The Design of a Distributed Database for Doctoral Studies Management

Enikö Elisabeta TOLEA, Aurelian Razvan COSTIN
Babes Bolyai University, Cluj-Napoca, Romania
eny_tol@yahoo.com, costin9razvan@yahoo.com

This paper aims to create a system that manages doctoral school requirements. The management of doctoral school implies administration of information like PhD personal information, Supervisors, Teachers, and other information that may be useful. We will debate on distributed database term as the proposed database project will have the same structure for four universities. Each university will be able to work on this database by accessing its own set of data and properly using the information received. This project will track the creation of a database to manage all the information needed and provide answers using these data.

***Keywords**: Distributed Database, Fragmentation, Top-Down Design Process*

1 Introduction

The goal of the project debated in this article is to create a system which has the ability to process information needed for management doctoral school. Its final utility will be to manage doctoral schools requirements. After designing the database we want to create a user friendly interface to facilitate its usage.

In order to discuss databases development we need to know what doctoral school requirements are. After analyzing this problem we concluded that the prerequisites are the following:

1. Recording all data necessary for the management, namely: data on doctoral students, tutors and school teachers, materials that are studied, exams, conferences and other information related to the doctoral school.
2. Data processing to derive information.
3. Information processing.
4. Creating query to display useful information. These queries will eventually be integrated into an application that will offer a friendly user interface that will provide an information management tool, along with reports and other evaluation methods tools needed to properly administrate a doctoral school.

This article is structured in two sections and includes the following:

a) First part: Problem, context and statements- in this section we will debate what a distributed database is and why it is better to create and use distributed databases instead of normal databases.

b) Second part: Proposed Model- in this section we will describe the architectural solution of this complex model along with designs and implementations details.

2 Problem Context and Statement

2.1 Related Works

Databases are extremely important because they stand behind any applications be it web pages or Enterprise Resource Planning projects.

What is a Database? A database can be defined as a set of basic or structured data, accessible to a community of users. More specifically, a database is a set of interrelated files that contain the core data required by a computer system (computer application).

Because we want to integrate this database in several universities we have concluded that we need a distributed database. One that by fragmenting by each university will provide access to its one information, but by rebuilding relationships will provide general information on all universities.

These definitions are representative for the distributed databases.

Distributed database is a set of logically related databases, but physically distributed on several machines, linked through a

communications network [1].
Very important is how to create the database and framing them in:

1. distributed databases
2. federal databases

This classification is made depending on how it wants to operate within their data.
We can define a distributed database as a collection of multiple, logically interrelated databases distributed over a computer network [2]. A distributed database management system is defined as a software system that permits the management of the Management System of Distributed Databases and makes the distribution transparent to the users [3]. System management software is necessary because it allows the database is distributed and the distribution is transparent to the user [4].
In distributed databases replication and fragmentation are two of the most important properties that a distributed database has. Fragmentation involves dividing a global relationship into logical drivers that can be placed optimally in distributed database. Fragmentation should be consistent semantic distributed database, to achieve a lossless decomposition of information and without duplicates, allowing reconstruction of the initial relationship before fragmentation. These can be achieved through three types of fragmentation: Horizontal, Vertical, Mixed. Replication involves copying fragments in several locations. For that it is required a structure for the database that fits all four university and by using this two properties this structure will have the best results.

2.2 Statement

The main reason of this project was to create a database that manages information not only from one university but from more. This database structure was designed so that all the needs of this consortium to be found in this structure.
First of all we analyzed all possible situations and we concluded that the top-down design process is the one that better answers to ours requirement. A framework for this process is shown in Figure 1 [5]. The activity begins with a requirements analysis that defines the environment of the system and "elicts both the data and processing needs of all potential users" [6]. The requirements are two parallel activities view design and conceptual design. The first activity deals with the defining of the interface for the final users, and the second activity is the process by which entity types and relationships are determined among entities.
We need a database to manage information not just for one university, but for four of them.
It takes such a database to manage all these information, precisely because so far every university manages its own data. But with the development of such a project, which aims to encapsulate more universities and to manage information from all of them this project will offer a wider range of information and users achieve a much better information management.
The problems that arise when designing this database were mainly related to structure.
This structure must match both the primary database or one that includes all four universities and the same structure must fit each university separately.
So we are talking about a structure which may well manage a single university data separately, but that is the same structure and information management behind consortium information management.
The solution to the above mentioned was to create a database that will be discussed further, a database that respects what we have said earlier.
Next we discuss our proposed model. How it was developed, designed, implemented and a SWOT analysis of it.

