International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

Integrating Fuzzy Concepts to Design a Fuzzy Data Warehouse

Kuyunsa Mayu Alain[a*], Kasoro Mulenda Nathanael[b], Mabela Matendo Rostin[c]

[a]*University of Kinshasa, Faculty of Sciences, Regional Center of Doctoral Education in mathematics and Computer Science, Kinshasa, D. R. Congo*

[b,c]*University of Kinshasa, Faculty of Sciences, Department of Mathematics and Computer Science, Kinshasa, D. R. Congo*

[a]*Email: alainkuyunsa2013@gmail.com*

[b]*Email: Kasoro_mulenda@yahoo.fr*

[c]*Email: mabelamatendorostin@gmail.com*

Abstract

In this study, we attempt to design a fuzzy data warehouse. The classification elements to design the concerned fuzzy data warehouse are presented through the following tasks: identification of the target attribute, the linguistic terms, attribute of class membership, definition of the functions of membership, degree of attribute membership, definition of table of fuzzy classification and fuzzy membership. From all the above tasks, we present a method allowing to design a fuzzy data warehouse data. What enables us to design, starting from a traditional data warehouse, a fuzzy data warehouse, which takes into account, inaccuracy and other uncertainties inherent in the digital data constituting the data warehouse. The Fuzzy logic enables us to deal with this type of data without affecting the base of this data warehouse.

Keywords: Target Attribute; Class Membership Attribute; Membership Degree; Membership Attribute Degree; Fuzzy Classification Table; Fuzzy Membership Table.

1. Introduction

The major transfers, which affect the world economic system, have marked repercussions at the regional and national levels. This evolution is irreversible. It is thus necessary to consider globalization as an essential data to take into account in the company strategy. Vis-a-vis globalization and with growing competition, the decision-making became crucial for company managers.

----

* Corresponding author.

The effectiveness of this decision-making is based on the provision of relevant information and adapted tools. Companies' task is to effectively exploit significant volumes of information, coming either from their operational systems, or external environment, to support decision-making. The traditional systems have been proven inadequate to such activity. In order to mitigate this disadvantage, of the decisional systems were developed. Most of these systems are based on a centralized storage space, called data warehouse whose role is to integrate and store useful information for the decision makers and also to preserve the data history to support analyses during the decision-makings. These systems of data warehouse receive data for the analysis of the surrounding operational systems. Because of the complexity and number of operating systems, the data received by the data warehouse are heterogeneous. This is why the data must be homogenized in at the first stage before the data warehouse starts processing them. The quantity of data, which must be processed in a data warehouse, increases daily and it is transformed into difficult tasks for the administration and analysis. Furthermore, the data coming from the operational systems are often incomplete, vague or uncertain (dubious). This quality problem cannot be completely eliminated in the data preprocess stage. Consequently, a certain quantity of inaccuracies influences directly the analysis and the decision-making based on information of a data warehouse. The fuzzy theory, proposed by Zadeh in [1] can face the lack of clearness, with uncertainty and imprecision. Contrary to probability systems, it is also optimized to interpret the imprecision in the human language and reasoning. Consequently, the application of fuzzy logic on technologies of data warehouse improves the analysis of the data and, this leads to a better decision-making. In this study, we propose an integration of fuzzy concepts in a classic data warehouse to obtain a fuzzy data warehouse. We present the steps to be followed to design a fuzzy data warehouse. This study is organized as follows: at section 2, the fuzzy data warehouse concept is presented. Section 3 presents the model Meta and the modeling method of a fuzzy data warehouse. Section 4 describes classical data warehouse and shows the fuzzy integration concepts to obtain a fuzzy data warehouse.

2. Fuzzy Data Warehouse

The concept of data warehouse was proposed by W.H. Inmon in 1990 to meet the needs for analysis of data which the transactional systems could not ensure [2, 5]. The models of data warehouses allow the design of data bases dedicated to the analysis. This analysis can be led to various levels of granularity, on huge amount of data represented in an aggregate way [3, 4, 6, 7,8]. Moreover, the models of fuzzy data bases are particularly interesting for the representation of the fuzzy and uncertain data and the taking into account of fuzzy requests. In this section, we present a concept of fuzzy data warehouse based on a structure of a Meta table.

2.1. The fuzzy data warehouse Concept

Fuzzy concepts can be integrated as the meta table structure without affecting the core of a DWH. Our approach is more flexible, since it enables the integration and definition of the fuzzy concept, without the need for redesigning the DWH core. The use of this DWH storing approach makes it possible to extract and analyse data simultaneously, in a classical form and a fuzzy manner. The purpose of this section is to represent some concepts of the meta tables, the modelling guidelines and the meta model of the fuzzy DWH approach. In order to integrate fuzzy concepts in a DWH, we start by identifying and analysing the elements which have to be fuzzily classified in the DWH. Such an element can be a fact in the fact table or the dimension attribute. An element which has to be fuzzily classified is called the *Target attribute* and the value range of that element instances is called the *Domain of attribute*. We define below some basic concepts linked to fuzzy DWH, concepts which will be used thereafter.

Domain of attribute: A set of potential values or the range of potential values of a dimension attribute or a fact is called *Domain of attribute or Universe of discourse* of a domain. The *Domain of attribute* of a Domain A is denoted $\mathbf{Dom_A}$.

Target attribute: A dimension attribute or a fact which is to be fuzzy classified is called a *Target attribute* (TA). Under fuzzy classification, the TA instances are classified over a set(S) represented by a linguistic variable. The linguistic variable consists of a set of non-numerical terms called *linguistic terms*, S = T_1,…, Tk.

The linguistic terms of a linguistic variable are captured in an attribute called *class membership attribute*.

Class membership attribute (CMA): a CMA for a target attribute TA, represented by CMA_{TA}, is an attribute which has a set of linguistic terms $T_1,\dots,T_k$ to which the target attribute may belong. In other words, for all the possible values of a TA (of domain of attribute, Dom_{TA}), there exists a corresponding relation to a CMA value. The values of CMA are the values in the set S.

All values of Dom_{AC} *to some fuzzy degree* belong to a CMA value. The degree of membership to a CMA value is called membership degree and it defines the relation of an instance TA to a CMA value.

The membership degree (MD$\in$ [0, 1]). It is the measure to which the values of a target attribute TA are linked to some linguistic terms $T_1,\dots,T_k$, respectively with the values of CMA. The MG is calculated using the membership function.

Membership function (MF): the MF of a CMA class is a function μ(TA) which is used to calculate the MD of a TA to a CMA μ: TA$\rightarrow$ [0,1]

The membership degrees generated by the membership functions are captured as membership degree attributes in the fuzzy data warehouse model. A membership degree attribute is defined as follows:

Membership Degree Attribute (MDA): the MDA of a TA is an attribute which has a set of MD of the TA. The value of a MD is calculated by a MF and is represented by μ_t(TA) = MG where MG is the membership degree of TA for the linguistic term t in CMA. An attribute which has to be fuzzily manipulated is prolonged by two meta tables. The first meta table contains a description of the fuzzy concept and the second meta table contains membership degrees of each instance with regards to the CMA. The two tables are defined as follows:

Fuzzy Classification Table (FCT): A table which consists of linguistic terms and and their unique identifiers is called FCT. There are two attribute tables which consist of an ID attribute and a CMA, where the ID attribute is a unique identifier of the table values. Formally,

FCT_{TA} = {identifier; CMA_{TA}}

A table which stores the values representing the measure to which a value is linked to a linguistic is called *Fuzzy membership table* (FMT). It is a four-attribute table: the Identifier table attribute, the target attribute identifier TA, the class membership attribute CMA_{TA} in the fuzzy classification table (FCT_{TA}) and the membership degree attribute (MDA) for TA. Formally,

FMT_{TA} = {identifier; identifier of TA; Identifier of CMA_{TA}; MDA_{TA}}

2.2. The Model of fuzzy data warehouse

The model of fuzzy DWH is a combination of four types of tables. They are dimension tables, fact tables, FMTs and FCTs.

