Especificación formal de cubos de datos aplicados a la administración de las actividades docentes

Formal specification of data cubes applied to the administration of teaching activities

Pedro Fletes Gudiño
Instituto Tecnológico de Colima
pfletes@itcolima.edu.mx

Nicandro Farías Mendoza
Instituto Tecnológico de Colima
nfarias@itcolima.edu.mx

Resumen

Durante los últimos años las bases de datos multidimensionales se han convertido en algo común en el mundo de los negocios y científico. En este trabajo se propone un modelo para la especificación formal de base de datos multidimensionales (BDMD) o cubos de datos, que nos permita a través de un proceso clasificar, filtrar información del área de docencia, seccionando los registros en ejes y capas. La meta que se pretende alcanzar a través de la propuesta de este modelo es la especificación formal del cubo de datos.

Palabras clave: base de datos multidimensionales, cubos de datos, docencia y modelo.

Abstract

During the last years the multidimensional databases have become something common in the business and scientific world. In this work is proposed a model for the formal specification of Multidimensional Database (MDB) or data cubes, which allows us to through a process sort, filter information in the area of teaching, sectioning records in

axes and layers. The goal that intends to achieve through the proposal of this model is the formal specification of the data cube.

Introduction

Background

During the last twenty years my main role has been as a teacher at the Technological Institute of Colima, always in the Department of Systems and Computer Science. One of the areas that has excited me always as a graduate in computer science are the data bases and for this reason this thesis project was elected to continue with the current theme of multidimensional databases.

Problem Statement

Currently at the national level, not the exception being our Institute, a problem is experiencing by the high rates of reproach of the students. Ways to reduce these rates has always sought, but at the institutional level there is a computational tool to keep track of these, which also provide statistical information from different years and different parameters. This led us to develop our proposal that will generate, through the formal specification, a computational scheme (data cube), capable of storing relevant information on previous rates of reproach and shed information so quickly and efficiently about the diverse causes giving rise to the reproach at the Institute.

Objective

Develop a formal model for the specification of Multidimensional Databases that can be represented with SQL1 for a study case focused on the administration of the teaching activities.

Specific objectives

• Define the formal model.

• Specify the case study.

• Detail the flow of data for the case study.

State of Knowledge

Historical Setting

This project had its beginnings in the Department of Systems and Computing Teaching Projects Headquarters, where each control half the plannings of those offered takes. Teachers at the beginning of the semester planning delivered each of the subjects taught and the teaching instrumentation of each of the units. During the course of the semester, the teacher grading and delivery progress at the end of the course delivery records of qualifications, where the rate of approval and disapproval of the group appears.

1.2 Contextual Framework

In this chapter, a brief description of other works presented in conceptual and logical modeling of multidimensional databases. To facilitate understanding of this work and unify terminology, previously an introduction to the structures and operations of multidimensional models will be presented.

Multidimensional data base

databases are designed to develop very specific applications, such as creating OLAP Cubes (Online Analytical Processing). Basically do not differ too much from the relational database (a table in a relational database could also be in a multidimensional database), the difference is rather a conceptual level; in the multidimensional database fields or attributes of a table can be of two types, or represent dimensions of the table, or represent metric that you want to study.

Vs. multidimensional databases OLAP cubes

Each of these tables can be likened to a hypercube or more specifically whether OLAP tools is an OLAP cube, where the dimensions of the fields correspond to the size of the table (fields 'di ...'), and value stored in each cell of the cube equals the metric or metrics (fields 'fi ...') stored in the table.

