

Toward Fuzzy Data Warehouses and Fuzzy Inference

Angélica Urrutia Sepúlveda¹ and Marcela Varas Contreras²

¹Universidad Católica del Maule, Chile
aurrutia@spock.ucm.cl

²Universidad de Concepción, Chile
mvaras@udec.cl

Abstract: A Data Warehouse is a data base that stores information for decision making. The characteristics of DWs make the data models and the strategy planning to be different from those used for operational data bases, requiring new techniques and design tools.

This work presents a study on the DW extension for the management of inaccurate information using the fuzzy conjunct theory. Here we display a methodology which incorporates a set of steps to design cube multidimensional components, generating a design star logic, or snow-flake arrangement. To apply, an SQL Server was used and it was extended to apply the FSQL scattered Fuzzy Database Relational engine, as well as incorporating some technical rules of knowledge using SCD, which are an extension of FSQL, to implement inference rules. Finally, we get a Fuzzy Data Warehouse, which implementation, by means of a data base, makes queries on decision making flexible

Key words: Fuzzy Data Warehouse, Fuzzy Data Base, Fuzzy Data Base inference.

1 Introduction

From a beginning, the data bases became fundamental tools for control and handling of commercial operations, reason why in a few years big companies and businesses got a considerable number of information stored in different data sources, reaching considerably higher size. In this way, the data bases have become part of a pillar in all the processes that participate in an organization, integrating different systems that provide quality information for the administration of such organization, which can cause some time problems in accessing, as critical factors in data administration. Figure 1 shows a graph of critical factors and a data base evolution.

A Data Warehousing (DW) is a data base that stores information for the decision making (Peralta 2001; Golfarelli, 1998). This information is built starting from data bases that register the organization's business transactions (operational bases). The DW's objective is to consolidate information from different operational bases making it available to understand the analysis of managerial type of data.

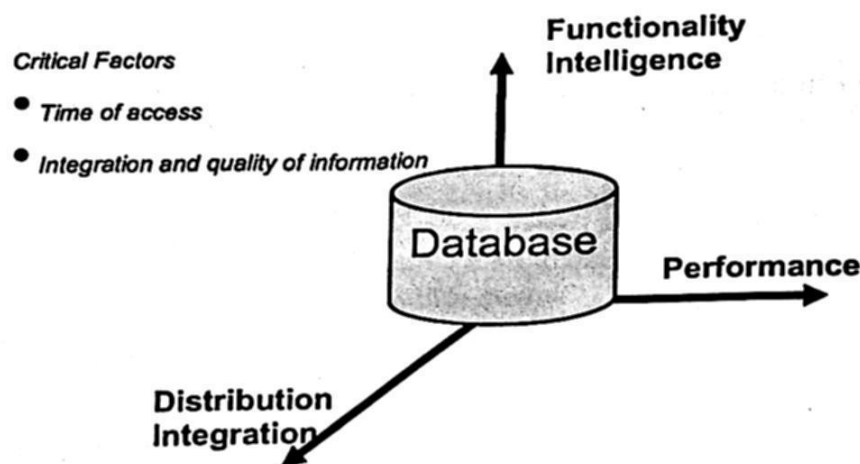


Figure 1: Database Integration and Critical Factors.

The interactive and immediate access to strategic information of an area of business should be treated as a priority. The predominant operations are not the transactions, as in the operational data bases, but the questions that involve great quantity of data and clusters of such transactions, which allow to support the decision making.

The characteristics of the DWs make the design strategies for the operational database not to be generally applicable to the design of a DW (Kimball, 1996; Inmon, 1996). The data models that represent the data stored in the DW are also different.

If the stored information on a Data Base is considered to be so precise as inaccurate and that the related systems of data bases work with the classical logic of covering only the type of precise information. Some proposals (Galindo, 1999) allow the use of the inaccurate information by means of Fuzzy Data Bases extending the traditional data bases with the theory of fuzzy groups, allowing the storage of information, treatment and consultation of inaccurate data.

Some jobs in fuzzy DW that are found in both investigations (Kumat et. al, 2005; Feng et. al, 2003) show the typified linguistic labels in some data, which allow to handle the information with extended fuzzy logic. In Server analysis Kumar discloses the data implementing with linguistic labels in SQL Server, while Ling reveals mathematical formality in generating cube's elements with a fuzzy logistics' extension, without reaching application

The investigation presented in this work is the result of the extension from a data warehouse to the inaccurate treatment as well as of its storage, in consultations that allow the managing of information with more flexibility in the decision making. The motors of data used here are SQL Server, FSQL and the inference software SCD, which is an extension of the FSQL. It is necessary to highlight that the implementation was one of the author's degree seminars.

The arranging consists of the theory elements of fuzzy groups, methodology and implementation of a fuzzy data warehouse, conclusions and bibliographical references.

Note: The software used by paper is in Spanish, for what the figures were shown in this language

2 Used Elements

The investigation presented in this document shows the work developed on an SQL Server, where its elements are expanded with the theory of fuzzy groups, types of diffused data and inference treatments using SCD for the DW components.

2.1 Theory of Fuzzy Groups

The concept "fuzzy" comes from the English term "fuzz" that means "confused, blurred, not defined or unfocused, although "fuzzy", in the academic world, is accepted as "ambiguous" and "vague", in the sense of human reasoning. The fuzzy logic, comes from the information that is managed in daily life and which is not always exact or precise, and due to this, it can possess uncertainty and inaccuracy.