A fuzzy DWH is a set of tables represented the following way

FDW = {Dim, Fact, FCT_{TA}, FMT_{TA}} Where

Dim = {a set of category attributes; level of category attributes}

Fact ={ a set of measures}

TA = {TA_1, TA_2,…,TA_n }, n is the number of FCAs.

It should be noted that the set of TAs is a subset of the dimension set and the set of facts. Formally, TA is a subset of Dim U Fact (i.e. $\forall TA_i \in$ Dim U Fact: $1 \leq i \leq n$).

For each TA_1, TA_2, ... ; TA_n :

FCT_{TAi} = {identifier; CMA_{TAi}} où $1 \leq i \leq n$.

FMT_{TAi} = {Identifier, Identifier of FCT, Identifier of TAi, MGA_{TAi}} où $1 \leq i \leq n$.

Guidelines for the modelling of the Fuzzy Data Warehouse

We present, here, a set of guidelines for the design of a fuzzy DWH model and the use of these guidelines for the elaboration of a meta model for the FDW using a real-life case.

i. Distinct fuzzy Classes / Linguistic conditions

A set of linguistic terms (also called fuzzy classes) is used to classify instances of a target attribute. In the simplest case, the linguistic terms are distinct, given that there is only one set of not repeated linguistic terms between them. In this case, one instance of a target attribute belongs to only one fuzzy class at a time and the degree of relation is measured by the MF. Formally,

TA—instance (1): Fuzzy classes (1)

Figure 10: Add a fuzzy classification table (FCT) and a fuzzy membership table (FMT) and a fuzzy membership table (FMT) for each target attribute TA, as indicated below.

TA—instance (1): Fuzzy classes (1)

But with different membership degree

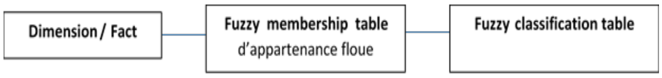

Figure 11: If an instance of a TA belongs to a fuzzy class, but with multiple degrees of membership, add a FCT and M number of FMTs, As indicated below, where M is the number of distinct membership degrees.

ii. Different linguistic conditions for a target attribute

An instance of a target attribute may belong to multiple linguistic terms as professional users can have more than one set of classes to which an instance of target attribute may belong i.e. multiple fuzzy classes and multiple MFs. Formally,

TA—instance (1): Fuzzy classes (M)

Guideline3. If an instance of a TA belongs to multiple fuzzy classes, but with the same membership degree, add M number of FCTs and a FMT, where M is the number of distinct linguistic terms. **Guideline 4.** If an instance of a TA belongs to more than one fuzzy category with different membership degrees, add a number M of FCTs and FMTs, as indicated below, where M is the number of distinct linguistic terms, and one FCT is linked to one FMT at the most.

3. Meta Model and method of modeling of a fuzzy data warehouse [13, 17, 19, 20]

3.1 Meta Model

According to Harel and his colleagues Al. [16], a meta model defines the elements of a conceptualization, as well as their relationships. Figure 1 shows the meta model of the proposed fuzzy DWH in which the right side shows the meta model of the classical DWH. The left side shows how the fuzzy concepts are integrated with a classical DWH as a structure of meta tables.

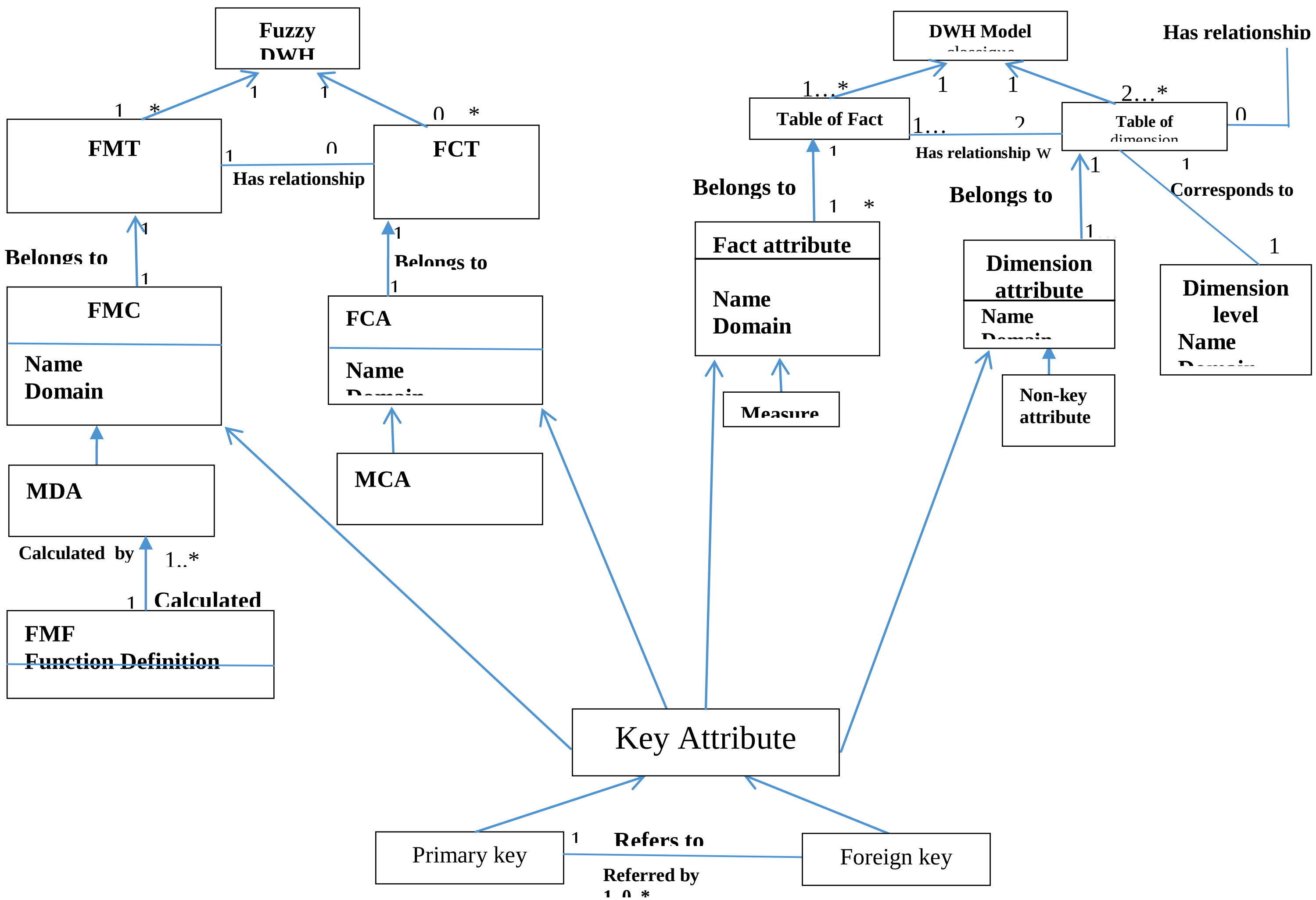

Figure 1: Meta model of fuzzy DWH

The model class of a DWH in the meta model refers to a DWH schema which is made up of one or more fact tables, and two or more dimension tables. A fact table is located at the centre of a DWH model and it essentially captures the commercial measures of the process (Kimball [17]). The relationship with dimension tables is realized with the help of fact attributes. A fact attribute could be a measure (also called fact) or a key attribute (primary or foreign key). A measure (a sub-class of the fact attribute) captures critical values of a business process i.e. whereas a set of key attributes are used to capture the relationship with dimension tables. In a classical DWH, two or more dimension tables surround a fact table. A fact table can also be linked to one or more other dimension tables to form hierarchies. In this case, each dimension table is at a different level of hierarchy (in order to comply with the snowflake schema). The level of hierarchy is referred to by the class of the dimension level in the meta model. A dimension table contains some dimension attributes which represent the category attributes or key attributes of a dimension table. The key attributes capture the relationship between dimension tables, respectively between fact and dimension tables. Moreover, other non-key attributes characterize the category attributes of a dimension table [19]. The model class of fuzzy DWH considered in the meta model refers to the fuzzy concept integrated within a DWH. For each target attribute identified a model of fuzzy DWH can be added. Thus, a classical model of DWH can have more than one fuzzy DWH model. Fuzzy concepts may exist without a linguistic variable. These fuzzy concepts are represented by fuzzy DWH models without a FCT. Each FCT has a relation to one or more FMTs. Therefore, the fuzzy DWH model is made up of one or more FMTs and zero or more FCTs. A FMT is built of FMAs. A FMA might be a key attribute to denote primary key or foreign keys. A second type of FMA is the MDA.The instances of the MDA are calculated by the FMFs of the fuzzy DWH model. The FCT contains the fuzzy CMAs which can be, similarly to FMAs, key attributes. Furthermore, the fuzzy class membership attributes can be a class membership attribute that describes the linguistic term of the fuzzy concept.