This type of database is applied to the system also called multidimensional OLAP cube or hyperlink. It consists of numeric facts called measures which are classified by size. The cube metadata is typically created from a star schema (see Fig. 1) or snowflake (see Fig. 2) and using
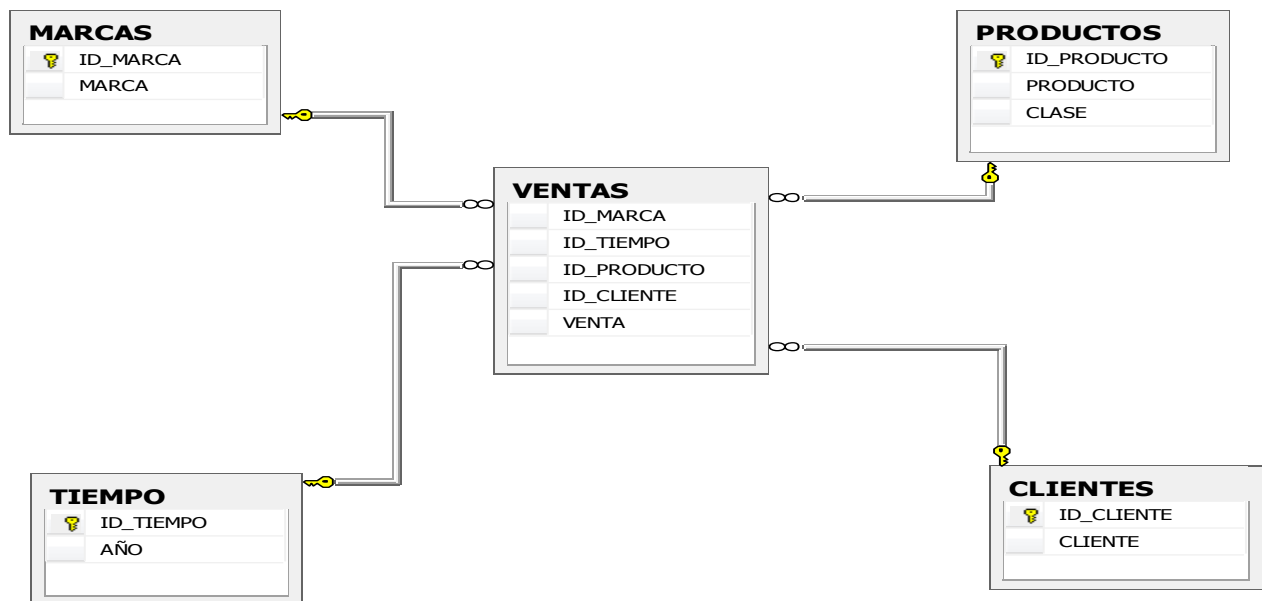
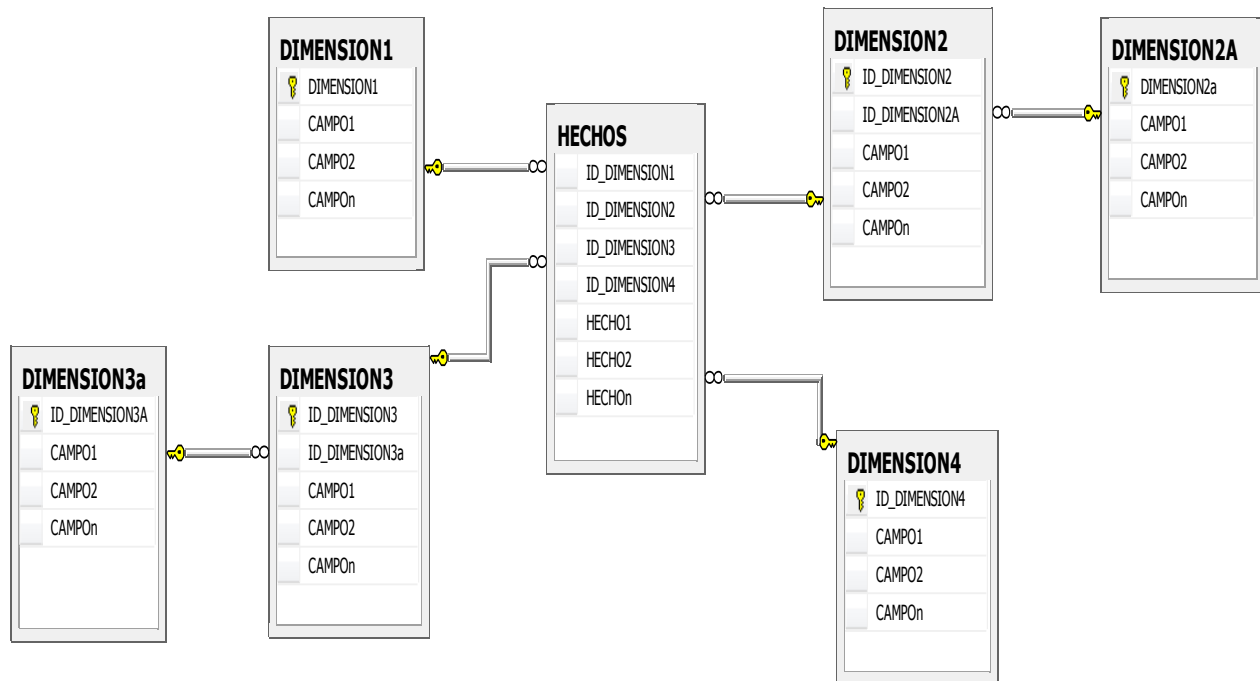tables in a relational database.

Fig.1 Star schema

Fig.2 Snowflake Schema

Related work and contributions

(Seng & Habibollah, 2013) KSOM intend to use to organize unstructured data for closed surfaces. KSOM used in this work by testing their ability to organize data KSOM medical imaging that is used primarily to build data engineering field. (Hsiao, Wo-Shun, & Petchulat, 2011) show a working prototype of a system OLAP client-centered, with a custom middleware on the server side and a client web that incorporates a light engine for OLAP data consultations in memory. They developed three tools for interactive data visualization running on the data engine in the client side. (Mansmann & Scholl, 2007) present an approach to explore multidimensional data cubes with hierarchical visualization techniques. A job whose main contribution is an approach called hierarchical difference scatterplot (HDS). This proposal allows multiple levels of hierarchy and explicitly displayed on the differences between them in the context of the absolute position of the pivoting values; It is presented by (Piringer, Buchetics, Hauser, & Gröller, 2009). (Lafon, Bouali, Guinot & Venturini, 2013) we have several ways of reorganizing an OLAP cube in terms of which all members of the reorganization is selected: the entire membership, members

will only show certain level (level level approach). The work aims to combine the route of cells the size and parametric statistical tests to identify different meaningful metrics between cube cells, is presented by (Ordonez, Chen, & Garcia-Garcia, 2011). (Takama & Yamada, 2009) offer us a bucket display for modeling interaction in exploratory data analysis of spatial and temporal trends. (Pitarch, Laurent & Poncelet, 2009) propose a conceptual model to model customized hierarchies multidimensional databases. The proposed construction of an algorithm based on ellipsoids (ER-Tree) presented by (Dankoand & Skopal, 2009) shows that they significantly affect the speed, indexing and query performance in large databases such as multimedia , medicine and geography. (Stolte, Tang, & Hanrahan, 2006) propose a visual query language that allows the graphical display of table-based data. This language commands compiles both consultations as needed to generate the graphics display, allowing us to design systems that tightly integrate analysis and visualization. (Yaghmaie, Bertossi, & Ariyan, 2012) propose the formalization of the path of the relational schema, which becomes the basis for obtaining dimensional repairs, here shown that relational ordinary star and schemes snowflake multidimensional snow databases are not the best choice for this process. TEXplorer a novel system allows users to search for keywords and OLAP aggregation and stylish, exploring the text objects in a building on a multidimensional text database presented to us (Zhao, Lin hub Ding, & Han, 2011). (Esch-Bussenmarkers & Cremers, 2004) propose a tutorial of an experimental system where the project objective was to evaluate the different types of access (voice, graphics and touch mode) to access and present certain types of information, and certain strategies Search to search and browse a database of musical multidimensional data using a simulated mobile.