A fuzzy group A on a speech universe " n " (finite interval or infinite inside which the fuzzy group can take a value) it is an even group, such as: $A = \{\mu_A(x) / x : x \in \Omega, \mu_A(x) \in [0,1] \in \mathbb{R}\}$ where $\mu_A(x)$ it is called the element's degree " x " belonging to the fuzzy group A . This level oscillates between the ends "0" and "1" in the domain of the real numbers:

$\mu_A(x) = 0$ indicates that " x " doesn't belong to the fuzzy group A at all.

$\mu_A(x) = 1$ indicates that " x " belongs completely to the fuzzy group A .

This definition allows to manage data with a degree of belonging to a group between the extremes "0" and "1", or to generate similar data, understanding the data types from a data base (Zadeh, 1965).

2.2 Linguistic Label

A **linguistic label**, in natural language, is such a word that expresses or identifies a fuzzy set, which may or may not be formally defined. Thus, the fuzzy set A membership function $\mu_A(x)$ expresses the degree in which x verifies the category specified by A .

With this definition, we can assure that in our everyday life we use several linguistic labels for expressing abstract concepts such as: "young", "old", "cold", "hot", "cheap", "expensive", etc. The intuitive definition of these labels, not only varies from person to person and depending of the moment, but it also varies with the context in which is applied. For example, a "high" person and a "high" building do not mean the same.

Example 1: If we express the qualitative concept "young" by means of a fuzzy set, where the x axis represents the discourse universe for "age" (in natural whole numbers) and the y axis represents the membership degrees in the interval $[0,1]$. The fuzzy set that represents such concept could be expressed as follows (considering a discreet universe):

$$\text{Young} = 1/0 + \dots + 1/25 + 0.9/26 + 0.8/27 + 0.7/28 + 0.6/29 + 0.5/30 + \dots + 0.1/34$$

The "age" (in total years) would be the discourse universe of "young". The linguistic label "young" would identify this fuzzy set represented by a membership function, if we consider an indiscreet discourse universe from others, such as "adult", "old", etc.

2.3 Types of Fuzzy Data

There are 3 types of attributes susceptible to inaccurate treatment (Galindo, 1999; Urrutia 2003)) and they are classified according to the type of domain underlying them and in that way, allowing to storage vague information or only allowing the imprecise handling of such data:

Type 1: Attribute that are traditional, without vagueness, but they also admit that in their domain it could define some linguistic label to be used in consultations.

Type 2: Attribute that admit data with and without indefinite in form of distribution possibility on an orderly underlying domain. Furthermore, it allows the representation of data type *Unknown, Undefined and Null*.

Type 3: Attribute that define some labels that are scaled with a defined similar relationship on them, in a way that this relationship shows the criterion by which they look alike on each couple of labels. The same as the previous type, this also allows to represent data labeled as *Unknown, Undefined and Null*.

2.4 Diffuse Expert System DCS (SCD)

DCS (Diffuse Control System) is a system created at the University of Malaga as a pre-graduate thesis (Escobar, 2003). It is an FSQL extension created by (Galindo, 1999) and it allows to produce inference rules and provides means to carry out data types T1, T2 and T3, with their respective linguistic labels.

It should be clear cut that such an expert system must be fed by an *Expert*, who is a person that has gradually acquired knowledge through a learning and experience period over a specific domain, or having acquired knowledge by using the system.

2.5 Data Warehouse Design

A Data Warehouse (DW) is an administration tool used for the decision making. According to (Inmon, 1996) a DW "is a group of data applied to topics, integrated, volatile and historical, organized to support a process of decision making", from which some transformation stages are considered, like:

1. Drawing out of data: It consists on drawing out data from the DB source and to load it into the DW.
2. Filtering: It consists on filtering unacceptable data into the DW.
3. Format or values modifying: It consists on adapting formats or values so that they comply with the guide lines defined in the DW.
4. Integration: It consists on integrating data coming from two sources.

The most used components in a DW design with Cubes or Hypercubes consisting of Dimensions (with Hierarchies formed by Levels) that are approaches for data analysis, independent Variables and axes in the hypercube, they are also the Measurements, which are values or indicators to analyze, dependent variables, besides variables in the dimensions intersection. On the whole, they are dimensions and measurements form a cube that are represented in a logical design called "shatters" or "snow-flakes", as it may be the case.

A DW has different technological alternatives: ROLAPs: they act directly on BD Relational. MOLAPs: they work on specialized storage. HOLAPs: they try to apply both strategies. Each one of them can be used in the design of a DW.

3 Fuzzy Data Warehouse Methodology and Implementation

Without a doubt (Salas, 2006), the methodology used in the construction of a Data Warehouse is a question of vital importance and it is for that reason that we propose a group of steps to develop a data warehouse that contemplates the extraction of data with the undefined, besides allowing a diffused inference motor. *The development methodology for the construction of the proposed Data Warehouse contemplates 10 stages, which are as follows:*

Situation

1. *Withdrawal and analysis of Requirements:* the necessary sources of data withdrawal are determined from the organization's information system (OLTP), as well as of external sources. Besides that, the user's requirements are established, this is: the Data Warehouse services, restrictions and objectives, obtained through interviews to users.
2. *T-A Transformation:* They are defined as linguistic labels that represent the attributes Type 1 and 2 and, in turn, they produce inputs for the generation of representing the attributes Type 3.
3. *Conceptual Design CMDM:* the conceptual Data Warehouse design is built on this stage, starting from the requirements specified by the users. As it's specified by the CMDM model, cubes are built from dimensions, criteria (standards) and the relationships grouping them.
4. *T-C Transformation:* the Linguistic Label transformations to Cube Dimensions are defined here.
5. *Cube Design:* Defines the activities subjected to analysis and the dimensions that characterize the activities for multidimensional modeling.
6. *T-B Transformation:* Considering the chosen labels from Step 2, the domain stages of each one of the sub-ensembles for attributes Type 1 and Type 2 and the similar relationships for attributes Type 3 that will be used by the inference motor.
7. *Logical design:* defines the outline of ROLAP or MOLAP, producing the logic star design and/or snow flakes and the ETL design (Extraction, Transformation, Load)
8. *Data Warehouse Inference:* defines the rules of knowledge, with acquired information through time and experts.