3.2 A method of modelling a fuzzy data warehouse [19]

In order to create a fuzzy DWH, a method is shown that guide the translation of a warehouse of a crisp DWH into a fuzzy DWH. The input of the method is a classical DWH and the output is a fuzzy DWH. The process is broken down into two phases: in the first phase, the first phase elements of classification are defined and in the second phase, the fuzzy DWH is built. Figure 2 shows the tasks and order in which they are carried out.

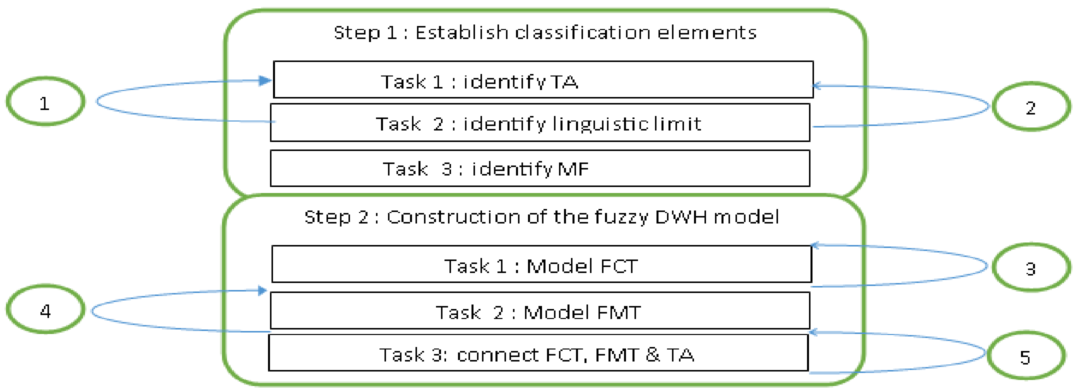

Figure 2: Graphical overview of the modelling method of a fuzzy DWH

Elements Classification

The purpose of this stage is to define the elements of classification which are used in the second stage to build the fuzzy DWH model. It involves three tasks: identify the target attribute, identify the linguistic terms and define the membership functions. Here are the details: **First task:** This task consists in identifying what has to be classified i.e. the TA which contains the values destined to be classified fuzzily. This will be done in a way that takes into account the input of the end user. In the simplest, a TA is identified. For Example 1, consider product price as a TA. **Second task:** It consists in determining how the values of the identified TA should be classified i.e. identifying the set of linguistic terms which are used to classify the instances of a TA. Repeat this task for all TAs. It is showed by Loop-through 1 in Figure 2. There are two possibilities:

***Case 1*–Distinct linguistic conditions:** it is the simplest case in which the linguistic terms are distinct i.e. there is only one set of linguistic terms. Formally,

TA— instance (1): Fuzzy classes (1)

For the example of the product price, let us consider the set of linguistic terms for {price_high; price_average; prix_low}.

***Case 2*–Different linguistic terms for a TA:** it is a case where there are more than one set of linguistic terms to classify the TA instances. In this case, instances of the TA belong to more than one linguistic term, as identified by professional users. Formally,

TA— instance (1): fuzzy classes (M)

For the example on the product prices, consider that the following sets of linguistic terms are identified. These sets are {price_low; price_average; price_high} and {cheaper; cheap; expensive}. The linguistic terms might already exist in a classical DWH modelling form of instances of a dimension category. In that case, these terms can be used for classifying the cases of TA. **Third task:** it consists in defining a MF (denoted μ) for each linguistic term. It is done in such a way that the values can be determined over a scale of 0 to 1. Repeat the task for each identified linguistic term. It is showed by Loop-through2 in Figure 2. It could be the case only for different users i.e. a TA belongs to the same set of linguistic terms with different membership degrees. The case is as follows:

The case is as follows:

***Case 3*–Different MDs for the same linguistic terms**: it is a case in which an instance of a TA belongs to a linguistic term with different membership degrees. It can be due to the fact that multiple business users have different interpretations of a single instance of a TA i.e. the multiple MFs are used for a TA. Formally,

TA— instance (1): fuzzy classes (1)

But with different membership degrees.

4. Application of the data warehouse

This paragraph presents a case study in the measurement of the output by using a brewery company. The case study aims at underlining the advantages of a warehouse of fuzzy data during the classification of the elements in a warehouse of data. First of all, the brewery company and its warehouse of initial and traditional data are presented in sub-section 4.1. In sub-section 4.2, the vague concepts are applied to the warehouse of data in order to build a warehouse of vague data.

4.1. The company of brewery product

We apply the method described here in a brewery company in Democratic Republic of Congo. This company is a manufacturing unit having its actions on all the Congolese territory. It produces and markets its products according to a certain logic imposed by the manager of this company. It works in collaboration with its customers. Each Monday, Tuesday, Thursday and Friday of the week, it puts in circulation its vehicles in order to supply its deposits. The deposits are on the one hand the customers for the brewery company and on the other hand, of the salesmen close to the ultimate consumers. The company delivers its products in racks, these racks are measuring units on which any calculation carried out on the sales of the company is based. All the deposits are equipped with products or the customers can come to buy. For each transaction of purchase a customer is registered in the warehouse of data. To measure the performance of the company, the company installed a warehouse of data. The company uses to evaluate its performance the sale (sales turnover) of each transaction of purchase. For the transaction of information on the customers, the product and the deposit are seized. Figure 3 presents the model in flake of the data of the marketing of the products of the warehouse of the data.

4.2. The fuzzy Integration of the concept dataWarehouse

In addition to the traditional analysis in the warehouse of the data, the brewery Company needs some features which are available by using the fuzzy concept.

Dimension product

The products are classified in various prices. In the warehouse of the traditional data, the products always have different prices. With classification in the warehouse of the data, the company classifies the product with a concept. On the basis of method presented in sub-section 3.2 initially the category "attribute of product" is defined like target attribute.

The second stage consists in identifying the linguistic terms. In this case, the linguistic terms are various prices to which the products belong. For the price of the product, the warehouse considered the linguistic terms for higher, price average and lower price.