3 Proposed Model

3.1 Design Details

The choice of creating this distributed database and its concrete implementation depended on:

- business needs, what information do this universities need to offer to future user;

- privacy of the database ;
- required degree of security, consistency and data integrity.

Another responsible question to ask at this point is: Why we chose to create an distributed database?

The classical answer to this question indicates that distributed processing better corresponds to the organizational structure of today's distributed enterprises, in this case universities, and that such a system is more reliable and more responsive. A lot of current applications of computer technology are inherently distributed. Because we want to integrate this database structure in more than one university, we need to create a distributed database.

As Chris J. Date [7] says a distributed system should look exactly like the one undistributed. This is the Fundamental Principle that Date sets in 1978, together with other twelve rules closely concerning distributed systems. There for this is why we chose to create such a database. We will discuss in the following about its actual creation of the database. Starting from the idea mentioned above will create database that will contain tables with attributes that we can offer users the information they need.

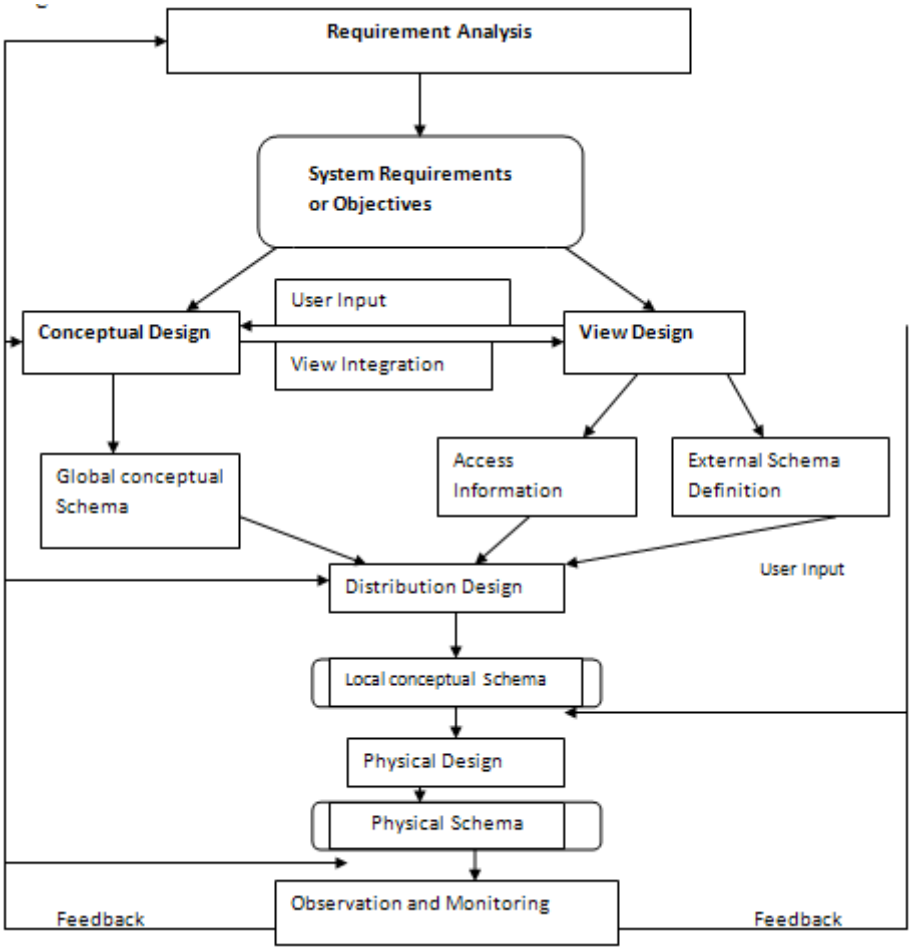

Fig. 1. Top-Down Design Process

3.2 Implementation Details

As you can see we tried to integrate in its structure all the information related to the doctoral school, namely: data on PhD students and tutors, teachers, subjects taught, exams, participating in conferences, articles and their quotes and other information that we considered useful.