Theoretical framework

Here are the items included in this project are detailed:

Database

A database management system (DBMS Data Base Management System) is a collection of interrelated data and a set of programs to access the data (Silberschatz, Kotrh, & Sudarshan, 2002, p. 1). A shared collection of logically related data, together with a description of these data, which are designed to meet the information needs of an organization (Connolly & Begg, 2005, p. 15). Multidimensional data base

Database-oriented based on complex queries with high performance structure. You can use an RDBMS star (multidimensional database logic level) or SGBDM (multidimensional database to logical and physical levels or Pura Multidimensional Database) (De la Herran Gascon, 2004).

Data cubes

The OLAP Cube, who coined its name from its multidimensional feature is a database that has several dimensions.

Data cubes are used in systems online analytical processing (OLAP) to support decision-making. Built from the databases of a business, this interactive display system presents a hierarchical conditional 1D cubic structure tree to represent data cubes and using 2D graphic icons to illustrate the data elements. Users can then interactively explore multi-dimensional data in hierarchical levels.

Development of data cube applied to the administration of teaching activities

System analysis

In all technological institutes nationwide, there are academic departments that manage the careers offered, and in each headquarters there teaching. Currently there is an automated system that allows us to make a record of data, then information like that today demand.

It is therefore not possible to obtain control information (statistics approval, failure and dropout rates) of each of the teachers (by subject, class time, half) of the races, the department and the Institute. Moreover, it is not possible to easily obtain information from previous semesters.

Conceptual model

The proposed multidimensional diagram (see Figure 1) shows how information from the SIITEC and different areas of teaching of academic departments, the information is undergoing transformations and segmenting (steps 1 and 2) to reach the data cube (step 3) and already there can be processed to make inquiries it (step 4) are required.
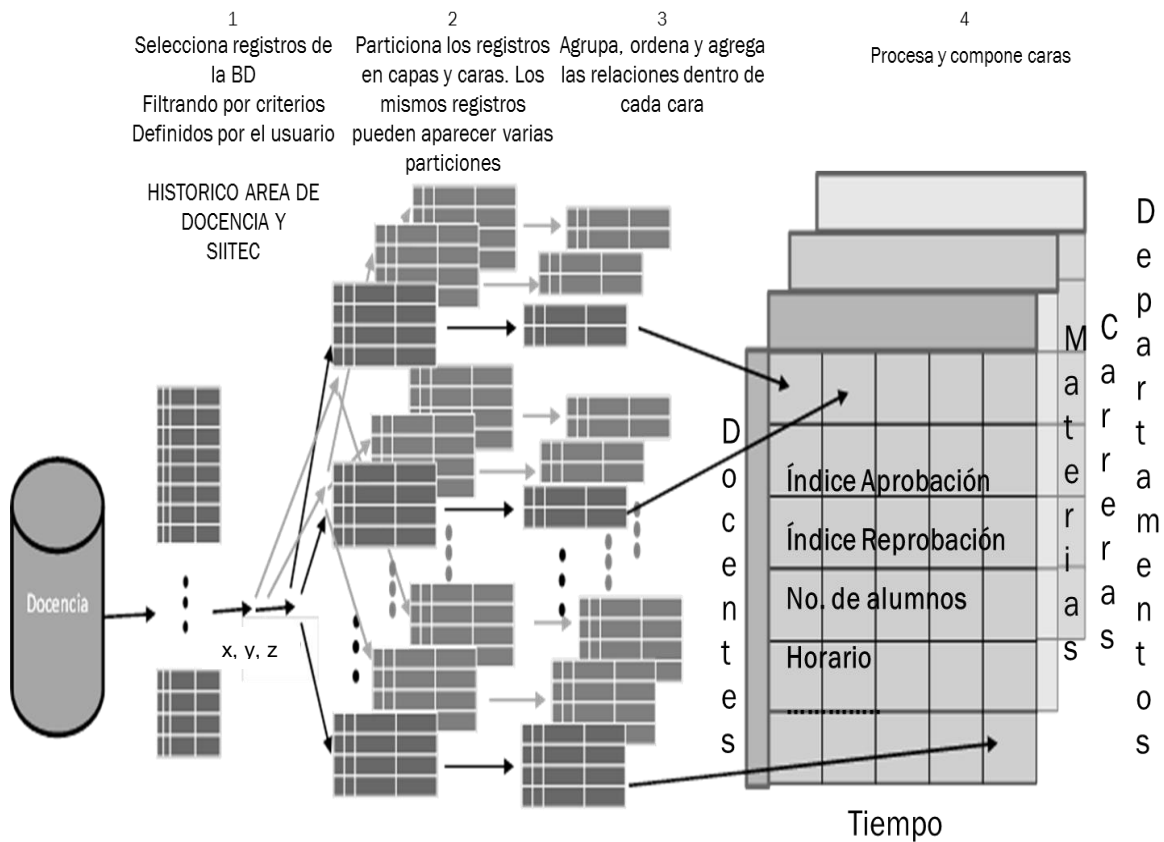
Figura 1 Big Picture del esquema multidimensional propuesto.

System Requirements

At this point considering all the information you can get the data cube to assess the aspects for which the disapproval of students in each of the subjects courses during his stay at the Institute and going from happening:

- Subjects with higher failure rate
- Schedule with the highest disapproval
- Teachers with higher failure rate
- Aspects that influence the disapproval
- Marital status of students.
- Employment status of students (if they have a job).

–   Gender of students.

–   Location students (local or foreign).

–   Grant (if the student has some support).

–   If you have any previous race.

–   The schedule of matter.

All this information can be obtained from the point of view of each of the cube dimensions: teaching, time, department, career and matter.

Formal specification model for Multidimensional Data Base (BDMD) or data cubes

The model for the formal specification of data cubes consists of the following five times:

MCD[1] = (C, OAT, OPAT, PR, OBD)

Where:

C = { ordinal fields, quantitative fields, dimensions, measurements} are the basics.