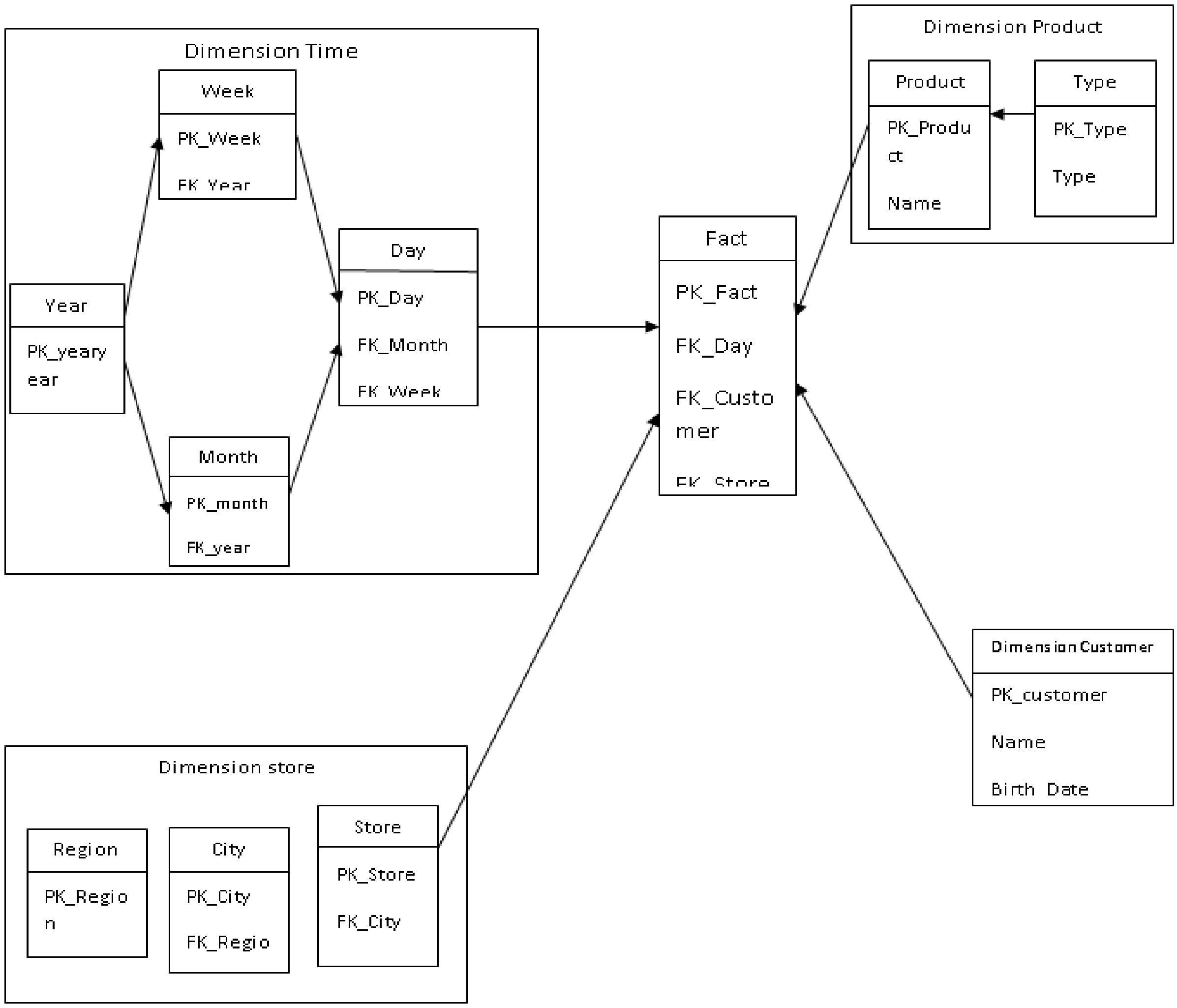

Figure 3: Classical Datawarehouse

By applying the third task of the method of modeling, the composition of the functions μ_{low}, $\mu_{average}$, μ_{high} becomes:

$$\mu_{low}(product\_price) = \begin{cases} if \quad \text{product\_price} \leq 10000, \text{MD\_PriceGroup} = 1 \\ if \quad 10000 \prec \text{product\_price} \prec 12000, \text{MD\_PriceGroup} = 1 - \left(\dfrac{\text{product\_price}}{12000}\right) \\ \text{else} \qquad , \text{MD\_PriceGroup} = 0 \end{cases}$$

$$\mu_{average}(product\_price) = \begin{cases} if \quad \text{product\_price} \geq 18000, \text{MD\_PriceGroup} = 0 \\ if \quad \text{product\_price} = 12000, \text{MD\_PriceGroup} = 1 \\ \text{if} \quad 12000 \prec \text{product\_price} \prec 18000, \text{MD\_PriceGroup} = \dfrac{18000 - \text{product\_price}}{18000 - 12000} \\ \text{if} \quad 10000 \prec \text{product\_price} \prec 12000, \text{MD\_PriceGroup} = \dfrac{\text{product\_price}}{12000} \\ \text{else} \qquad , \text{MD\_PriceGroup} = 0 \end{cases}$$

$$\mu_{high}(prodcut\_price)=\begin{cases} if \quad \text{product\_price} \quad \geq 18000, \text{MD\_PriceGroup}=1 \\ if \quad \text{product\_price} \quad \leq 12000, \text{MD\_PriceGroup}=0 \\ \text{else} \qquad\qquad\qquad , \text{MD\_PriceGroup}=1-\left(\dfrac{18000-product\_price}{18000-12000}\right) \end{cases}$$

After having identified the attributes target, the linguistic terms and their functions of membership, the structure of the table meta for the fuzzy concept can be defined according to the second stage of the process of modeling. According to the first and the second task, a table of fuzzy classification and a table of fuzzy membership must be created. The table of fuzzy classification holds the kinds like attributes of membership of class. The table of fuzzy membership contains degrees of membership for each attribute targets corresponding to the attributes of membership of class. The size of a fuzzy belonging is a Cartesian product of the product table and a fuzzy classification table. For the third and last task, the table of fuzzy classification, the table of fuzzy membership and the target table of attribute must be related to each other. This last task is finished by bringing back the attributes of foreign key of the table of fuzzy membership with the attributes of primary key of the table of attribute of target and the table of fuzzy classification. figure 4 shows the product dimension and fuzzy concept of the product price.

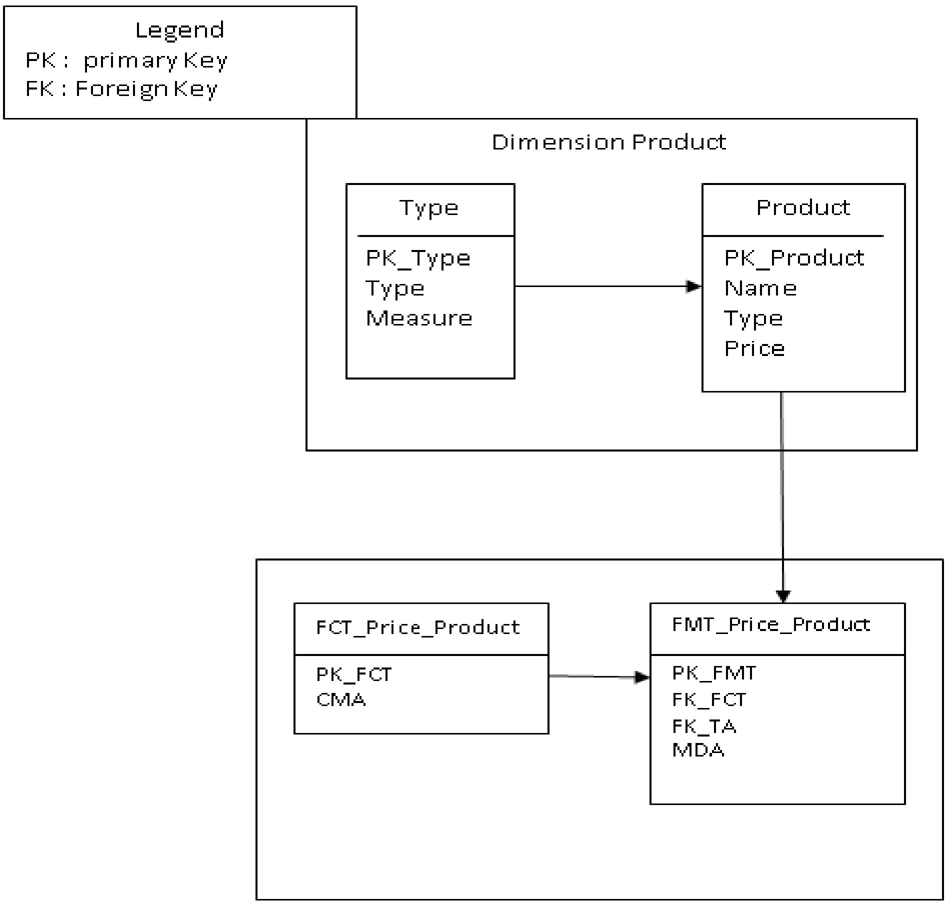

Figure 4: fuzzy concept of price product

Dimension customer

The brewery company is interested in the analysis of the amount of the sales on the basis of classification of the customer. The use of a fuzzy concept which classifies the customers in the good, means and bad customer is a more suitable classification. The fuzzy concept allows a classification finer grain of the customers. The category of attribute customer in dimension customer is defined like target attribute in the first task of the method of modeling. In the second task, a set of linguistic terms is defined as {good, moderate, bad}. An adaptive fuzzy concept is used and the functions of membership are defined in the third tries as follows:

