must belong to a Fact class. • It may also belong to Dimension class, if that Dimension class has the hierarchies. • It may also belong to the Hierarchy class, if that Hierarchy class has further levels. • The lowest level of Hierarchy class must not have this key

There are some cases in which the dimension is not considered explicitly because the majority of its characteristics are previously characterized through other essentials. However, it is still believed that some of its characteristics are needed in the fact to exclusively recognize fact occurrences. When this occurs, this dimension is called as degenerated dimension and this degenerate dimension is represented by a degenerate key in the fact class.

Therefore, a degenerate dimension is that which identifier presents merely in a fact, but that is not appear as a real dimension. This provides other fact features in addition to the measures for analysis. Consequently, degenerate dimension stereotype is defined using Attribute model element of the UML. It is named as Degenerate Key and used to represent degenerate dimension in the MD model. The icon of this stereotype is derived from its name as well. Table 7 shows the tabulated form of the Degenerate Key stereotype and all the constraints related to this stereotype.

Table 7. Stereotype of degenerate key.

Name	Degenerate Key
Model Element	Attribute
Description	Attribute of this stereotype represents degenerate dimension in MD model
Icon	DK
Constraints	• A Degenerate Key can only belong to a Fact class stereotype

Measures are the key properties of the Fact class. The derived measure is the one which is derived from other properties that is why this stereotype of Attribute model element also consist a tagged value which is known as Derivation Rule. This tagged value will be placed somewhere outside the Fact class if Fact class consists the Derived Measure. Table 8 elaborates the Derived Measure stereotype and shows that it is also extended from Attribute model element of UML. The Derived Measure does not have any constraints as it has specific tagged value.

Table 8. Stereotype of derived measure.

Name	Derived Measure
Model Element	Attribute
Description	Attribute of this stereotype represents derived measures of the fact class in MD model
Icon	**DM**
Constraints	N/A
Tagged Value	Derivation Rule

On the other hand, the Atomic measure is the measure which is independent. Table 9 shows the details of the atomic measure that it is extended from Attribute model element, represents the atomic measures in the Fact class, represented by a specific icon and specific constraints are also listed. It can only belong to a Fact class.

Table 9. Stereotype of atomic measure.

Name	AtomicMeasure
Model Element	Attribute
Description	Attribute of this stereotype represents atomic measures of the fact class in MD model
Icon	**AM**
Constraints	• Atomic Measure can only belong to a Fact class.

Surrogate keys links the table of fact with the table of dimension. It provides an important resource of recognizing each occurrence inside of a dimension. Surrogate Key stereotype is introduced to represent the surrogate keys of the dimensions in the Dimension

classes. It can only belong to Dimension and Temporal Dimension classes and represented by a specific icon. Table 10 shows the tabulated form of this stereotype.

Table 10. Stereotype of surrogate key.

NAME	Surrogate Key
Model Element	Attribute
Description	Attribute of this stereotype represents surrogate keys of the dimensions and temporal dimensions in MD model.
Icon	SK
Constraints	• A Surrogate Key can only belong to a Dimension or Temporal Dimension.

Dimensions consist of specific attributes called dimension attributes. In the proposed extension of UML, Attribute element of UML is extended for the representation of dimension attribute in the form of a stereotype. These stereotypes can belong to Dimension, Temporal Dimension and Hierarchy classes. Table 11 elaborates the tabulated form of the Dimension Attribute stereotype.

Table 11. Stereotype of dimension attribute.

Name	Dimension Attribute
Model Element	Attribute
Description	Attribute of this stereotype represents dimension attributes of the dimensions, temporal dimensions and hierarchies in MD model
Icon	DA
Constraints	• A Dimension Attribute can only belong to a Dimension, Temporal Dimension, or Hierarchy.

The last stereotype of the proposed extension is Identifying Attribute. Table 12 elaborates the identifying attribute stereotype. This stereotype is used to represent the hierarchy levels of the Dimension class in a unique form.

Table 12. Stereotype of identifying attribute.

NAME	Identifying Attribute
Model Element	Attribute
Description	Attribute of this stereotype represents identifying attributes of hierarchy classes in MD model
Icon	ID
Constraints	Identifying Attribute can only belong to Hierarchy classes.