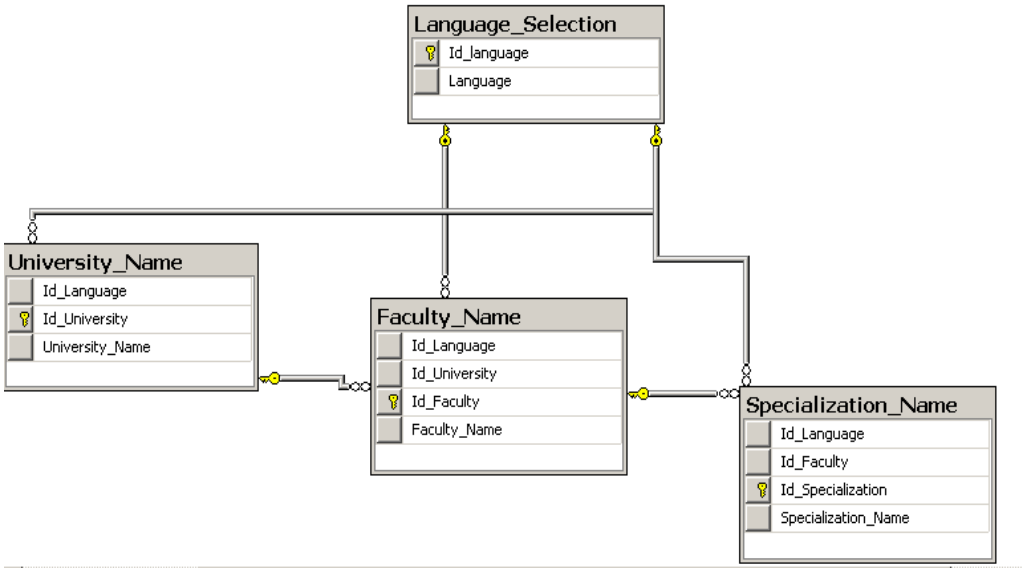

Fig. 2. Tables and their contents form our database

In Figure 2 you can see some tables. The table Language_selection has two attributes, Id_language and Language (the language name) and this table is very important because when we are going to implement it we will have the possibility to chose in which language to be shown the information. For the begging in this table we added the following languages: Romanian, English, French, German and Hungarian, with the possibility of extending them by adding other languages. In Figure 3 you will see an ERD (entity-relationship diagram) Diagram in which you will find a sort description of the database structure. There are included some tables, not all of them, and the relation between them. And in Figure 4 you will see an example of a query and its result.

With the input data we made some queries (the tables were created in Microsoft SQLServer 2008 R2), and here is an example. This code for example displays the University, Faculty and Specialization names.

```
SELECT
University_Name.University_Name,
Faculty_Name.Faculty_Name,
Specialization_Name.Specialization_Name
FROM         Faculty_Name INNER JOIN

Specialization_Name                    ON
Faculty_Name.Id_Faculty                =
Specialization_Name.Id_Faculty      INNER
JOIN
                      University_Name  ON
Faculty_Name.Id_University             =
University_Name.Id_University
For this code we have as a result
the fallowing (Figure 4):
```