9. *Implementation*: construction of Data Warehouse Fuzzy and preparation of user's views through Systems Expert.
10. *Representation*: View of a Data Warehouse through a Front End application.

Figure 2 shows a description of the proposed steps for methodology that we have called MFDW (Methodology for Fuzzy Data Warehouse) numbered with each step and their corresponding sequence.

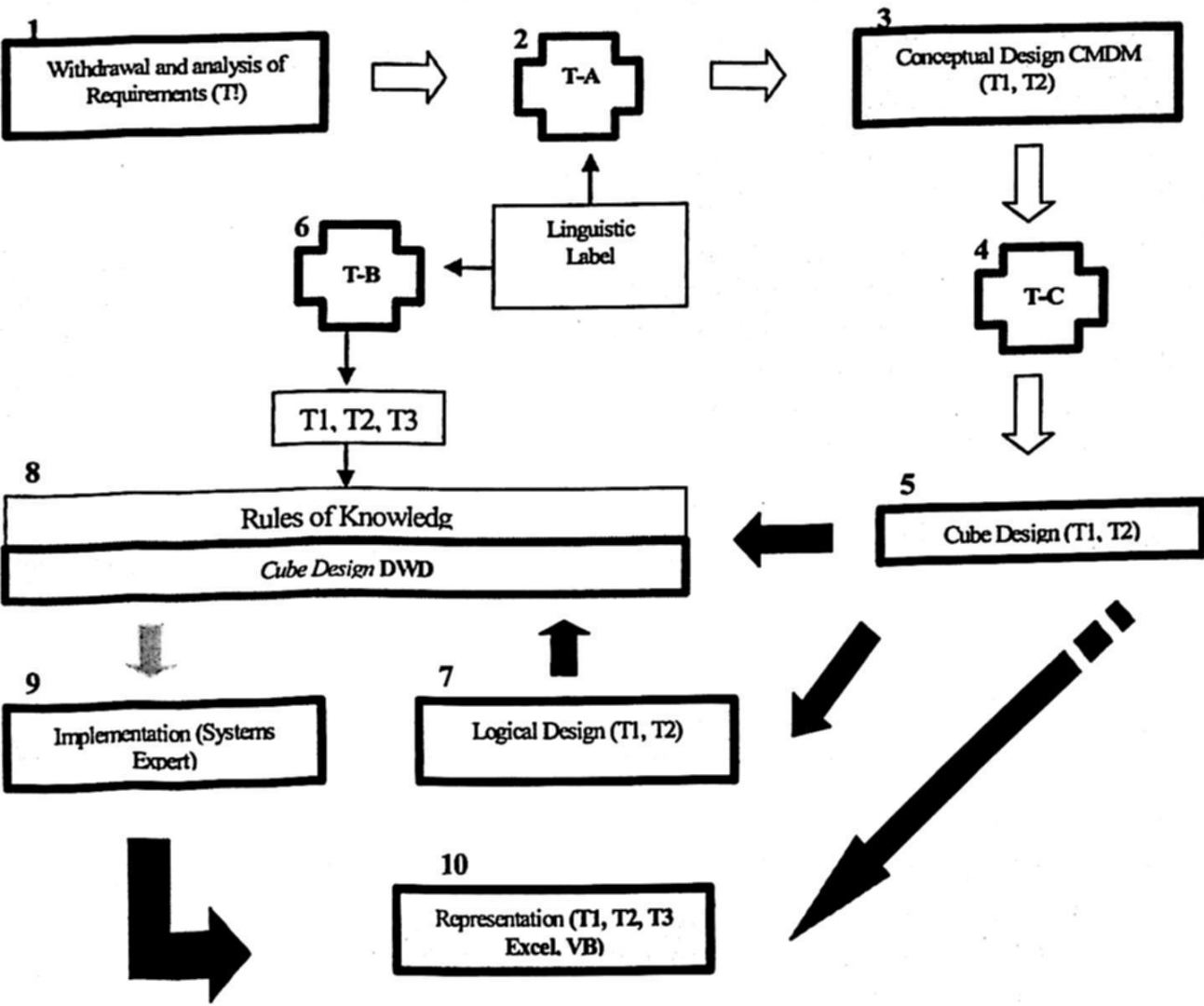


Figure 2: MFDW Methodology for the Development of a Data Warehouse.

Just as mentioned in the section 2.2 the undefined tried with fuzzy logic bears to three types of data T1 (Type 1), T2 (Type 2) and T3 (Type 3), each one of them with a different type of representation. Table 1 displays the stages proposed in the design and implementation of a DW, also, shows the incorporation of inference elements for the administration of information and decision making (Salas, 2006).

Next we show a case where the methodology MFDW is applied and that was part of a degree thesis in Engineering Sciences at the UCM (Universidad Católica Del Maule).

Table 1: Description stages MFDW according to data T1, T2, T3 and inference.