OAT = {A, B, C, P, Q, R} are the operators of the algebra of tables.

OPAT = {+, X, /} are the operators of the algebra of tables.

PR = { Filter Row (i), Column (j), Layer (k)} are the predicates.

OBD = { Aggregates, dim, SUM, AVG} Operators Database.

This proposed model is based on an algebra of tables that captures the structure of the tables in relational model and performs the conversion of these to a format of data cubes formalism. Concepts used:

Ordinal fields: The ordinal scale is discreetly displayed as headers or different kinds.

---

[1] Modelo de cubo de datos.

Quantitative fields: quantitative continuous and fields are shown as axes or as a continuously variable values.

Algebra tables

We define an algebra as the formal mechanism for specifying configuration tables. A complete configuration consists of three separate tables algebra expressions. Two expressions define the configuration of X-axis and Y-axis table sectioning the table axes and columns. The third term defines the Z axis of the table, which part in the unfolded layers. The expressions X, Y and Z are clauses in language.

The operands in the algebra of tables names are ordinal or quantitative fields in the database. It is using A, B and C to represent ordinal and P, Q and R fields to represent quantitative fields. Sequences assign values to each field symbol as follows: ordinal fields to assign members of an ordered domain of the field and sets quantitative fields of individual items containing the field name is assigned.

Ordinal and quantitative fields generate tables with different structures (1) - (2). Ordinal segment table fields in rows and columns using headers, while quantitative fields generate axes.

$$A = domain (A) = \{a_1,...,a_n\} \qquad (1)$$
$$P = \{P\} \qquad (2)$$

A valid expression in algebra consists of one or more symbols with operators between each pair of adjacent operands and parentheses used to alter the precedence of operators.

Operators algebra tables

Concatenation (+) operator + (3) - (5) concatenates two sequences as follows:

$$A + B = \{a_1, ..., a_n\} + \{b_1, ..., b_m\} \qquad (3)$$
$$= \{a_1, ..., a_n, b_1, ..., b_m\}$$

$$A + P = \{a_1, ..., a_n\} + \{P\} \qquad (4)$$
$$= \{a_1, ..., a_n, P\}$$

$$P + Q = \{P\} + \{Q\} \qquad (5)$$
$$= \{P, Q\}$$

Product: (X) The operator Product (6) - (7) takes the Cartesian product of two sequences.

$$A \times B = \{a_1, ..., a_n\} \times \{b_1, ..., b_m\} \qquad (6)$$
$$= \{a_1 b_1, ..., a_1 b_m,$$
$$a_2 b_1, ..., a_2 b_m, ...,$$
$$a_n b_1, ..., a_n b_m\}$$

$$A \times P = \{a_1, ..., a_n\} \times P \qquad (7)$$
$$= \{a_1 P, ..., a_n P\}$$

Projection (/) The projection (8) operator is similar to the operator product, but only creates sequences for which records exist.

$$A / B = \{aibj \mid \square r \square R \ st \ A(r) = ai \ \& \ B(r) = bj\} \quad (8)$$

The intuitive interpretation of the projection operator B in A.

For example, given the quarter and month fields, the expression quarter / month could be interpreted as those months within each quarter, resulting in three tickets each quarter. In contrast, X month quarter could be 12 entries for each quarter. Data cubes represent hierarchies and need not explicitly compute the mapping relationship.

Data processing and generating queries in the database

An important aspect of the proposed formalism is the transformation of data. So the formalism must support the full range of data transformations in a query in a query language like SQL, including common relational operators: selection, filtering, grouping, and aggregation system.

Total flow of data in the MCD model

The formalism must be compatible with the full range of possible data transformations in a query language like SQL, including common relational operators: selection, filtering, grouping, aggregation and classification. It can be shown that any SQL query defined can be expressed as a specification formalism MCD.

In this proposal, as fields are added while the dimension fields are inserted into a statute GROUP BY with additional dimensions fields specified in detail. Each dimension can also be ordered, and

the different aggregation functions can be associated with each measure and these options can be chosen. There is also a filter that represents the elements in the WHERE clause. Finally, general and joins calculations are also presented.

It reference to Figure 1, it is explained in detail each of the steps shown in the figure described by the model MCD.

Step 1. Selecting records. The first phase of the data stream retrieves records from the database, using user-defined filters to select subsets of data.

For an ordinal field A, the user can specify a subset of the domain of the field like filter (A) is a subset selected by the user. Then a relational predicate expressing the filter for A is:

A filter in (A)

For a quantitative field P, the user can define a subset of the domain of the field as valid if min (P) and max (P) are extensions of user defined subset. The relational predicate expressing the filter for P is:

$(P \geq \min (P)$ and $P \leq \max (P))$

We can define the relational predicate filter as a filter combination of all the individual fields. So the first stage of data processing network is equivalent to the following SQL statement:

```
SELECT  *
WHERE {filtros}
```

It is possible within the whole formalism to define a more sophisticated filter such as filter product cross multiple field units or system filters (filter A is calculated relative to the filter B).

Step 2. Partition panels records: The second phase of partitions data stream retrieved records in groups corresponding to each panel in the table. The table is divided into rows, columns and layers corresponding to the entries in these sets.

The ordinal values in each entry define the criteria established by which the records are arranged in each row, column, and layer. Given the row (i) it is the predicate that represents the selection criteria for the i-th row, column (j) be the predicate of the j-th column, and Layer (k) the predicate for the k-th layer . For example, if the Y-axis table is defined by the standardized set:

{A1b1P, a1b2P, a2b1P, a2b2P}

Below are four rows of the table, each defined by an entry in this group, and the row is defined as:

Fila (1) = (A = a_1 and B = b_1)
Fila (2) = (A = a_1 and B = b_2)
Fila (3) = (A = a_2 and B = b_1)
Fila (4) = (A = a_2 and B = b_2)

Given these definitions, the records in which the panel is divided is the intersection of the i-th row and j-th column and the k-th layer can be recovered with the following query:

SELECT *

WHERE {Row (i) and column (j) and Layer (k)}

Step 3. Transformation of records within the panels: The last phase of the data stream is the transformation of the records in each panel. If the visual specification includes aggregation, then every measure in the database schema must be assigned to an aggregation operator.

