CENTRO *UNIVERSITÁRIO* DE BARRA MANSA
ACADEMIC PRO-RECTORY

COMPUTER ENGINEERING COURSE

TOP-DOWN APPROACH IN DISTRIBUTED DATABASES

By:

Leniel Braz de Oliveira Macaferi

Barra Mansa
November 2007

UBM

CENTRO *UNIVERSITÁRIO* DE BARRA MANSA
ACADEMIC PRO-RECTORY

COMPUTER ENGINEERING COURSE

TOP-DOWN APPROACH IN DISTRIBUTED DATABASES

By:

Leniel Braz de Oliveira Macaferi

Paper presented to the Computer Engineering course at Centro Universitário de Barra Mansa, as a partial requisite to the obtention of the second grade in the Distributed Databases discipline, under prof. Hélio Camargo Soares supervision.

Barra Mansa
November 2007

ABSTRACT

A distributed database is formed by a collection of multiple databases logically interrelated in a computer network.

The top-down approach when used in distributed databases correlates a series of stages for the construction of a distributed database project beginning from the ground and is employed in homogeneous systems. The emphasis in the case of distributed databases is given to the data distribution project.

This work presents the stages of the top-down approach through a schema, which gives a macro vision of the process. After that the inherent details of each process's stage are described.

Keywords: database, distributed database, top-down approach, distributed database construction

TABLE OF FIGURES

CONTENTS

1 INTRODUCTION

1.1 Paper objective

This paper aims to present and describe the top-down approach stages when it is employed in distributed databases.

1.2 Distributed database definition

According to [1], a distributed database is formed by a collection of multiple databases logically interrelated in a computer network. A distributed database management system (DDBMS) is used as a system that enables the management of the individual database systems (DBMS). The DBMS distributes the data through a transparent way to the user.

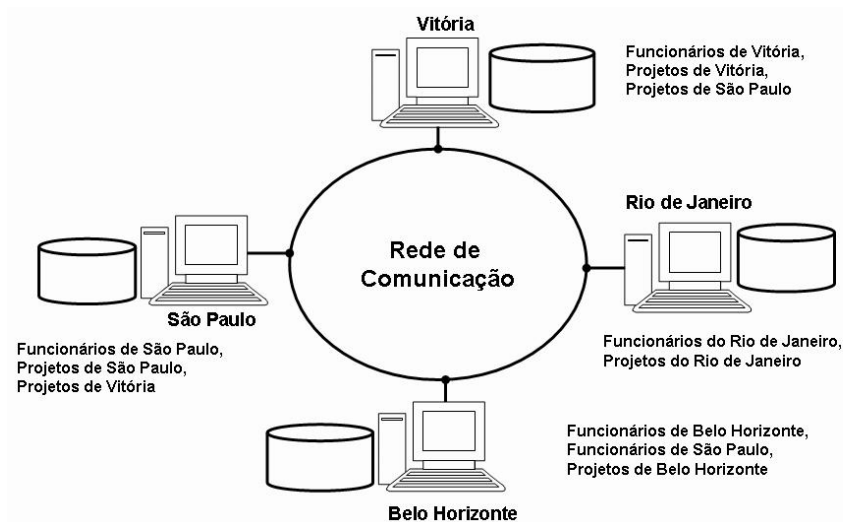Figure 1 shows a common example of a distributed database.

Figure 1 - Distributed database example

The transparence provided by a DDBMS can be understood as the high level semantic separation of the details inherent to the physical implementation of a distributed database.

The focus is to provide data independency in a distributed environment. This way, the user sees only one logically integrated image of the distributed database as if it weren't physically distributed. Figure 2 shows the logical view that a user has (or should have) of the distributed database presented in Figure 1.
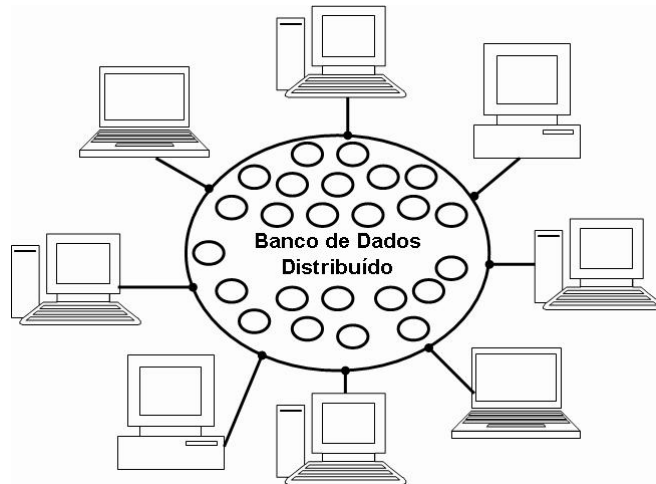
Figure 2 - Logical view of a distributed database user

In the next chapter the top-down approach stages will be presented and described. It's important to mention that such approach is directly linked to the methodology adopted in the construction of distributed databases.