Stages	Type 1 Attributes	Type 2 Attributes	Type 3 Attributes	Inference
analysis of requirements	Receiving Data Reception	Define of data continue	Define of data not continue	None
Conceptual Design	Storing Data	Linguistic Labels	None	Knowledge Rules
Cube's Design	Dimensions, Measurements	Dimensions and Measurements	Dimensions and Measurements	None
Inference Motor	None	None	None	Knowledge Base
Logia Design	Snow Flakes, Star	None	,Snow Flakes, Star	Expert System
Implementation	Excel, Data Report, etc.	Excel, Data Report, Expert System	Expert System	Expert System
Presentation	Normal Data Warehouse	Diffuse or Normal Data Warehouse	Diffuse Data Warehouse	Diffuse Data Warehouse

3.1 Practical Case

The Chilean Telemedicine is in need of periodically getting reports with ECG (Electro-Cardiograms) received data. These reports are requested mainly from the Ministry of Health, which is the main Company client. On the other hand, these statistics are necessary to define the need of present operators at the call center on every schedule and, in this way, plan the shift systems, about which we may say that they have an incidence in the decision making of the Company operations. Also, the statistics throw the quantity of ECGs sent by each client, needed information to collect for the service (billing).

Actually, these reports are carried out through different statistics that are taken based on the ECGs. These statistics are made monthly, checking the received ECGs one by one and analyzing the different points of interest, such as ages, schedules of ECG arrivals, symptoms and histories, among others. An important statistics portion produced in this case allow to collect information for the decision making, being necessary to establish a store for data (DW) to register this information and the storing in time. Some of the data required in this DW have an imprecise behavior, so it is necessary to build a fuzzy DW and, in a special way, some inference rules, mainly for symptoms, so the decision making becomes more efficient.

3.1.1 Collection and Analysis of Requirements

It is the MFDW number one step, if not the most important. They were picked up based on consultation to experts and in the experience on the topic of one of the authors of this work. Not to go into details, in a simplified form, we show the user's requirements according to their *Gender Classification* (Number of ECG received, by sex, for each client), *Symptoms*, *Hour* (Requested ECG received by the time of arrival, number of operators per Schedule in function of demand), *History* (Patient's

history influence on consulting) and *Age* (Number of consulting appointments depending on the patient's age). Some indicators that can be measured are displayed next:

- Type of diagnosis depending on the patient's type, ECG's timetable, hospital to be sent.
- Deduce diagnosis based on certain rules of patient's behavior through time.
- Cardiovascular diagnosis starting from variables, such as hypertension, tobaccoism, etc.

The data base source has the related pattern, which is shown in the figure 3.