For example, if the database contains the profits, sales and payroll quantitative fields, and the user has specified explicitly that the average sales must be calculated, then the aggregates are defined as:

Aggregate = SUM (Profit), AVG (sales), SUM (Payroll)

Filters aggregate fields (eg SUM (Profit> 500), could not be assessed in Step 1 with all other filters, because the aggregates had not yet been calculated. Therefore, the filters should be applied in this phase. We define the relational predicate filters as in step 1 for non aggregate fields.

Also, define the following lists:

G: The field names in the platform group,

S: The field names in the sorting platform, and

Dim: Dimensions in the database.

The necessary transformation can be expressed by the SQL statement:

```
SELECT {dim}, {agregados}
GROUP BY {G}
HAVING {filters}
ORDER BY {S}
```

If the added fields are not included in the specification, the remaining processing simply sorts the records in order:

SELECT *

ORDER BY {S}

Design

Because the proposed scheme is the generation of a data cube for the exploitation of information in the area of teaching through a formal specification, it takes as input a schema relational database. In Figure 2 the multidimensional scheme which specifies each of the proposed dimensions and measures of the fact table (GROUP) is shown.

Architectural design

It is proposed that the fact table group also contains the attributes already mentioned (Index approval, failure rate, number of students and hours matter), additional data to complement the information and we offer the possibility of extracting information from different points of view. Figure 2 shows the conceptual model of the proposed scheme is shown. Below all the attributes of each of the dimensions and the fact table are specified.
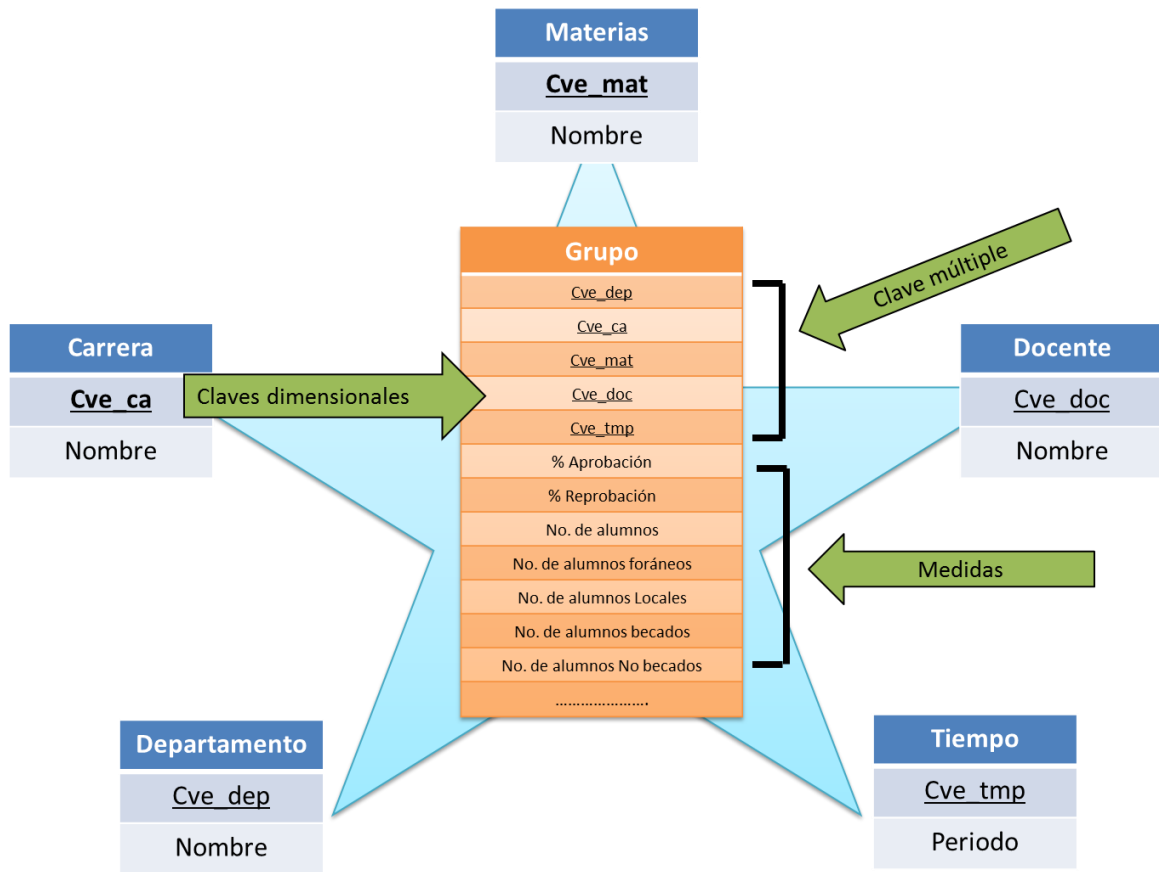
Figura 2 Conceptual Model.