2 DEVELOPMENT

2.1 Top-down approach

The top-down approach is employed in different computer areas. In distributed databases it correlates a series of stages to the construction of a distributed databases project beginning from the ground and is employed in homogeneous systems [1].

The emphasis in the case of distributed databases is given to the data distribution project.

The top-down approach is the more usual given the fact that in the majority of the cases there isn't a distributed database already implemented. If there is a system already established another approach called bottom-up is used.

Figure 3 shows the stages of the top-down approach employed in distributed databases:
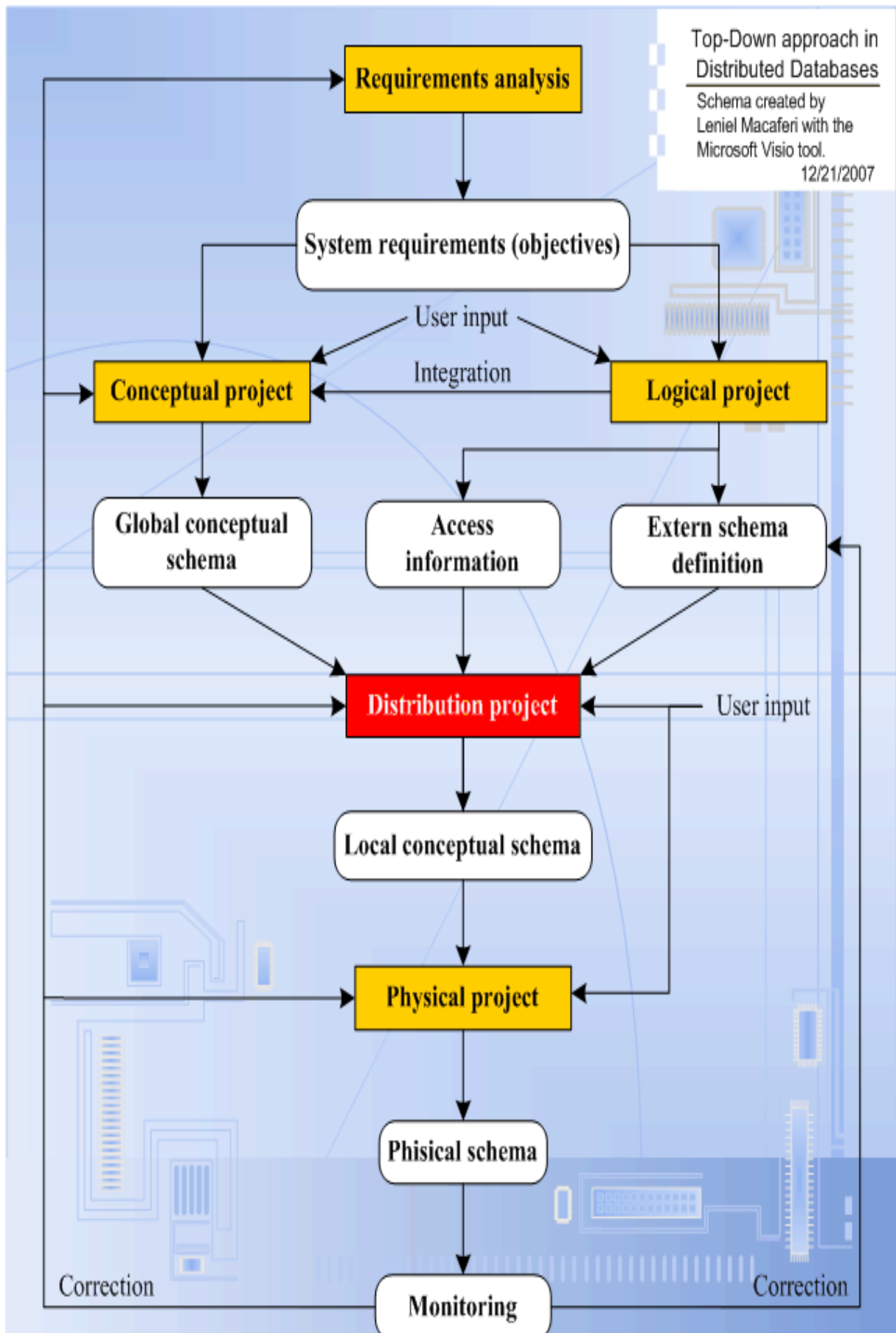
Figure 3 - Stages of the top-down approach in distributed databases