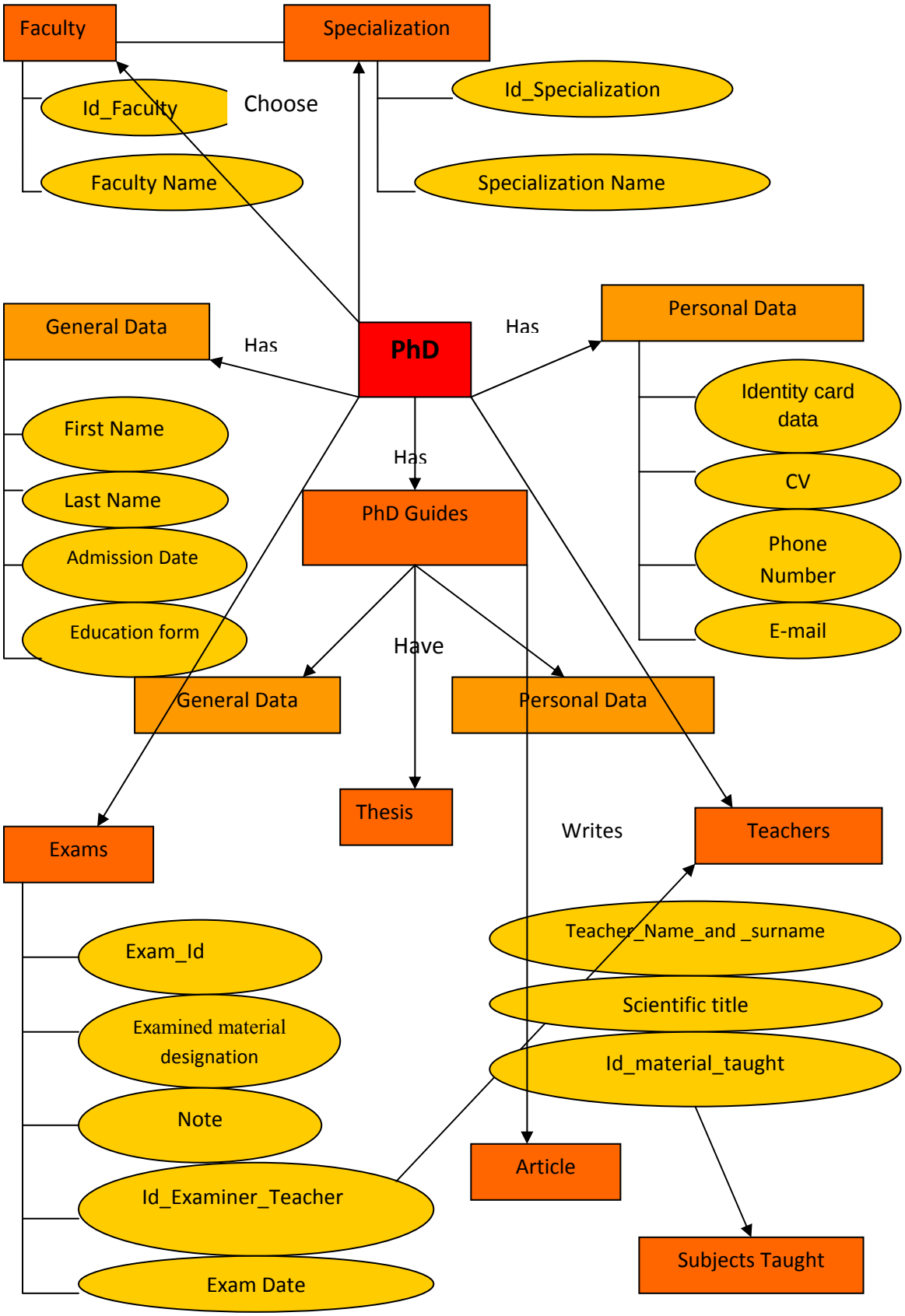

Fig. 3. ERD Diagram

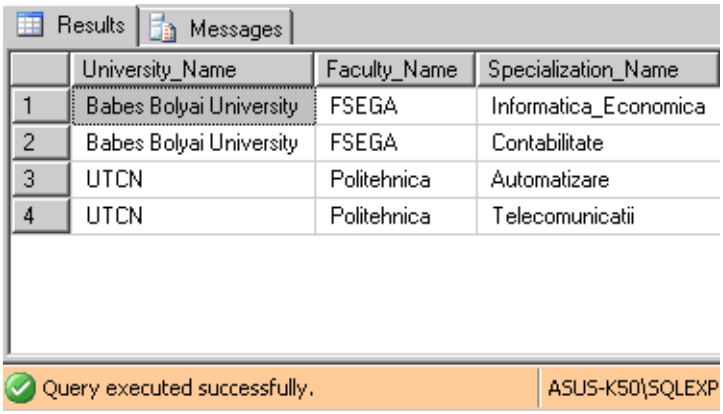

Fig. 4. Result for the interrogation

4 Evaluation

This project hasn't reached total development. It is still in the implementation process. Next we perform a SWOT analysis of the project described above (Figure 5). As you can see in this analysis this project offers advantages and disadvantages, pros and cons for this project.

It requires an analysis of indicators that can be found in Figure 5. Strengths are very important because this project had a requirement, and that was the possibility to integrate this application in four universities and the information had to be accessible to all users. Also, we need a database schema that can be extended if needed by adding new information that not appears in the initial structure, although here we found a weakness, namely the difficulty of changes within the structure.

We thought we would have problems in implementation because there could be differences between how universities implement this structure in their university; but it was also a challenge because it was something new, a new structure, a new information management system.

4.1 Subsequent Developments

1. We will create queries using the same SQL Server environment, which will attempt to answer by listing all questions that may arise on database.
2. Will try to correlate this DB with an object approach programming environment.
3. It wants to create an interface through which you can access the database, and providing a series of operations such as displaying, adding data, display reports and other things.
4. All these future developments will form a complete and fully functional project that will meet all requirements of future users.