A number of possible representations for a stereotyped class are defined in order to model the MD aspects. These representations are used to model the corresponding MD property described in the definition of the stereotypes. There are three representations that can be used in the proposed profile to model the MD properties.

- *Standard Label*: This is the standard form of representing the stereotype and proposed by OMG. This stereotype representation is mostly used in different UML profiles. In this type of representation, the stereotype name is shown between a pair of guillemots (<<>>) to represent the defined stereotype of class. In Figure 2, this representation of stereotype for different aspects is shown.

<<Fact>> Sales	<<Dimension>> Product
<<ReferenceKey>> ATR <<DegenerateKey>> ATR <<AtomicMeasure>> ATR	<<SurrogateKey>> ATR <<DimensionAttribute>> ATR

Figure 2. Label Representation of fact (sales) and dimension (product).

- *Decoration*: In this type of representation, the defined icon for the stereotype is used. The icon of the stereotype is positioned in the upper right of the consequent model. This representation of stereotype for different aspects is shown in Figure 3.

Figure 3. Decoration representation of fact (sales) and dimension (product).

- *Icon*: In Figure 4, another representation is shown. In this type of representation, the stereotype is depicted with the icon. In addition, the attributes are placed below the stereotype icon.
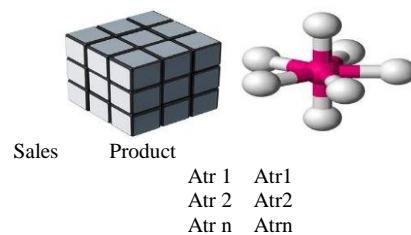

Figure 4. Iconic Representation of Fact (sales) and Dimension (product).

4. Stanley Pharmaceutical: A Case Study

In this section a specific case study is developed through which it is illustrated that how the extended UML profile can be applied. This case study is used to show how the proposed MD modelling profile is properly applicable to the MD properties. The case study which is used is medicine industry related case study. Stanley pharmaceutical company has employed 400 employees. The company produces 100 various product items. The product information is recorded as product name, product group, product pack/size/presentation, unit price, registration number and batch number. The company distributes its products to 52 various distributors nationwide. Each distributor is licensed with the government and is one per district. The distributors are recorded as distributor company name, shipping address, district, province and contact number. When talking about the average, about 45000 items are prepared and about 44000 items are sold daily. Averagely 15 to 20 orders are received

daily. The order detail about each product is recorded in a particular fashion.

4.1. Identified Fact and Dimensions

The major orientation is towards the SALE of our products that is either not satisfactory in a season (winter) of some products or is not satisfactory by some organization (distributor) or spoiled in a location. The proposed Fact according to the complete study and thorough review is given in Table 13. This Fact is identified through a proper analysis with Database Administrator of Stanley Pharmaceutical. Moreover, there could also be some other subject areas, but sale is one of the most common and more affective subject is in the mentioned case study.

Table 13. Identified fact of stanley pharmaceutical.

SUBJECT AREA	DEFINITION
Sales	The product may be sold by a distributor at a specific time

To analyze the major subject of analysis (Fact) from different point of views, it is important to identify the most common entities (Dimension) that can analyze the Sales (Fact). Table 14 lists all the dimensions and their short description which are used in our case study to apply our MD modelling profile. These dimensions are Product, Distributor, Date, Location and Customer. The Customer dimension is considered as degenerate dimension.

Table 14. Identified dimensions of stanley pharmaceutical.

DIMENSION	DESCRIPTION
Product	The actual product which is sold.
Distributor	The distributors who distribute the products of the company.
Date	The date (time period) when a specific product is sold.
Location	List of locations where the product is sold.