$$lcr : lowest\ customer\ revenue$$
$$hcr : highest\ customer\ revenue$$
$$cr : customer\ revenue$$
$$base = hcr - lcr$$

$$\mu_{good}(cr) = \begin{cases} cr - lcr \geq 0{,}8 \times base, MD_{cr} = 1 \\ cr - lcr \leq 0{,}6 \times base, MD_{cr} = 0 \\ else, \qquad MD_{cr} = \dfrac{cr - lcr - 0{,}6 \times base}{0{,}8 \times base - 0{,}6 \times base} \end{cases}$$

$$\mu_{\mathrm{mod}erate}(cr) = \begin{cases} cr - lcr \geq 0.8 \times base, MD_{cr} = 0 \\ cr - lcr \leq 0.2 \times base, MD_{cr} = 0 \\ 0.4 \times base \leq cr - lcr \leq 0.6 \times base, MD_{cr} = 1 \\ 0.2 \times base \leq cr - lcr \leq 0.4 \times base, MD_{cr} = \dfrac{cr - lcr - 0.2 \times base}{0.4 \times base - 0.2 \times base} \\ else, \qquad MD_{cr} = \dfrac{0.8 \times base - cr - lcr}{0.8 \times base - 0.6 \times base} \end{cases}$$

$$\mu_{bad}(cr) = \begin{cases} cr - lcr \geq 0.4 \times base, MD_{cr} = 0 \\ cr - lcr \leq 0.2 \times base, MD_{cr} = 1 \\ else, \qquad MD_{cr} = \dfrac{0.4 \times base - cr - lcr}{0.4 \times base - 0.6 \times base} \end{cases}$$

The base is calculated by subtracting the lowest revenue from the highest revenue.

The amount of the sales by customer depends on the amount in fact. Consequently, each time that a new authority in fact of the amount in fact is added, the income of the customer is modified too. The recalculation of the income of the fuzzy concept to each change of the table of facts will affect the fuzzy performance of the warehouse of data in a negative way. So the amount of the fuzzy concept of the customer is adapted on a monthly basis by the brewery company.

Moreover, the customers are classified according to their age, in order to analyze the products according to the age of the customers. The target attribute is the category of attribute customer like front. The customers are classified with a vague concept containing the old, average and young linguistic terms (second task of the method of modeling).

In the third task, the company defines the function of membership in the following way :

$$\mu_{old}(customer\ age) = \begin{cases} customer\ age \geq 65,\ MD_{customer\ age} = 1 \\ customer\ age \leq 40,\ MD_{customer\ age} = 0 \\ else, \qquad MD_{customer\ age} = \dfrac{customer\ age - 40}{65-40} \end{cases}$$

$$\mu_{middle}(customer\ age) = \begin{cases} customer\ age \geq 65,\ MD_{customer\ age} = 0 \\ customer\ age \leq 40,\ MD_{customer\ age} = 0 \\ 30 \leq customer\ age \leq 40,\ MD_{customer\ age} = 1 \\ 20 \leq customer\ age \leq 30,\ MD_{customer\ age} = \dfrac{customer\ age - 20}{30-20} \\ else, \qquad MD_{customer\ age} = \dfrac{65 - customer\ age}{65-40} \end{cases}$$

$$\mu_{young}(customer\ age) = \begin{cases} customer\ age \geq 30,\ MD_{customer\ age} = 0 \\ customer\ age \leq 20,\ MD_{customer\ age} = 1 \\ else, \qquad MD_{customer\ age} = \dfrac{30 - customer\ age}{30-20} \end{cases}$$

In the first task of the second stage of modeling, two tables of fuzzy classification are created for the fuzzy concepts and a table of fuzzy membership for each fuzzy concept is created. Each table of fuzzy membership will be the size of the Cartesian product of the table of fuzzy classification and the table of customer. In the third stage, the table target customer of attribute, the tables of fuzzy membership and the tables of fuzzy classification are connected the ones to the others by using the relations of foreign key. Figure 5 illustrates the diagram of the table of dimension customer, the fuzzy concept of the revenue customer and the fuzzy concept of the age customer.

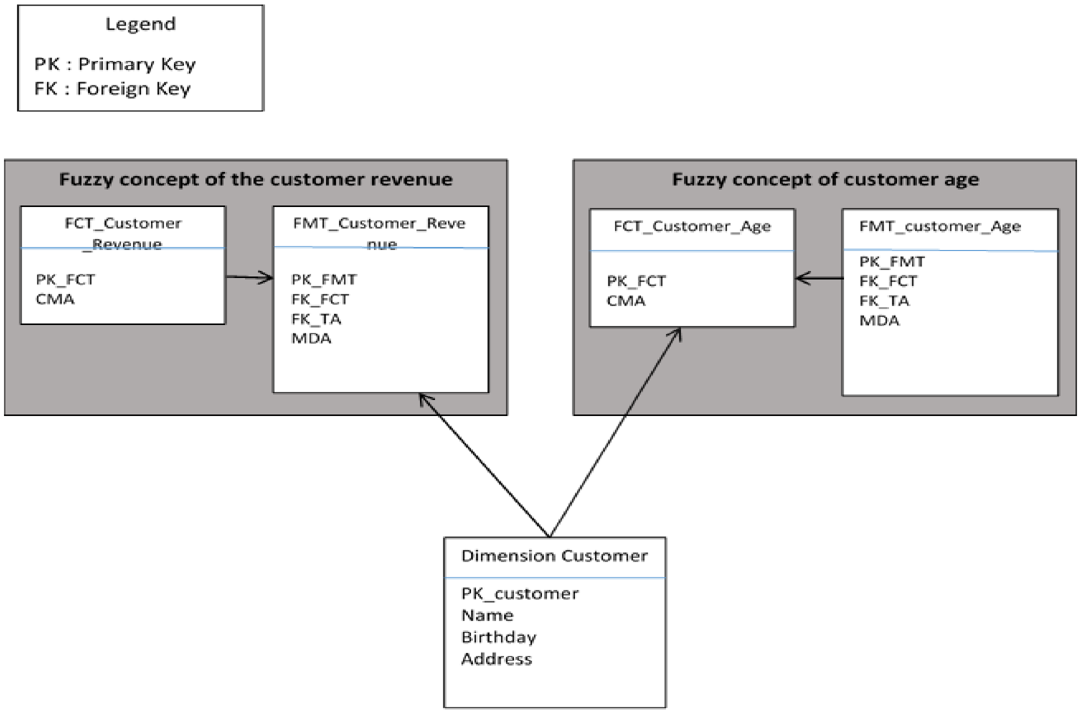

Figure 5: fuzzy concept of costumer revenue and the costumer age

Dimension store

The surface of each store is recorded in the warehouse of data. The brewery company is interested to know the performance of the small and large sales of store. Consequently, the company creates a fuzzy concept to classify the surface of store in large the, average ones and small surfaces. For the first task in the first stage, the table store is selected like target attribute and for the second task, the linguistic terms large, average and small are selected. The fuzzy concept has the characteristic of a limited vague concept that the company will not be able to rent the stores with a surface lower than 40 square meters or larger than 250 square meters.