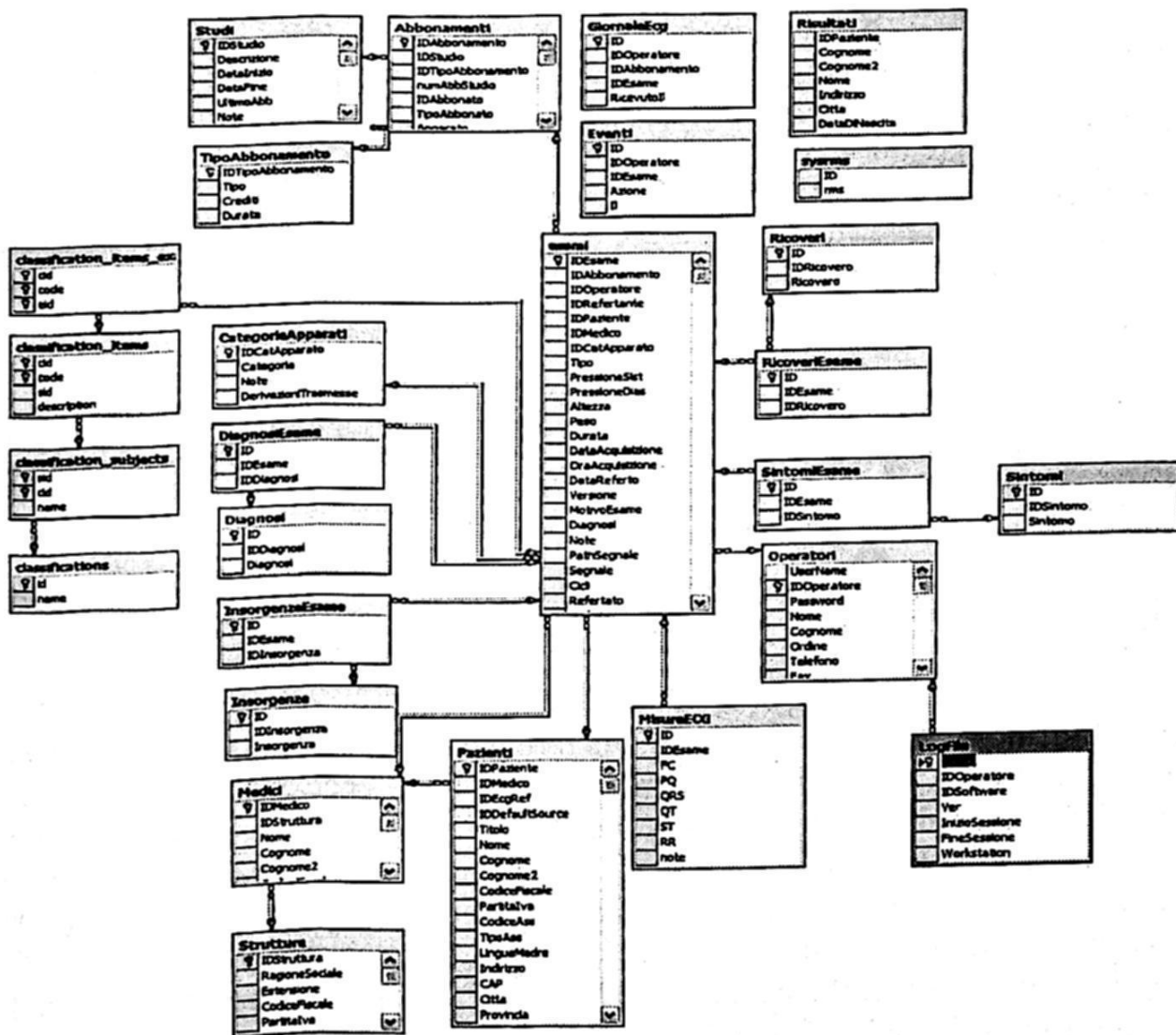


Figure 3: Data Base source (Salas, 2006).

3.1.2 Definition of Linguistic Labels (T-TO)

Under the classification of this Data Base outline the tables "classification_subjects, classification_items, classification_items_exams", were repeated in tables Diagnosis, Symptoms, History and Classification, to manage a descriptive name for such tables. On them, these processes are synchronized via trigger, so that every time an insert is

received in one of the classification tables "classification_subjects, classification_items, classificaton_items_exam", are repeated on of Diagnostic, History and Symptoms tables, which keep the same structure as the tables mentioned previously, but they add a relationship with defined tables to receive fuzzy data, through linguistic labels, which are described next.

STATURE: Linguistic Label: {Under, Normal, High}. Where, the representation is done the same way as in the age label, where the trapezoidal values are:

$$S(x) = (0, 1.45, 1.65, 2.15).$$

PHYSICAL STRUCTURE: Linguistics Label: {Small, Average, Big}

Where: Structure Index "Thin" of Corporal Mass is represented as:

$$P(x) = (0, 18, 25, 40).$$

AGE: Linguistics Label: {Boy, Young, Adult} where,

$L(x) = "1"$ If the Boy's Age < 8 years, " $(8 - \text{Boy's Age})/(10-8)$ " If the Boy's Age is more than 8 and less than 10 years, " 0 " If the Boy's Age is more than 10 years.

$L(x) = "0"$ If the Youth's Age is smaller than 8, or more than 30 years, " $(\text{Age of the Youth} - 8)/(10-8)$ " If the Youth's Age is between 8 and 10 years, " 1 " If the Youth's Age is between 10 and 28 years, " $(30 - \text{Youth's Age})/30-28$ " If the youth's Age is between 28 and 30 years.

$L(x) = "1"$ If the Adult's Age is 28 years, " $(28 - \text{Age Adult})/(100)-30$ " If the Adult's Age is more than 30 and less than 100 years, " 0 " If the Adult's Age is more than 100 years

Other labels, as time, behave in the same way described previously. Summer - winter - autumn and spring, have their respective months assigned. In the same way "hour", among others.

3.1.3 Conceptual Design.

CMDM will be used (Carpani, 2000) to perform the conceptual design. CMDM distinguishes among dimensions that identify reality objects and relationship dimensions that represent the existent multidimensional relationships among this objects.

- Even: A level representing a group of data
- Dimensions: They represent the approaches for analysis
- Relate Dimensional: They represent cross-links among dimensions.

DIMENSIONS: One of the requirements settles down as a discriminating approach to the diagnosis for the ECGs. This dimension is obtained from the chart of diagnosis. The dimension Diagnosis is composed by a level "diagnosis".

The user requests, within the requirements, to know the main symptoms for which an ECG is requested from the patient. The dimension Symptom is formed by a level "symptoms". The Patient dimension relates to the patient's data and it is composed by a level "patients".

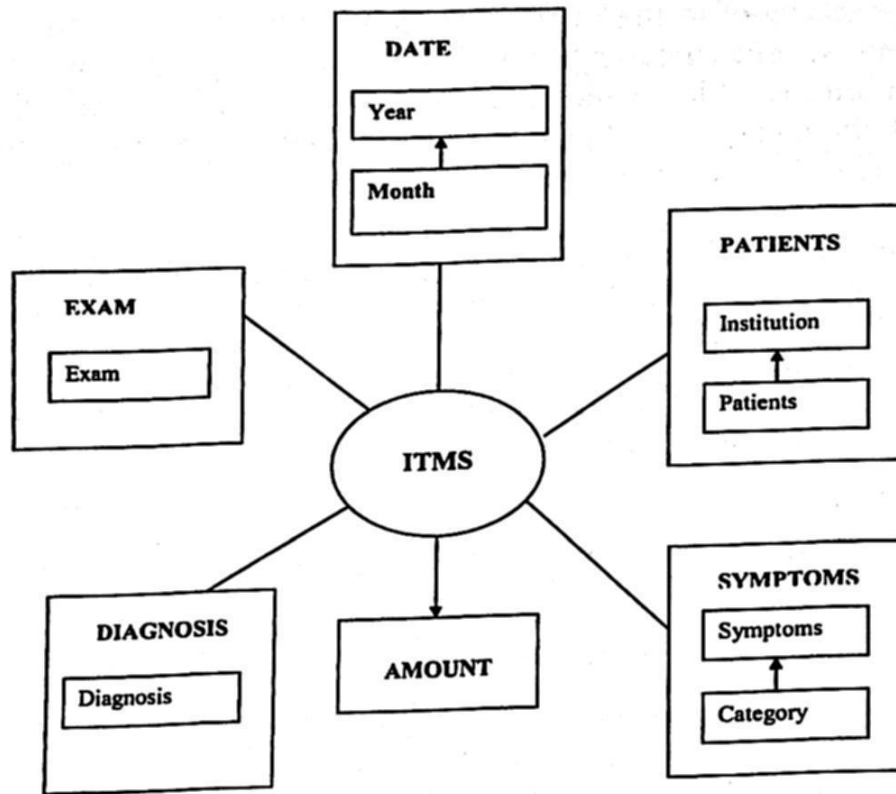


Figure 4: Graphic representation of Dimensional Relational ITMS.