4.2. Stanley Pharmaceutical's MD Model

Figure 5 shows the MD model of Stanley Pharmaceutical using the proposed MD modelling profile. In this model there is a Fact class called Sales and the corresponding four Dimensions which are Date, Product, Distributor and Location. Where, Date is a Temporal Dimension. In addition, there is a degenerate dimension called Customer which is represented through a Degenerate Key (DK) attribute in the Fact (Sales) class. The Fact (Sales) also has Reference Keys (RKs) shown in Figure 5. The measures of the Fact (Sales) are also identified by their corresponding icons in the Fact class. In addition, there is a Derived Measure (DM) called total-price. This attribute is dependent on product-price and quantity of the products. So, the corresponding tagged value (DerivationRule) of this derived measure is put outside the Fact class. Derivation Rule is a rule through which the derived measure is derived to be evaluated. Moreover, all the dimensions have their corresponding

Surogate Key (SK) attributes which are used to represent their dimensions in a unique way shown in Figure 5. In addition, the corresponding attributes (properties) of the dimensions are represented through Dimension Attribute (DA) stereotypes. The product dimension can further be extended to its hierarchy levels which are represented by the Hierarchy class stereotypes shown in Figure 5. The case study has shown that the proposed profile is applicable in real situation of multidimensional modelling and DW. Through this case study it is discussed and shown in Figure 5 that all the properties of multidimensionality can be modelled using the proposed profile.
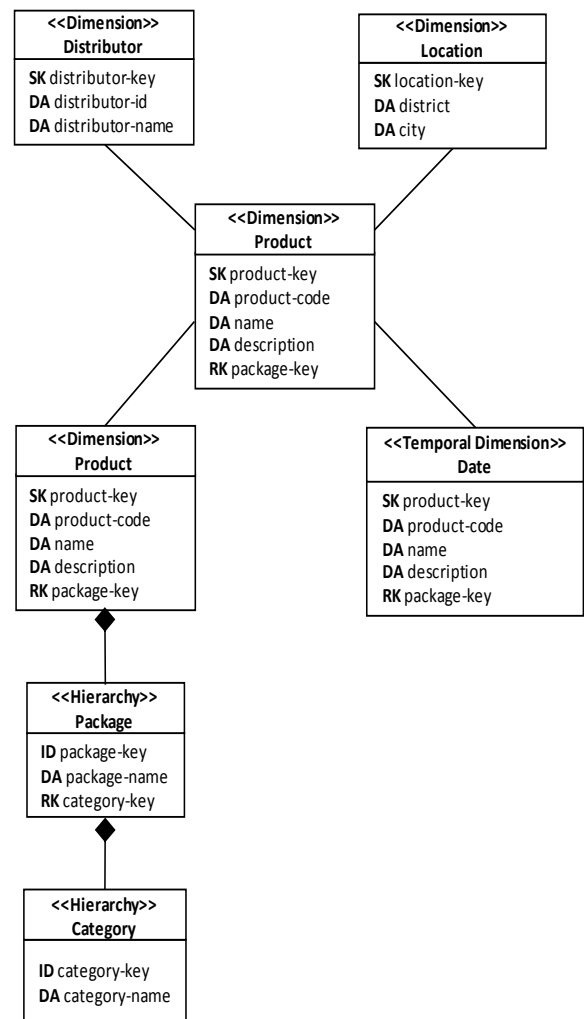
Figure 5. MD Model of Stanley Pharmaceutical using Proposed Framework.

5. Comparative Study

In this section, a comparative study of existing MD models with respect to main MD properties is provided. In literature a number of proposals are found which are used to model the MD systems. However, in this comparative study, only the most relevant conceptual multidimensional models are evaluated and analyzed which are based on UML. These models provide a high level of abstraction for assessing multidimensional properties. Table 15provides the

comparison of MD properties discussed by various authors and highlights the efficiency of proposed work. The properties of MD model are taken as the criteria through which different models based on UML are compared. These criteria are taken because in literature it is found to analyze and discuss the MD models. These criteria include the following main properties of the MD modelling.

- Iconic Representation.
- Atomic Measure.
- Derived Measure.
- Degenerate Key.
- Reference Key.
- Surrogate Key.
- Temporal Dimension.
- Classifications of Hierarchy.
- Many-to-Many Relationship.
- Aggregation.
- Composition of Dimension and Hierarchies.

Table 15 shows the main MD properties row wise and the different models column wise. Only [15] provides the iconic representations, but even that is also not according the definition. Therefore, the proposed framework provides accurate iconic representation. Moreover, the atomic measure is the only MD property which is covered by all the models. But even some models discussed it as a fact attribute which is not a proper term. In addition, a measure could be derived as well which should be separated from other measures. Models [11, 15, 24] identify this issue but a specific stereotype is not proposed. Degenerate dimension is taken into account by [15] but none of the rest of the models considers it. Even [15] models it as a degenerate dimension but this aspect of multidimensionality should be represented as a key.