For the three tasks, the functions of membership are defined as follows:

$$\mu_{big}(store\ surface) = \begin{cases} if\ \ store\ surface \geq 25,\ MD_{store\ surface} = 1 \\ if\ \ store\ surface \leq 15,\ MD_{store\ surface} = 0 \\ \text{else}, \qquad MD_{store\ surface} = 1 - \left(\dfrac{25 - store\ surface}{25 - 15} \right) \end{cases}$$

$$\mu_{medium}(store\ surface) = \begin{cases} if\ \ store\ surface \geq 25,\ MD_{store\ surface}\ = 0 \\ if\ \ store\ surface = 15,\ MD_{store\ surface}\ = 1 \\ if\ 15 \leq store\ surface \leq 25,\ MD_{store\ surface}\ = \dfrac{25 - store\ surface}{25 - 15} \\ if\ 8 \leq store\ surface \leq 15,\ MD_{store\ surface}\ = \dfrac{store\ surface}{15} \\ \text{else,} \qquad\qquad\qquad \mathrm{MD}_{store\ surface} = 0 \end{cases}$$

$$\mu_{small}(store\ surface) = \begin{cases} store\ surface \leq 8,\ MD_{store\ surface}\ = 1 \\ 8 \prec store\ surface \prec 15,\ MD_{store\ surface}\ = 1 - \left(\dfrac{store\ surface}{15}\right)1 \\ \text{else,} \qquad\qquad\qquad \mathrm{MD}_{store\ surface} = 0 \end{cases}$$

For the second step of the modeling method, the fuzzy classification table from task one and the fuzzy membership table from task two are defined to integrate this fuzzy concept in the fuzzy data warehouse. The fuzzy membership table of this limited concept has the size of the Cartesian product of the fuzzy classification and the store table. This is due to the fact that the company will always rent new stores in this range of surface, so all stores are classified. The relation of the tables is established by the foreign key relations. Figure 6 presents the table schema of the fuzzy concept of store surface.

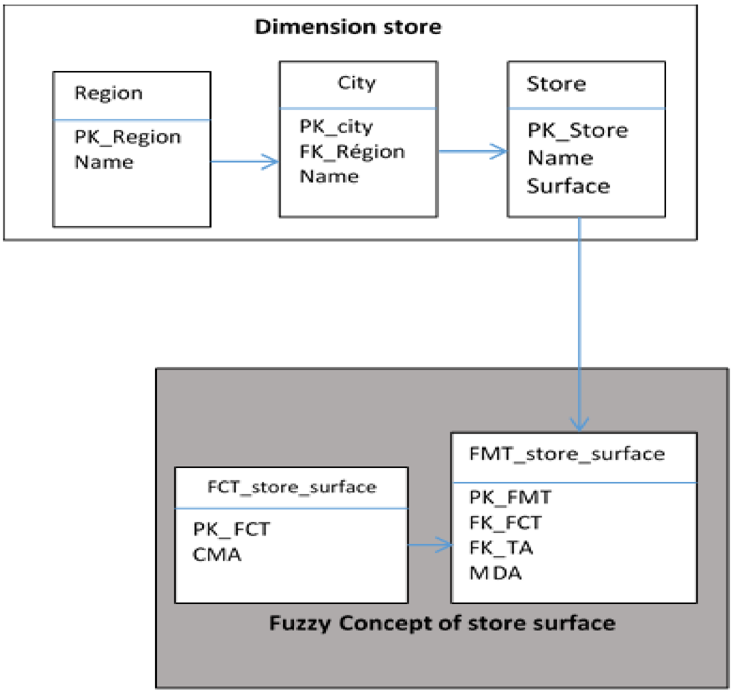

Figure 6: fuzzy concept of store surface

The fuzzy concept of store surface can be propagated to the dimension levels city and region. The propagated fuzzy concepts are adaptive fuzzy concepts and the membership functions calculate the membership degrees based on the specific domain of the target attributes. Therefore, the membership function uses a ranking on the ordered set of the domain of attributes. The membership functions are as follows:

$$S = \{x_1,...,x_n\}: \forall x_j \in dom_{TA} \wedge 1 \leq j \leq n \wedge x_j \leq x_{j+1}$$

$$rank(x_j) = \frac{j-1}{|S|-1}$$

$$\mu_{big}(store\ surface) = \begin{cases} rank(store\ surface) \geq 0.8,\ MD_{store\ surface} = 1 \\ rank(store\ surface) \leq 0.6,\ MD_{store\ surface} = 0 \\ \text{else}, \qquad \mathrm{M}D_{store\ surface} = \dfrac{rank(store\ surface) - 0.6}{0.8 - 0.6} \end{cases}$$

$$\mu_{medium}(store\ surface) = \begin{cases} rank(store\ surface) \geq 0.8,\ MD_{store\ surface} = 0 \\ rank(store\ surface) \leq 0.2,\ MD_{store\ surface} = 0 \\ 0.4 \leq rank(store\ surface) \leq 0.6,\ MD_{store\ surface} = 1 \\ \text{else}, \qquad \mathrm{M}D_{store\ surface} = \dfrac{0.8 - rank(store\ surface)}{0.8 - 0.6} \end{cases}$$

$$\mu_{small}(store\ surface) = \begin{cases} rank(store\ surface) \geq 0.4,\ MD_{store\ surface} = 0 \\ rank(store\ surface) \leq 0.2,\ MD_{store\ surface} = 1 \\ \text{else}, \qquad \mathrm{M}D_{store\ surface} = \dfrac{0.4 - rang(store\ surface)}{0.4 - 0.3} \end{cases}$$

The propagated fuzzy concepts reuse the fuzzy classification table. Hence, for each dimension level on which the fuzzy concept is propagated, an additional fuzzy membership table has to be defined. Figure 7 illustrates the fuzzy concept store surface with the propagated fuzzy concepts on level city and region. The fuzzy membership tables for the propagated fuzzy concepts are represented as dashed table objects.

Fact sale

The company creates also a classification on the amount of sale in the table of facts. According to the process of modeling, the table of information is selected in the first stage a task of the target attribute. The sales must be classified like high, middle and low. In the second task, the terms linguistic high, middle and low are defined. The sale is defined by the amount of sale per day. A fuzzy concept of open end with discrete functions of membership can be used in this case.