The dimension exam is necessary to obtain the data from the exam. This dimension consists of a level "exam"

The dimension "date" is one of the requirements for the Data Warehouse development, since most of the data is consulted according to a date or a range of dates. This hierarchy dimension is formed by two levels "Year and Month".

The defined dimensions cross-links generated the dimensional relationship "Integrate" (in figure 4 equal ITMS), which links all the dimensions previously defined and from which it could be generated as many cubes as the required measures.

3.1.4 Changing Labels to Dimensions (T-C) and Cube Design.

The MFDW steps 5 and 7 are analyzed in this section, using the relationships shown in figure 4. The Diagnostic cube is believed to represent the measure *Diagnosis Quantity*, which is formed by the Dimensions Diagnosis with the hierarchy Category and the sub-hierarchy Symptoms and History; and the dimension Exam, with the patient's hierarchy.

This cube was implemented in the SQL Analysis Server, just as it's shown in the figure 5, only for accurate data. The cube is built from a star model and a snow flake is made with the diagnosis dimension hierarchies. All this derived from the conceptual model structure change (transformation) in figure 5 as a change to logic design. The implementation for inaccurate data is shown in the section 3.1.5.

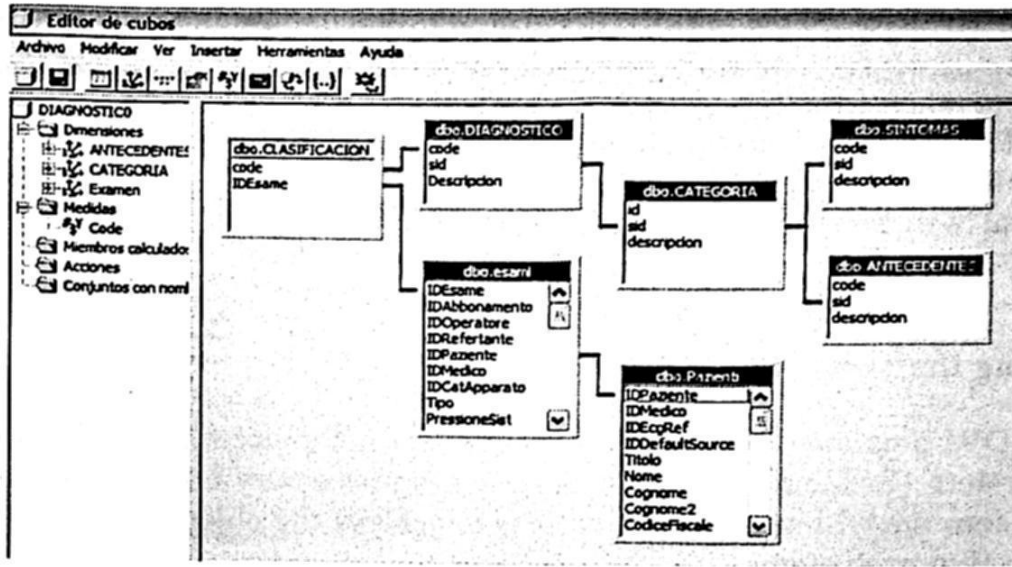


Figure 5: Star Shape Model for Cube Diagnosis.

3.1.5 Representation of a Fuzzy OLAP Cube in a SQL Server

The MFDW step 10 is shown like an extension of the cube generated in the section 3.1.4 using the motor of Galindo's FSQ. To implement FSQ in the SQL Server, which incorporates, as part of the catalog, the relational pattern shown in figure 6, called FIRST: Fuzzy Interface for Relational SysTems (Medina et.al, 1995; Galindo et. al, 2005) in Galindo's FSQ (Galindo, 1999).

Data-Base Implementing in FIRST and the FMB: The Data Base implementing in FIRST (figure 6) in the Data Base SQL Server allows to incorporate the way how the fuzzy and classic attributes are represented in a BDRD.