Table 15. Comparative analysis.

MD PROPERTIES	MODELS					
	[11] 2006	[15] 2006	[17] 2019	[24] 2011	[25] 2010	Proposed Profile
Iconic Representation	NO	YES	Partial	NO	NO	**YES**
Atomic Measure	YES	YES	YES	YES	YES	**YES**
Derived Measure	YES	YES	YES	YES	NO	**YES**
Degenerate Key	NO	NO	NO	NO	NO	**YES**
Reference Key	NO	NO	Partial	NO	NO	**YES**
Surrogate Key	NO	NO	NO	NO	NO	**YES**
Temporal Dimension	YES	NO	Partial	NO	NO	**YES**
Classifications of Hierarchy	YES	YES	NO	YES	YES	**YES**
Many-to-Many Relationship	NO	YES	YES	YES	YES	**YES**
Aggregation	NO	YES	Partial	YES	YES	**YES**
Composition of Dimension and Hierarchies	NO	NO	NO	NO	NO	**YES**

Table 15 shows that none of the models includes surrogate key and the reference key combination. The proposed framework provides complete details of these properties. The proposed framework provides a specific stereotype with specific constraints to distinguish temporal and non-temporal dimensions. Only [11] gives some identification of temporal level of dimension but that mostly focuses on physical level and also do not specify it as a specific stereotype. With regard to hierarchy, fact/dimension relationship and aggregation, some partial modelling is seen. Few models cover all these properties, but they have some deficiencies [24, 25]. These models do not consider the composition association between dimension and hierarchy levels.

After comparing, evaluating and analyzing proposed extension of UML, a proposal with the desired characteristics is achieved. These characteristics show the efficiency of proposed framework compared with existing models.

- Precise: The proposed profile allows the designer to symbolize all significant features of MD modelling at the conceptual level accurately.
- Consistent: The new defined elements in this profile are mutually exclusive and complete in terms of modelling all the properties of MD systems. It means that this approach covers the complete domain space of the MD modelling.
- Simple: The graphical notations of proposed profile are limited to a smallest subset of UML elements that allow the designer to describe the main MD properties in effortless way.
- Comprehensible: The proposed profile is understandable for the desired audience (both DW designers and end users).
- Tool Support: The proposed MD modelling profile has a rich set of CASE tools to be supported.

6. Conclusions

The most common area of the DW which is MD modelling is elaborated in our research. An improved solution is provided to cover the designing and modelling aspects of the MD. UML Profiles are seen as attractive solutions for various specific domains other than modelling, documenting, and specifying object-oriented artefacts. This work addresses a UML extension based on UML class meta-model need for multidimensional modelling of DW applications. In this research an attempt is made to present in depth study of the current MD models using UML extensions. This study reveals that the current models are not accurate and complete regarding the properties and features of the MD model. A framework is proposed based on UML extension mechanism, which can be used to model all the features and aspects of the MD model in a precise way. In addition, the proposed

framework is consistent, precise, simple and comprehensible. The developed model provides generic way because UML notations are familiar and acceptable among the designers and architect widely.

References

[1] Abelló A. and Samos J., "YAM/Sup 2/(Yet Another Multidimensional Model): An Extension of UML," *in Proceedings of International Database Engineering and Applications Symposium*, Edmonton, pp. 172-181, 2002.

[2] Amaral G. and Guizzardi G., "On the Application of Ontological Patterns for Conceptual Modellingin Multidimensional Models," *in Proceedings of European Conference on Advances in Databases and Information Systems*, Bled, pp. 215-231, 2019.

[3] Azzaoui A., Rabhi O., and Mani A., "A Model Driven Architecture Approach to Generate Multidimensional Schemas of Data Warehouses," *International Journal of Online and Biomedical Engineering*,vol.12, no. 12, pp. 18-31, 2019.

[4] Abdelhedi F., Brahim A., and Zurfluh G., "Formalizing the Mapping of UML Conceptual Schemas to Column-Oriented Databases," *International Journal of Data Warehousing and Mining*, vol. 14, no. 3, pp. 44-68, 2018.