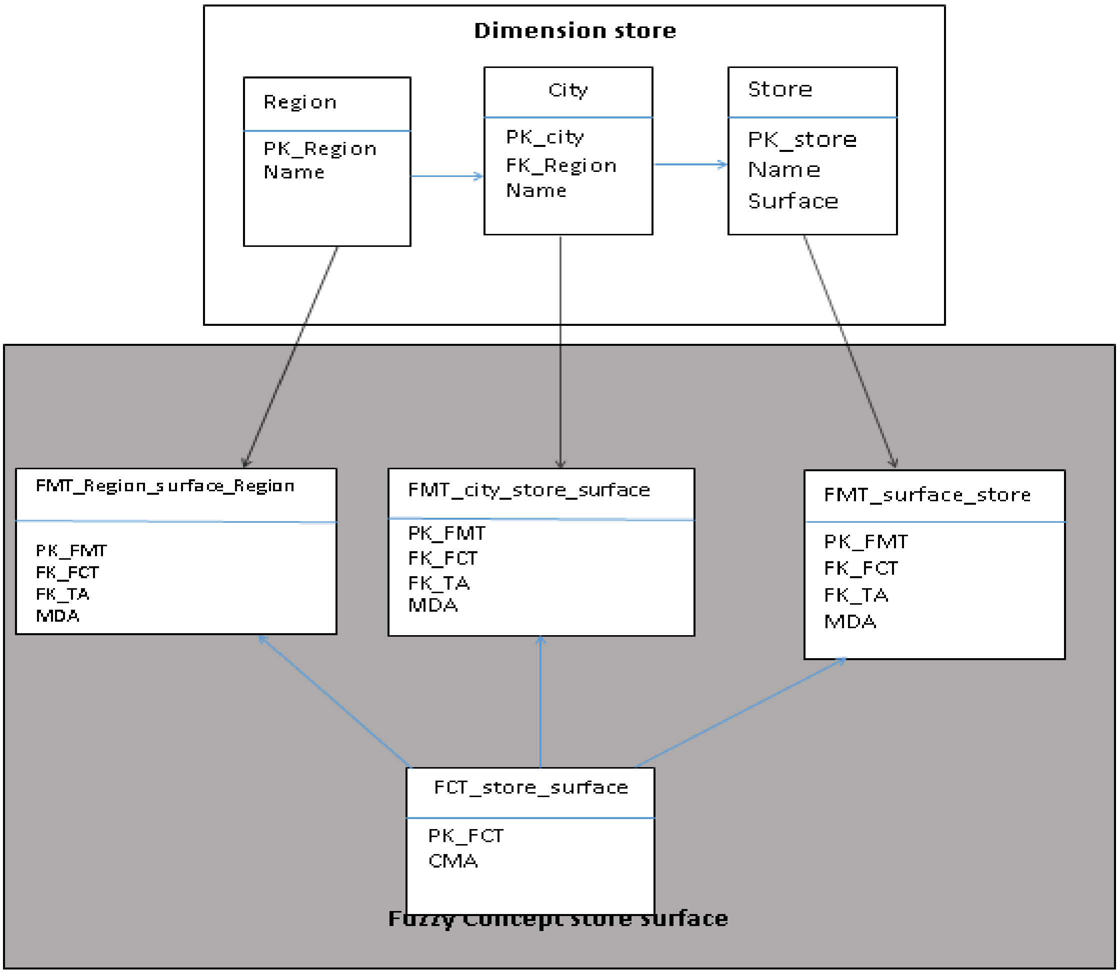

Figure 7: Propagated Fuzzy Concepts City Store Surface and Region Store Surface

In the three tasks, the functions of membership are defined as follows:

$$\mu_{high}(sale) = \begin{cases} if\ \ sale \geq 30,\ MD_{sale} = 1 \\ if\ \ sale \leq 15,\ MD_{sale} = 0 \\ else\ MD_{sale} = 1 - \left(\dfrac{30 - sale}{15}\right) \end{cases}$$

$$\mu_{middle}(sale) = \begin{cases} if\ \ sale \geq 30,\ MD_{sale} = 0 \\ if\ \ sale = 15,\ MD_{sale} = 1 \\ if\ 15 \leq sale \leq 30,\ MD_{sale} = \dfrac{30 - sale}{15} \\ if\ \ 5 \leq sale \leq 15,\ MD_{sale} = \dfrac{sale}{15} \\ else\ \ MD_{sale} = 0 \end{cases}$$

$$\mu_{low}(sale)=\begin{cases} if\ \ sale \le 5,\ MD_{sale}\ =1 \\ if\ \ 5 \le sale \le 15,\ MD_{sale}\ =\dfrac{sale}{15} \\ else\ \ MD_{sale}\ =0 \end{cases}$$

In the second stage of the method of modeling, a table of fuzzy classification for the task one and a table of fuzzy membership for the task two are defined. The foreign keys of the table of fuzzy membership are in relation to the primary keys of the table of facts and the table of fuzzy classification in order to build relations at the stage three. Figure 8 presents the diagram of the table of fuzzy concept of the sales.

The fuzzy concept of sale is a significant fuzzy concept for the brewery company. Consequently, it is propagated with all the attributes of each category of dimension. The total categories of the attributes exist in the data warehouse and for each category attribute a fuzzy membership table has to be created.

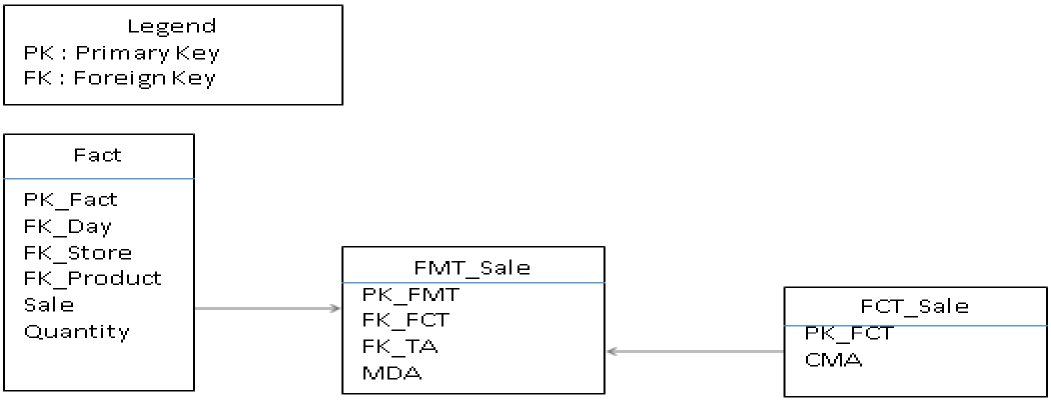

Figure 8: fuzzy concept of sale

The propagated fuzzy concepts are defined as adaptive fuzzy concepts and the membership functions are described, in store dimension, as a ranking of the domain of the target attributes. The membership functions for all propagated fuzzy concepts are as follows:

$$S=\{x_1,...,x_n\}:\forall x_j \in dom_{TA} \wedge 1 \le j \le n \wedge x_j \le x_{j+1}$$

$$rank(x_j)=\frac{j-1}{|S|-1}$$

$$\mu_{big}(sale)=\begin{cases} rank(sale) \ge 0.8,\ MD_{sale}\ =1 \\ rank(sale) \le 0.6,\ MD_{sale}\ =0 \\ \text{else}, \qquad \text{MD}_{sale}=\dfrac{rank(sale)-0.6}{0.8\text{-}0.6} \end{cases}$$

$$\mu_{medium}(sale)=\begin{cases} rank(sale)\geq 0.8,\ MD_{sale}=0 \\ rank(sale)\leq 0.2,\ MD_{sale}=0 \\ 0.4\leq rank(sale)\leq 0.6,\ MD_{sale}=1 \\ \text{else,} \qquad \mathrm{M}D_{sale}=\dfrac{0.8-rank(sale)}{0.8-0.6} \end{cases}$$

$$\mu_{small}(sale)=\begin{cases} rank(sale)\geq 0.4,\ MD_{sale}=0 \\ rank(sale)\leq 0.2,\ MD_{sale}=1 \\ \text{else,} \qquad \mathrm{M}D_{sale}=\dfrac{0.4-rank(sale)}{0.4-0.3} \end{cases}$$

The brewery company has decided to consider the fuzzy concept of customer dimension as a distinct fuzzy instead of redefining it as propagated fuzzy. Fuzzy sale concept is still propagated at the customer dimension. The customer dimension can be classified into two different revenue concepts.

The target attribute revenue of the base fuzzy concept revenue is modified as soon as new data is added into the data warehouse. The actual data from the source systems is loaded every night by the ETL process of the data warehouse.

The classification of the propagated fuzzy concept revenue has to be recalculated as soon this event takes place in order to always guarantee a real time classification. The brewery company decides to trigger the recalculation of the fuzzy concept revenue after every daily load.

Fuzzy Data Warehouse Schema

Finally, the data warehouse schema is extended with the meta tables of the fuzzy concepts. In Figure 9 the fuzzy data warehouse schema is presented including the meta tables for all fuzzy concepts shaded grey. For each fuzzy concept, the fuzzy classification table and the fuzzy membership tables are presented.

The linguistic terms for a fuzzy concept are stored in the attribute CMA in the fuzzy classification table and the membership degrees of a target attribute to the corresponding linguistic term is stored in the attribute MDA in the fuzzy membership table. The fuzzy membership tables of propagated fuzzy concepts are presented as dashed fuzzy membership table objects within the fuzzy concepts.

For readability, the thirteen propagated fuzzy membership tables of the fuzzy concept revenue are only symbolized by one fuzzy membership table object and the relations to the category attributes are omitted. For simplicity, the dimensions which do not have any fuzzy concept are only depicted as boxes with the dimension name. The structures of these simplified dimensions are retained as shown in Figure 3.