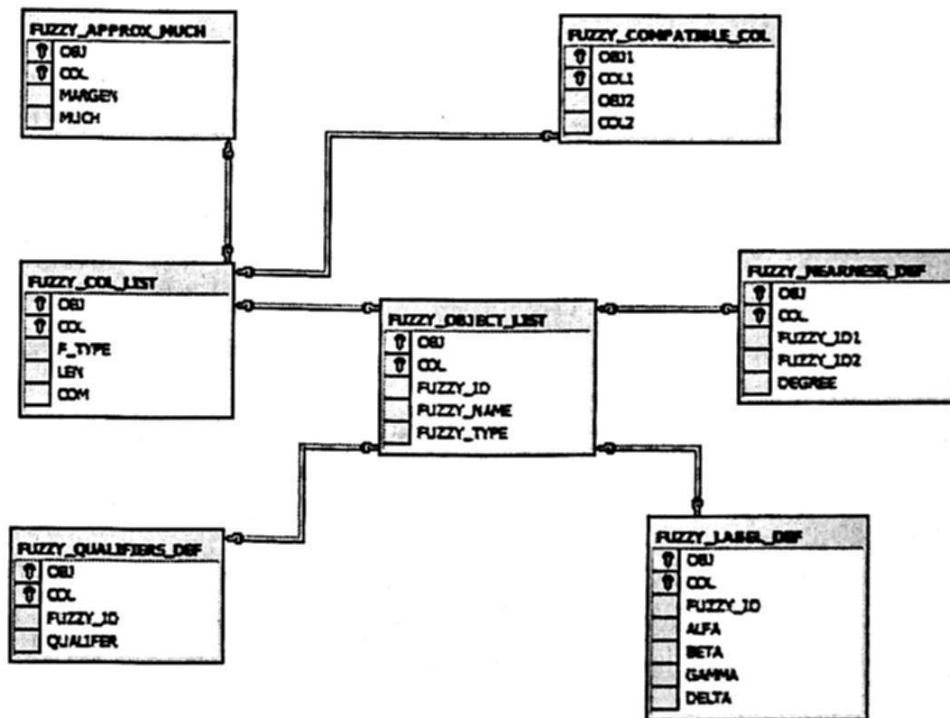


Figure 6: Tables in FIRST that incorporates to the SQL Server in Catalog.

Each one of these tables is filled according to the definition of the linguistic labels for data T1, T2 or T3. The filling of each one of them is discussed next (Salas, 2006) with the data taken from the Data Base sources.

The tables will be filled with the data of the labels shown in the section 3.1.2. With these definitions you are under conditions to implement the BDRD. On this respect, the data that will be stored in FIRST is the one that has been shown in the previous points.

3.1.6 Using the Fuzzy Control System

In the MFDW data steps 9 and 10 we will use a computer system called of Diffuse Control System (Escobar, 2003), DCS, to get the necessary inference for the system. To this system we will make modifications to achieve the objective, which takes us to develop an Expert System.

Such an Expert System should be fed by an expert who can be a person whose knowledge has been acquired gradually through a period of learning and experience in a specific field, or, fed by a system that offers him knowledge through its use (practice). In the application DCS, the attributes Type 2 and Type 3 will be implemented. For Type 2 we will use the symptoms and a patient's history, while in Type 3 we will use diagnosis (show section 2.3). The knowledge rules are obtained through the Data Warehouse.

To the traditional SQL language are added: the linguistic labels, the fuzzy comparators, the logical connectors, the execution threshold and the fuzzy constants (Galindo, 1999). In the application DCS to be used, this SQL extension is found in the code source, where each one of the comparators, connectors, etc., are defined (Salas, 2006)

Next, the development and the use of this system are shown. The first input data input was to create a Project applying DCS, where the user will be *Seminar*; the title, *Cardiology* and as observation, we will define *inference of Cardiovascular Diagnosis*.

Entrance Variables. We will use the Entrance Variable "*Hypertension*", called by their acronym HTA and *Tobaccoism*, as history of a patient. Both with fuzzy grade of incidence, Low, Average, and High. According to Step 2 of the MFDW pattern proposed in Step 3.1.2 the defined linguistic labels are now implemented.

Output Variables. As Output Variables we will use the possible Diagnosis, according to the inputs in this case. Our possible output variables will be the category *Necrosis* and inside those we will find *Isquemia*. The attributes type 3, *Diagnosis*, are noticed as **output attributes**.

We will create the rules of knowledge showing high HTA with High Tobacconism, getting Necrosis, corresponding to Isquemia Anteroseptal. Figure 7 A) shows a picture of this stage.

Inference of Diagnosis. The inference process of diagnostic is carried out through acquired knowledge providing diagnostic probabilities according to the input variables and to the rules inserted in the inference motor. Figure 7 B) shows a picture of this stage.

The system shown here was of great advantage to the final user, allowing to have pre-defined diagnosis and, over all, providing information more related to the human

treatment, such as uncertain data. A good portion of the grades of domain and the created linguistic labels were obtained from the System's Expert (Escobar, 2003).