Design data cube

Drawing on the conceptual model in Figure 2, the proposed design is shown in the ER diagram (see Figure 3).
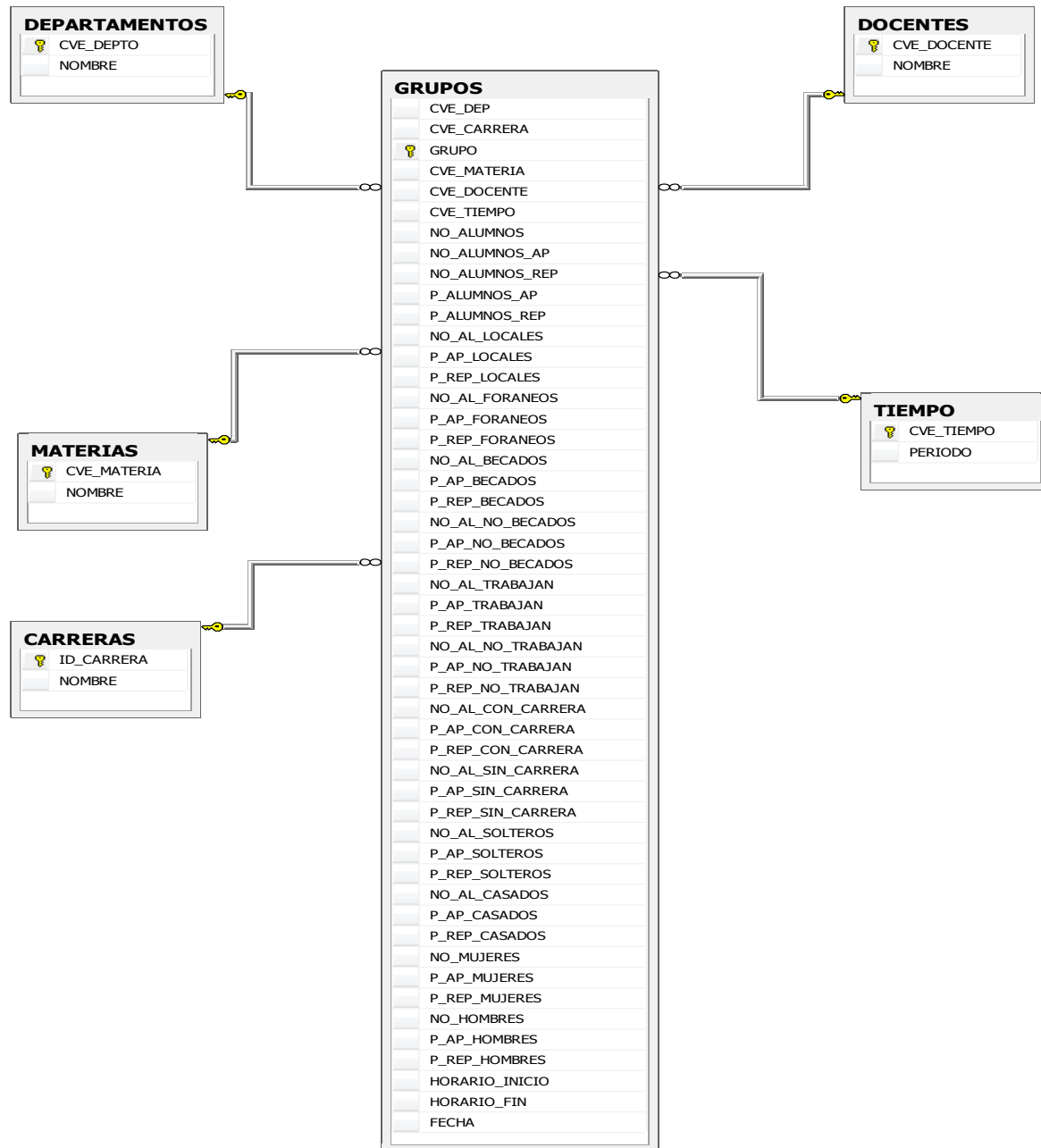
Figura 2 E-R diagram data cube.

No = Number, P = Average, AP = aprobation, REP = Failure.

Study case

The following case shows how from the relational schema in Figure 4 and using the MCD model reaches the data cube Figure 3.
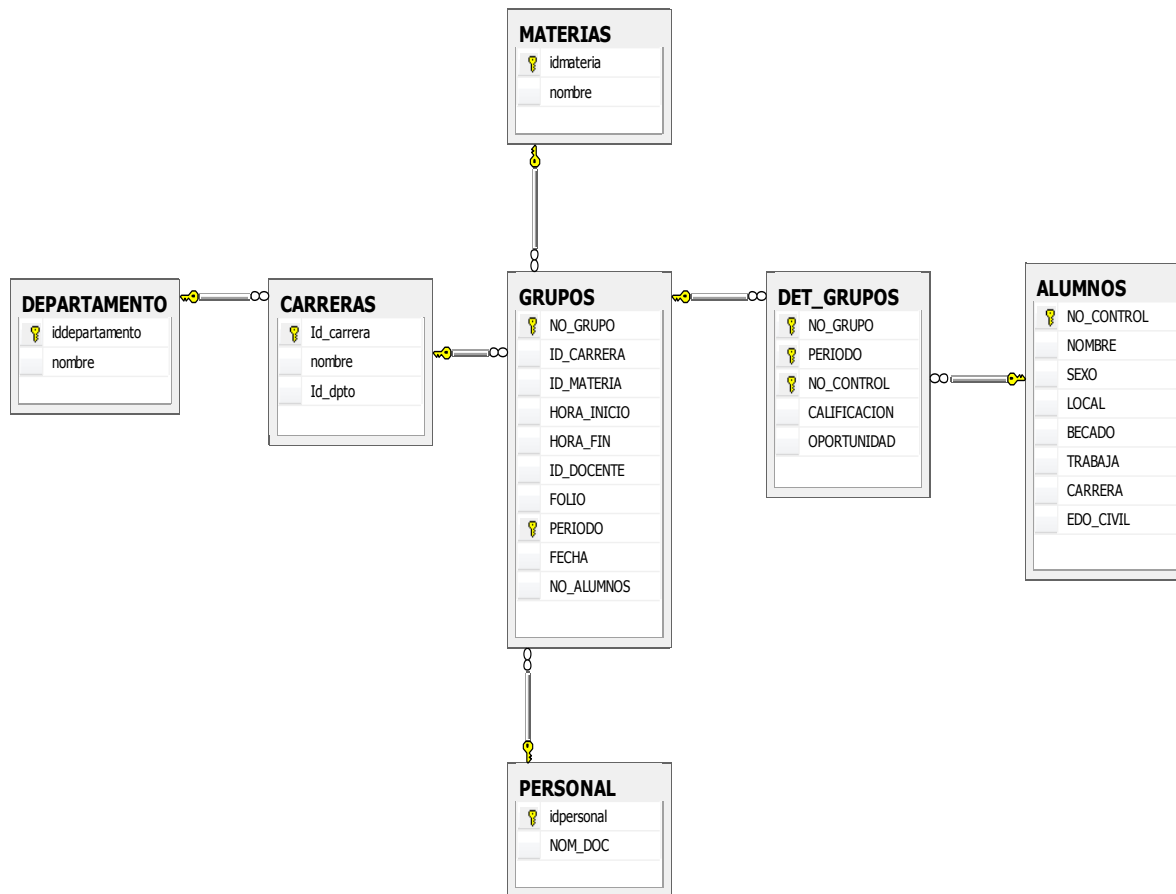