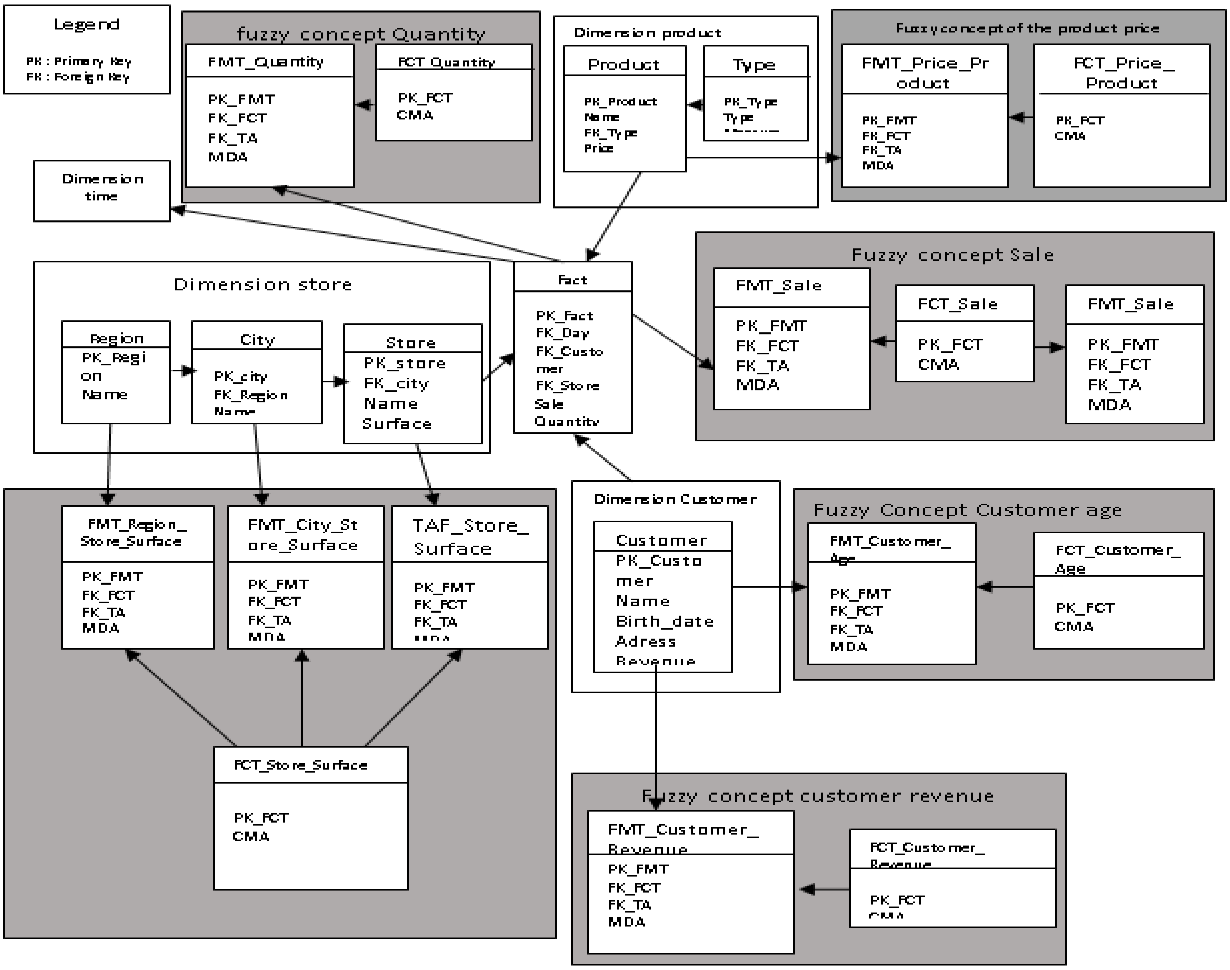

Figure 9: Fuzzy Data Warehouse Schema for the brewery company

5. Conclusion

Data warehouses are developed for an effective management of large volumes of data at ends of analysis and are used for the performance of the companies. In general, professional users of data warehouses interpret the numerical values in an incorrect way. The use of fuzzy logic makes possible to exploit in depth the contents of data warehouse and require an interpretation in significant terms, non numerical. The fuzzy representation allows the integration of the fuzzy concepts in dimensions and the facts, while preserving the structures of clear data using fuzzy logic. In this study, we proposed an approach of modeling of a data warehouse of a brewery company by taking into account data presenting some inaccuracies. We use the fuzzy set theory in order to allow the integration of the fuzzy concepts in dimensions, dimension product, dimension customer, dimension store and the fact sale, and that, while preserving the structures of clear data. This integration of the fuzzy concepts leads to the design of a fuzzy data warehouse.

References

[1] Zadeh, L.A. (1973). The Concept of has Linguistic Variable and its Application to Approximate Reasonning-1, Informations Sciences vol. 8/3, p. 199-249.

[2] Inmon W., Building the datawarehouse, John Wiley & Sounds, 1996.

[3] Kimball R., The Datawerouse Toolkit, John Wiley & Sounds, 1996.

[4] A. KUYUNSA M. , N. Kasoro M. and R. Mabela M., "Modelling a Structure of a Fuzzy Data Warehouse", Applied Engineering. Vol. 1, No. 3, 2017, pp. 81-89.

[5] Inmon W., "The operational Dated Blind", 1995, http://www.evaltech.com/wpapers/ODS2.pdf.

[6] [6] Imhoff C., `' The operational Dated Blind: Hammering Away ", DM Review, http://www.dmreview.com, 1998.

[7] Codd E., Codd S., Salley C., Providing OLAP (On-line Analytical Processing) to To use-Analysts: Year IT Mandate, Report, Arbor Software White Paper, 1993.

[8] Pendse N., Creeth R., Database explosion, Carryforward, The OLAP Carryforward, 1995.

[9] [9] Pedersen T. B., Jensen C. S., and Dyreseon C. E. Supporting Imprecision in Multidimensional Databases Using Granularities. International Eleventh Conference one Scientific and Statistical Database Management, 1999.

[10] Inmon W., Building the datawarehouse, John Wiley & Sounds, 1996.

[11] Kimball R., The Datawerouse Toolkit, John Wiley & Sounds, 1996.

[12] Jarke M., Lenzerini M., Vassiliou Y., V assiliadis P., Fundamentals of Data Warehouses, Springer-Verlag, 1998.

[13] Sapir L., A., Shmilovici, and Rokach L. In Methodology for the Design of Fuzzy Data Warehouse has. In Intelligent Systems, 2008. IS' 08. 4th International IEEE Conference, volume 1, 2008.

[14] Inmon W., "The operational Dated Blind", White Paper, www.billinmon.com/library/whiteprs/earlywp/ttods.pdf, 2000.

[15] Codd E., Codd S., Salley C., Providing OLAP (On-line Analytical Processing) to use-Analysts: Year IT Mandate, Report, Arbor Software White Paper, 1993.

[16] L. A. Zadeh. The Concept of has Linguistic Variable and its Applicqation to Approximate Reasoning – Part I. Information Science, (8): 199-249, 1975.

[17] D. Harel and B Rumpe. Meaningful Modeling: What' S the Semantics of "Semantics"? Computer, 37(10):64 – 72, October 2004.

[18] Ralph Kimball and Joe Caserta. The Dated Warehouse LTE Toolkit. Wiley Publishing, Inc, 2004.

[19] K. V. N. N. Pavan Kumar, P. Radha Krishna, and Supriya Kumar Of. Fuzzy OLAP Cubes for Qualitative Analysis. In Intelligent Sensing and Processing information, pages 290 –295, 2005.

[20] Heiko Schepperle, Andreas Merkel, and Alexander Haag. Erhalt von Imperfektion in einem Dated Warehouse. International Symposium: Dated Warehouse-system und Knowledge-Discovery, 2004.

[21] Vassiliadis P., Sellis T., `' A survey of logical Models for OLAP Databases ", SIGMOD Record, vol. 28, n°4, 1999.

[22] C. haudhuri S., Dayal U., `' An overview of Data Warehousing and OLAP Technology ", ACM SIGMOD record, vol. 26, n°1, p. 517-526, March 1997.

[23] Balaschka M., Sapia C., Hofling G., Dinter B., `' Finding you way through Multidimensional Dated Models ", Proc. International workshop one Dated Warehouse Design and OLAP Technology (DWDOT, in connection with DEXA), Vienna, Austria, p. 198-203, 1998.