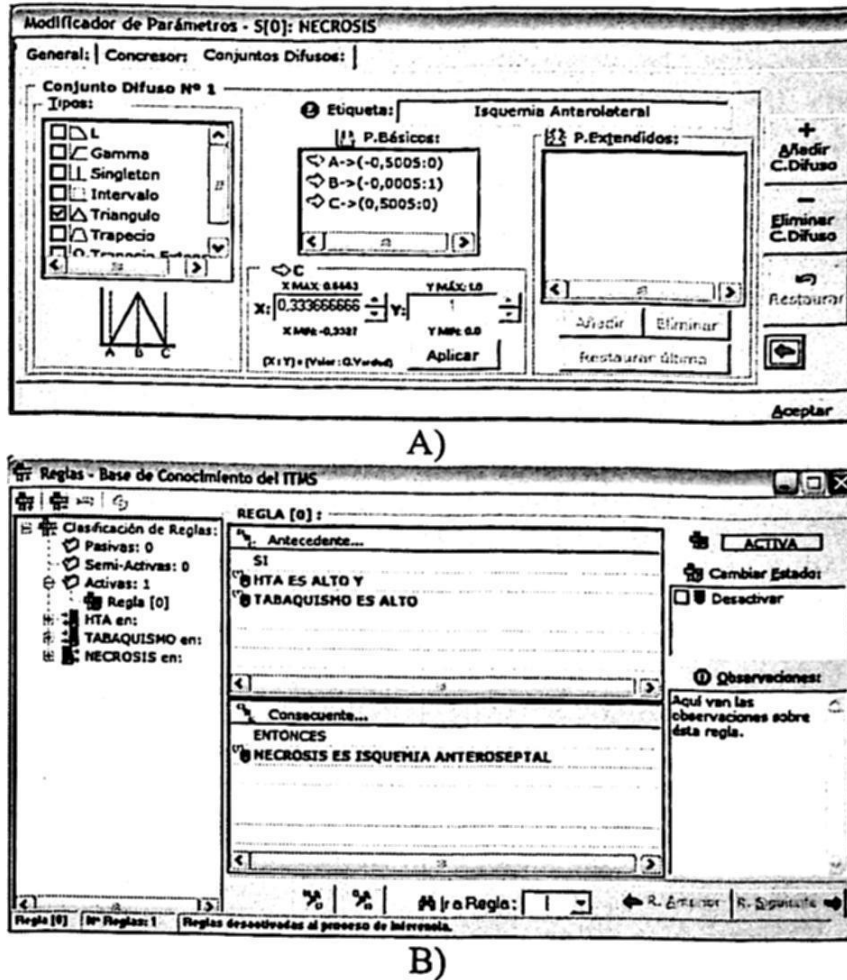


Figure 7: A) Necrosis Output Variables. Linguistic Ischemia Labels Anterolateral.
B) Rule of Knowledge.

As seen in this case, the use of vague data is much used in complex systems with information from doctors in medicine, among others. The DW are of great advantage when comes the time to make decisions about certain thematic. The use of fuzzy logic can provide a good way to work with vague data. In this case, the use of an extension of a fuzzy related model FSQI to a fuzzy DW model has been shown, besides using fuzzy inference.

4 Conclusion and Future Work

The methodology based on a set of steps presented here may serve as a guide for similar tools, mainly to create data storehouses that allow to handle the information management with vague data by means of the theory of fuzzy ensembles.

Proposing the use of Type 1, 2 and 3 attributes, linguistic labels and inference rules is a contribution to supporting the decision making and, in some instances, satisfy

more efficiently the user's requirements than the traditional DW systems. FSQ created by Galindo can be incorporated to different SQL platforms with related models (Escobar, 2003), including data stores, o DW (Salas, 2006).

One of the tasks is to represent in a concept model for DW that includes Type 1, 2 and 3 attributes, besides the linguistic labels shown on this work. Produce a clear method to specify the user's requirements when considering imprecise data which could be handled with fuzzy logic.

Acknowledgments: Catholic University of the Maule Internal Project Number 81201(2006-2007).

References

1. Carpani Fernando (2000), "*CMDM: Un modelo conceptual para la Especificación de Bases Multidimensionales*", Tesis de Maestría, Universidad de La Republica, Uruguay.
2. Escobar Calixto (2003), "*Software para control difuso de todo tipo de sistemas (SCD): Aplicación al Control de Invernaderos Industriales*". Tesis de título de Ingeniero Técnico Industrial (Electrónica).
3. Galindo J., Urrutia A., Piattini M. (2005): "*Fuzzy Databases: Modeling, Design and Implementation*". To publish by Idea Group Publishing Hershey, USA.
4. Galindo J. (1999): "*Tratamiento de la Imprecisión en Bases de Datos Relacionales: Extensión del Modelo y Adaptación de los SGBD Actuales*". Ph. Doctoral Thesis, University of Granada (Spain). (www.lcc.uma.es).
5. Golfarelli, M. Rizzi, S.: "Methodological Framework for Data Warehouse Design.", DOLAP'98, USA, 1998.
6. Ling Feng, Tharam S. Dillon, Fellow, IEEE (2003), "*Using Fuzzy Linguistic Representations to Provide Explanatory Semantics for DataWarehouses*", IEEE Transactions on Knowledge and Data Engineering, Vol. 15, N° 1.
7. Inmon, W (1996): "*Building the Data Warehouse*". John Wiley & Sons, Inc. 1996.
8. Kimball, R.: "*The Datawarehouse Toolkit*". John Wiley & Son, Inc., 1996.
9. Kumar Pavan, Krishna Radha, Kumar Supriya (2005), "*Fuzzy OLAP Cube for Qualitative*" Institute for Development and Research in Banking Technology, IDRBT-2005
10. Marotta, A.: "*Data Warehouse Design and Maintenance through Schema Transformations*". Master Thesis. Advisor: Raúl Ruggia. Pedeciba, Universidad de la República, Uruguay, 2000.
11. Medina J.M., Pons O., Vila M.A. (1995): "*FIRST. A Fuzzy Interface for Relational SysTems*". VI International Fuzzy Systems Association World Congress (IFSA'1995). Sao Paulo (Brasil).
12. Peralta Verónica (2001), "*Diseño Lógico de Data Warehouse a partir de Esquemas Conceptuales Multidimensionales*", Tesis de Maestría, Universidad de La Republica, Uruguay.
13. Salas Yosselin (2006): "*Extensión del Diseño e Implementación de un Data Warehouse Difuso, Aplicado a Enfermedades Cardiovasculares*". Tesis para Ingeniero en Computación e Informática Universidad Católica del Maule, Chile.
14. Urrutia A. (2003): "*Definición de un Modelo Conceptual para Bases de Datos Difusas*". Ph. Doctoral Thesis, University of Castilla-La Mancha (Spain).
15. Zadeh L.A., "Fuzzy sets". Information and Control 8, pp 338-353, 1965.