Figura 3. Relational model of teaching.

In the case shown below the generation of the cube dimensions and the fact table is displayed (here you get only the key group of -carrera key dimensions, materials, teacher and department- time, when start, end time class, folio, number and percentage of passing students).

SQL statements using the code would be:

Select L.No_Grupo, L.Id_Carrera, L.Id_Materia, L.Hora_inicio, L.Hora_fin,
        L.Id_Docente, L.Folio, L.Fecha, L.No_Alumnos, P.Iddepartamento,
        Count (*), Count (*)/L.No_Alumnos
From Grupos L join Det_Grupos M   on L.No_Grupo   = M.No_Grupo
                join Alumnos N   on M.No_control = N.No_Control
                join Carreras O   on L.Id_Carrera = O.Id_Carrera
                join Departamento P   on O.Id_Dpto   = P.Iddepartamento
Group By L.No_Grupo
Having M.Calificacion >= 70

MCD using the proposed method, the specification would be as follows:

Select A_1, A_2, A_3, B_1, B_2, A_4, B_3, B_4, P_1, A_5, Q_1, Q_2
From Grupos L / Det_Grupos M / Alumnos N / Carreras O / Departamento P
Group By A_1
Having Filtro$_1$

Where:

A_1 = L.No_Grupo	A_4 = L.Id_Docente	Q_1 = COUNT (*)
A_2 = L.Id_Carrera	B_3 = L.Folio	$Q2$ = COUNT (*) / P1
A_3 = L.Id_Materia	B_4 = L.Fecha	Filtro$_1$ = {M.Calificacion >= 70}
B_1 = L.Hora_inicio	P_1 = L.No_Alumnos	
B_2 = L.Hora_fin	A_5 = P.Idepartamento	

y L, M, N, O y P They are tuple variables or alias tables Groups, Det_Grupos, Students, Racing and Department respectively. Displayed in this query is generated:

- The Departments with P.Iddepartamento dimension attribute,

- The Career dimension with L.Id_Carrera attribute,

- The size materials with L.Id_Materia attribute,

- The Teachers dimension attribute with L.Id_Docente

- The time dimension with L.Fecha attribute.

For the attributes No_alumnos_Rep y P_Alumnos_Rep, the filter is changed to:

$Filtro_1 = \{M.Calificacion < 70\}$

For the attributes No_Al_Locales, P_Ap_Locales y P_Rep_Locales, the filter is changed to:

$Filtro_1 = \{M.Calificación >= 70\ and\ N.Local = 'L'\}$ y
$Q_2 = COUNT\ (*)/P_1$
$Q_3 = 100 - Q_2$ y se añadiría al Select

For the attributes No_Al_Foráneos, P_Ap_Foráneos y P_Rep_Foráneos, the filter is changed to:

$Filtro_1 = \{M.Calificación >= 70\ and\ N.Local = 'F'\}$ y
$Q_2 = COUNT(*)/P_1$
$Q_3 = 100 - Q_2$ y se añadiría al Select

For the attributes No_Al_Becados, P_Ap_Becados y P_Rep_Becados, the filter is changed to:

$Filtro_1 = \{M.Calificación >= 70\ and\ N.Becado = 'S'\}$ y
$Q_2 = COUNT(*)/P_1$
$Q_3 = 100 - Q_2$ y se añadiría al Select

For the attributes No_Al_No_Becados, P_Ap_No_Becados y P_Rep_No_Becados, the filter is changed to:

$Filtro_1 = \{M.Calificación >= 70\ and\ N.Becado = 'N'\}$ y
$Q_2 = COUNT(*)/P_1$
$Q_3 = 100 - Q_2$ y se añadiría al Select

For the attributes No_Al_Trabajan, P_Ap_Trabajan y P_Rep_Trabajan, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Trabaja = 'S'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Al_No_Trabajan, P_Ap_No_Trabajan y P_Rep_No_Trabajan, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Trabaja = 'N'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Al_Con_Carrera, P_Ap_Con_Carrera y P_Rep_Con_Carrera, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Carrera = 'S'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Al_Sin_Carrera, P_Ap_Sin_Carrera y P_Rep_Sin_Carrera, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Carrera = 'N'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Al_Solteros, P_Ap_Solteros y P_Rep_Solteros, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Edo_Civil = 'S'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Al_Casados, P_Ap_Casados y P_Rep_Casados, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Edo_Civil = 'C'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Mujeres, P_Ap_Mujeres y P_Rep_Mujeres, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Sexo = 'F'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select

For the attributes No_Hombres, P_Ap_Hombres y P_Rep_Hombres, the filter is changed to:

$Filtro_1$ = {M.Calificación >= 70 and N.Sexo = 'M'} y
Q_2 = COUNT(*)/P_1
Q_3 = 100 – Q_2 y se añadiría al Select.

With the above exposed it is formally specified the data cube, using the MCD model.

Total data flow of the Big Picture using the case study

To explain each of the steps of the Big Picture, Figure 1, use the example for obtaining the relational data model with the data cube in the previous section.