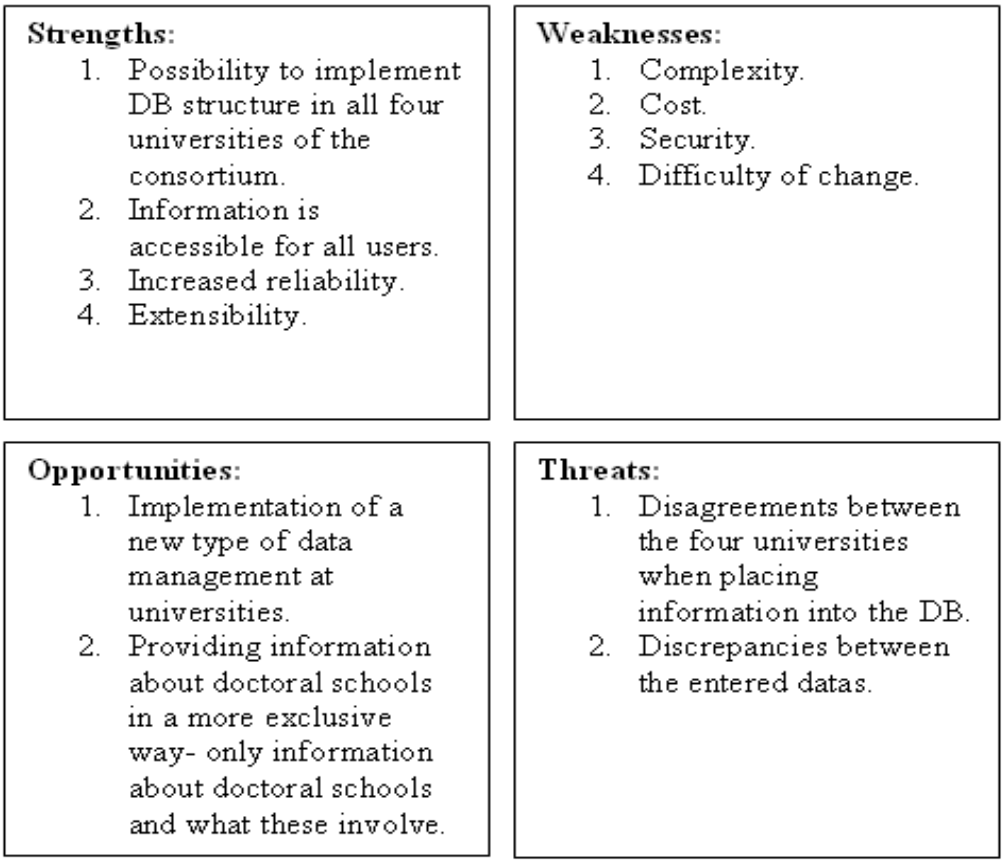

Fig. 5. SWOT Analysis

5 Conclusions

We created such a database because we need to fragment it and we need to manage these fragments as well as we can.

This project was extremely important to develop since that time there wasn't an exclusive management of PhD.

So we tried to develop the structures embedded in it all the necessary information for the management of PhD.

This is a project that requires time and multiple re-evaluations for the structure to be one that any university to be able to integrate it without problems. Therefore we proposed an approach that in time, after several evaluations we hope that will reach that level.

Acknowledgments

This work was supported by ANCS-CNMP, project number PNII – 91037/2007.

References

[1] I. Popescu, *Modelarea bazelor de date*, Editura Tehnica, Bucuresti, 2001.

[2] T. Ozsu, P.Valduriez, *Principles of Distributed Databese Systems*, Second Edition, 2002.

[3] T. Connolly, C. Begg, A. Strachan, *Baze de date Proiectare. Implementare. Gestionare*, Teora, 2001

[4] T. Ozsu, P. Valduriez, *Principles of Distributed Databese Systems*, Second Edition, 2002

[5] S.B. Yao, S. B. Navathe and J-L. Weldon, "An Integrated Approach to database design", *In Data Base Design Techniques I: Requirements and Logical Structures, Nyu Symposium*, New-York, 1982.

[6] C. J. Date, *Baze de date*, Ediția a VIII-a, Editura Plus, Bucureşti, 2005

Enikö Elisabeta ŢOLEA is a 24 years old PhD candidate. She took her Master of Science degree in 2010 and started her publishing carrier in February 2010 with a first article in the Journal of Information Systems & Operations Management. She started studying economical informatics at the Faculty of Economics and Business Administration in 2005, graduating in June 2008. After this period she followed the courses of SADE (Systems to Assist Economic Decision) mastery school in the same institution. Now she is a young research member and she has been accepted in the doctoral school, studying Cybernetics and Economical Statistics under the guidance of Prof. Dr. Ştefan Ioan Niţchi at the same faculty.

Aurelian Răzvan COSTIN is a 24 years old PhD candidate. He took his Master of Science degree in 2010 and started his publishing carrier in February 2010 with a first article in the Journal of Information Systems & Operations Management. He started studying economical informatics at the Faculty of Economics and Business Administration in 2005, graduating in June 2008. After this period he followed the courses of E-Business mastery school in the same institution. Now he is a young research member in a project named SICOMAP, and has been accepted in the doctoral school, studying Cybernetics and Economical Statistics under the guidance of Prof. Dr. Niţchi Ioan Ştefan at the same faculty.