[5] Campeau-Lecours A., Lamontagne H., Latour S., Fauteux P., Maheu V., Boucher F., Deguire C., and L'Ecuyer L., "Kinova Modular Robot Arms for Service Robotics Applications," *International Journal of Robotics Applications and Technologies*, vol. 5, no. 2, pp. 693-719, 2019.

[6] Das I., Roy S., Chatterjee A., and Sen S.,"A Data Warehouse Based Schema Design on Decision-Making in Loan Disbursement for Indian Advance Sector," *in Emerging Technologies in Data Mining and Information Security*, vol. 2, pp. 603-614, 2019.

[7] Dhawan B. and Gosain A., "Extending UML for Multidimensional Modeling in Data Warehouse," *International Journal of Computer and Communication Technology*, vol. 2, no. 7, pp. 59-64, 2011.

[8] Debnath N., Garis A., Riesco D., and Montejano G., "Defining Patterns Using UML Profiles," *in Proceedings of IEEE International Conference on Computer Systems and Applications*, Dubai, pp. 1147-1150, 2006.

[9] El Hayat S. and Bahaj M., "A Temporal Data Warehouse Conceptual Modelling and its Transformation into Temporal Object Relational Model," *in Proceedings of International Conference on Advanced Intelligent Systems for Sustainable Development*, Marrakech, pp. 314-323, 2018.

[10] El Moukhi N., El Azami I., Mouloudi A., and ElMounadi A., "Requirements-Based Approach for Multidimensional Design," *Procedia Computer Science*, vol. 148, pp. 333-42, 2019.

[11] Fernández L. and Vallecillo A., "An Introduction to UML Profiles," *UML and Model Engineering*, no. 2, pp. 6-13, 2004.

[12] Gosain A. and Mann S., "Object Oriented Multidimensional Model for a Data Warehouse with Operators," *International Journal of Database Theory and Application*, vol. 3, no. 4, pp. 35-40, 2010.

[13] Kochar B. and Chhillar R., "An Effective Data Warehousing System for RFID Using Novel Data Cleaning, Data Transformation and Loading Techniques," *The International Arab Journal of Information Technology*, vol. 9 no. 3, pp. 208-16, 2012.

[14] Luján-Mora S., Vassiliadis P., and Trujillo., "Data Warehouse Design with UML," *International Conference on Conceptual Modeling*, Salvador, pp. 191-204, 2005.

[15] Luján-Mora S., "Multidimensional Modeling using UML and XML," *in Proceedings of 12th Workshop for PhD Students in Object-Oriented Systems, 16th European Conference on Object-Oriented Programming*, Málaga, vol. 2548, pp. 48-49, 2002.

[16] Luján-Mora S., Trujillo J., and Song I., "A UML Profile for Multidimensional Modeling in Data Warehouses," *Data and Knowledge Engineering*, vol. 59, no. 3, pp. 725-69, 2006.

[17] Nguyen T., Tjoa A., and Wagner R., "An Object Oriented Multidimensional Data Model for OLAP," *in Proceedings of International Conference on Web-Age Information Management*, Beijing, pp. 69-82, 2000.

[18] Naiburg E. and Maksimchuck R., UML for Database Design, Addison-Wesley Professional, 2001.

[19] Platt R. and Thompson N., The Past, Present, and Future of UML, in Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics, IGI Global, 2019.

[20] Prat N., Akoka J., and Comyn-Wattiau I., "A UML-Based Data Warehouse Design Method," *Decision Support Systems*, vol. 42, no. 3, pp. 1449-73, 2006.

[21] Prat N., Comyn-Wattiau I., and Akoka A., "Representation of Aggregation Knowledge in OLAP Systems," *in Proceedings of 18th European Conference on Information Systems*, Pretoria, pp. 7-9, 2010.

[22] Specification OMG Infrastructure "Object Management Group," Needham MA, USA 2.2 2006.

[23] Schacher M., "Integrated Enterprise Modelling Based on OMG Specifications," *in Proceedings of 12th International IEEE Enterprise Distributed Object Computing Conference*, Munich, 2008.