If we have the following formalization:

Select A_1, A_2, A_3, B_1, B_2, A_4, B_3, B_4, P_1, A_5, Q_1, Q_2
From Grupos L / Det_Grupos M / Alumnos N / Carreras O / Departamento P
Group By A_1
Having $Filtro_1$

Where:

A_1 = L.No_Grupo	A_4 = L.Id_Docente	Q_1 = COUNT (*)
A_2 = L.Id_Carrera	B_3 = L.Folio	$Q2$ = COUNT (*) / P1
A_3 = L.Id_Materia	B_4 = L.Fecha	$Filtro_1$ = {M.Calificacion >= 70}
B_1 = L.Hora_inicio	P_1 = L.No_Alumnos	
B_2 = L.Hora_fin	A_5 = P.Idepartamento	

y L, M, N,O y P are tuple variables or alias tables Groups, Det_Grupos, Students, Racing and Department respectively.

Step 1. Select the database records filtering by user-defined criteria.

This step is defined through Filter1 = {M.Calificación> = 70}, in this case with the filter the number and percentage of students approved are obtained.

Step 2. Segment records layered and expensive. The same records can appear in multiple partitions.

In this way the data are grouped into faces through the clause:

Select A_1, A_2, A_3, B_1, B_2, A_4, B_3, B_4, P_1, A_5, Q_1, Q_2
From Grupos G / Det_Grupos D / Alumnos A / Carreras C / Departamento P

In the Select the A5, A3, A2, A4 and B4, variables determine the dimensions of Department, Matter, Career, Teacher and Time respectively. Linking data through the projection operation in clause From.

Step 3. Group, sorts and adds the relationships within each face.

In this step grouped, they ordered to be added on each side through the instruction:

Select A_1, A_2, A_3, B_1, B_2, A_4, B_3, B_4, P_1, A_5, Q_1, Q_2

Where each attribute is defined above.

Step 4. Processes and made faces.

And with the information stored in the data cube through SQL statements can be obtained indicators as required.

Analysis of results

The scope of this project is limited to the formal specification of data cube through the MCD model, which through the case study demonstrated.

Another result is that any design of a data cube can be represented and viewed through this specification.

Furthermore, without any designer implement the hub can capture and display the final outline data cube.

Conclusions and recommendations

Observing the results we can say that any expression represented by the SQL language can be formally specified using the data cube model developed in this research. By the above, it is observed that the objectives are met in the beginning and therefore is fully adopted the working hypothesis for this investigation.

The importance of this project lies in the design show a data cube through a formal specification using the MCD model, in this case as a proposed solution to be indicators of failure rates that reveal the causes of these rates in different careers offered by the Technological Institute of Colima and fold.

It is also important to note that the multidimensional diagram obtained is analytical and able to store both current and historical information, giving the opportunity to efficiently manage the information derived from the teaching activities of the Institute.

To continue in the future with this research suggests:

• Use the other representation scheme as a data cube it is the snowflake to display the advantages or disadvantages of this approach.

• Increase the size of the cubes, which gives us the ability to manage other technological grouped into geographical regions.

Bibliography

Connolly, T. & Begg, C. (2005). *Sistemas de Bases de datos.* Madrid: Addison Wesley.

Dankoand, T. & Skopal, T. (2009). *Elliptic Indexing of Multidimensional Databases.* Wellington, New Zealand: Australasian Databases Conference.

De la Herrán Gascón, M. (2004). *Red Científica.* Recuperado el 28 de Noviembre de 2013, de http://www.redcientifica.com/oracle/c0001p0005.html

Esch-Bussenmarkers, V. & Cremers, A. (2004). *User Walkthrough of Multimodal Access to Multidimensional Databases.* Pennsylvania, USA: ICMI.

Hsiao, T.; Wo-Shun, L.; & Petchulat, S. (2011). *Data Visualization on Web-based OLAP.* Glasgow,Scotland, UK: DOLAP'2011.

Lafon, S.; Bouali, F.; Guinot, C. & Venturini, G. (2013). *Hierarchical Reorganization of Dimensions in OLAP Visualizations.* Francia: IEEECS.

Mansmann, S. & Scholl, M. H. (2007). *Exploring OLAP Aggregates with Hierarchical Visualization Techniques.* Seoul, Korea: SAC'07.

Ordonez, C.; Chen, Z. & García-García, J. (2011). *Interactive Exploration and Visualization of OLAP.* Glasgow, Scotland, UK: DOLAP'11.

Piringer, H.; Buchetics, M.; Hauser, H. & Gröller, E. (2009). *Hierarchical Diffrence Acatterplots.* París, Francia: VAKD'09.

Pitarch, Y; Laurent, A. & Poncelet, P. (2009). *A Conceptual Model For Hamdling Personalized Hierarchies in Multidimensional Databases.* Lyon, France: MEDES.

Seng, P. & Habibollah, H. (2013). *Cube Kohonen Self-Organizing.* Malaysia: IEEE Transactions onNeural Networks and Learning Systems Vol.24.

Silberschatz, A.; Kotrh, H. & Sudarshan , S. (2002). *Fundamentos de Bases de datos.* España: McGraw-Hill.

Stolte, C.; Tang, D. & Hanrahan, P. (2006). *Polaris: A System for Query, Analysis and Visualization of Multidimensional Databases.* ACM.

Takama, Y. & Yamada, T. (2009). *Visualization Cube: Modeling Interaction for Exploratory Data Analysis of Spatiotemporal Trend Information.* Tokyo, Japan: IEEE/WIC/ACM.

Yaghmaie, M.; Bertossi, L.; & Ariyan, S. (2012). *Repair-Oriented Relational Schemas for Lultidimensional Databases.* Berlín, Alemania: EDBT.

Zhao, B.; Lin, x.; Ding, B. & Han, J. (2011). *TEXeplorer: Keyword-based Object Search and Exploration in Multidimensional Text Databases.* Glasgow, Scotland, UK: CIKM.