[24] Song E., Yin S., and Ray I., "Using UML to Model Relational Database Operations," *Computer Standards and Interfaces*, vol. 29, no. 3, pp. 343-54, 2007.

[25] Thalheim B., *Entity-Relationship Modelling: Foundations of Database Technology*, Springer Science and Business Media, 2013.

[26] Trujillo J., Soler E., Fernández-Medina E., and Piattini M., "A UML 2.0 Profile to Define Security Requirements for Data Warehouses," *Computer Standards and Interfaces*, vol. 31, no. 5, pp. 969-83, 2009.

[27] Theodoratos D., Ligoudistianos S., and Sellis T., View Selection for Designing the Multidimensional Structure, *Data and Knowledge Engineering*, pp. 219-240, 2001.

[28] Unified Modeling Language: Superstructure version 2.4.1. OMG Group. 2016.

[29] Villarroel R., Fernández-Medina E., Piattini M., Trujillo J., "A UML 2.0/OCL Extension for Designing Secure Data Warehouses," *Journal of Research and Practice in Information Technology*, vol. 38, no. 1, 2006.

[30] Zubcoff J. and Trujillo J.,"A UML 2.0 Profile to Design Association Rule Mining Models in The Multidimensional Conceptual Modeling of Data Warehouses," *Data and Knowledge Engineering*, vol. 63, no. 1, pp. 44-62, 2007.

[31] Zubcoff J., Pardillo J., and Trujillo J., "A UML Profile for the Conceptual Modelling of Data-Mining with Time-Series on Data Warehouses," *Information and Software Technology*, vol. 51, no. 6, pp. 977-92, 2009.

Muhammad Babar is currently associated with IQRA University Islamabad Campus. He received hisPh.D degree in Computer Software Engineering from National University Sciences and Technology (NUST), Islamabad, Pakistan. He did his Masters of Sciences in Computer Software Engineering from National University Sciences and Technology (NUST), Islamabad, Pakistan in 2012. He receives his Bachelors in Computer Sciences with distinction from University of Peshawar, Pakistan in 2008. His research area includes but not limited to Big Data Analytics, Internet of Things (IoT), Smart City Design and Planning, and Social Web of Things (SWOT). He has published his research work in various IEEE and ACM/Springer International conferences and journals. He is an active reviewer and guest editor in the reputed journals.

Akmal Khattak received MS Computer Science degree from National University of Computer and Emerging Sciences (NU-FAST), Islamabad, Pakistan, in 2006. He got his Ph.D. in Engineering–Computer Science) from University of Leipzig, Germany, in 2014. He later joined Department of Computer Sciences, Quaid-i-Azam University, Islamabad, Pakistan, as an Assistant Professor in Sep2014. He also served as Assistant Professor in Iqra University, Islamabad from Sep 2017 to Aug 2018. Since September 2018, he has been with the Quaid-i-Azam University, Islamabad, Pakistan, as an Assistant Professor. His current research interests include information retrieval systems, natural language processing, machine learning, text mining and recommender systems. Dr. Khattak was a reviewer in NUST Journal of Engineering Sciences (NJES), Journal of the American Society for Information Science and Technology (JASIST).

Fahim Arif receives his Bachelors in Telecommunication from College of Telecommunication Engineering (UET Lahore) in 1995 and Master in Sciences in Computer Software Engineering from National University Science and Technology, Islamabad in 2003. He has won NUST Endowment fund scheme scholarship for NUST in 2003 and International Research Support Initiative Program Fund from HEC in 2007. He worked as international research scholar in System and Computer Engineering Department, Carleton University, Ottawa, Canada in 2007 and participated in numerous research and academic activities. He is principal investigator (PI) for a project funded by NUST. Recently, his biography has been published by South Asian Publication Who's Who in the World 2008 Edition and awarded with Star Laureate 2008 in recognition to his contributions to knowledge and research.

Muhammad Tariq has more than 13 years' experience in industry and academia. He earned his PhD in Management with distinction from Calsouthern USA. He holds a Bachelor's and Master's of Science in Computing, with a specialization in Software Engineering. His research interests include management, IoT, six sigma, knowledge management, information technology, economics, organizational change, facial recognition, biomedical devices, and computer science. Dr. Tariq has a passion for learning and development, project management, and training that made him achieve 4 patents.