As can be seen, in yellow are the traditional stages of the top-down model: requirements analysis, conceptual project, logical e physical project. In read is the distribution project stage that is specific in distributed databases. Note that the physical project in this case is implemented after the distribution.

The following describes the stages of the top-down approach in distributed databases.

2.1.1 Requirements analysis

In this stage takes place the collect of information about the data, and its restrictions and relationships within the organization. The requirements analysis is realized through meetings with the users where is observed how the organization operates. In the end of the analysis a document with the requirements specification is generated.

2.1.2 Conceptual project

In this stage takes place the modeling of the data and its relationships independently of the structure of representation regarding the distributed database system DDS (conceptual modeling). The conceptual project is realized through the analysis of the requirements specification. In the end of the conceptual project a conceptual schema (diagram) with de correct data integrity restrictions is obtained.

2.1.3 Logical project

In this stage takes place the conversion of the conceptual project to the representing schema of a DDS (logical schema). The logical project is realized through the application of conversion rules, translation to the relational model of the distributed database. In the end of the logical project a logical schema with tables, stored procedures, views, access authorizations, etc. is obtained.

2.1.4 Distribution project

In this stage is taken the decision of how the data and programs must be allocated, fragmented through the nodes of the computer network. In some cases the network itself is designed and built to satisfy the necessities of the distributed database project. This stage is considered the most critical in the project of a distributed database.

2.1.5 Physical project

In this stage the logical schema is defined in a DDS suitable to the data model. The physical project is realized by means of SQL instructions. The result is a physical schema in concordance with the established in the distribution project. After finishing the physical project of each node of the computer network the distributed database is ready for the use. A monitoring process is initialized and aims to discover possible errors. Such errors are the system feedback and are sent to the people responsible for the construction of the distributed database.

3 CONCLUSION

At the beginning of a distributed database project it's extremely important to assess the organizational environment of the company that holds the data. Obviously when the company doesn't have a distributed database back-end and legacy systems, the top-down construction approach will be necessarily employed.

As a macro overview it's possible to infer that in the first stage a document with the requirements is generated and after that the logical and conceptual projects start. Logical and physical schemas start to be generated. With the logical and physical schemas already defined the distribution project starts and this is the most complex. After the definition of the local schema of each node of the computer network, the implementation of the local physical schema starts. At this point each network node is given the responsibility for determined tasks of the company. This means that some objects (views, stored procedures, etc.) of the database are created specifically according to each physical local schema. In the last level is the distributed database monitoring process, which helps in the discovering of bugs and propitiates at the same time the possibility of correction forwarding the bugs to the superior levels.

The top-down approach aims to structure the creation process of a distributed database. Defining and separating the construction stages in a correct manner, the database architects and other people involved in the construction of a distributed database will have more chances of achieving success in a given project. It'll only happen if the stages are accomplished with strictness and in the established order.

4 REFERENCES

[1] Ozsu, M. Tamer e Valduriez, Patrick. **Principles of Distributed Database Systems**. 2nd Edition. Upper Saddle River : Prentice Hall, 1999.

[2] Zhou, Li-Zhu. **Distributed Database System Course**. 2002. Available at <http://dbgroup.cs.tsinghua.edu.cn/ddb/2002/ddbHandout.zip>. Accessed on November 16, 2007.

[3] Mello, Ronaldo S. **Projeto Top-Down de Banco de Dados**. 2007. Available at <http://www.inf.ufsc.br/~ronaldo/ine5623/1-introdProjBD.pdf>. Accessed on November 16, 2007.

[4] Ozsu, M. Tamer e Valduriez, Patrick. **Notes for "Principles of Distributed Database Systems"**. 1999. Available at <http://softbase.uwaterloo.ca/~tozsu/ddbook/notes/Design/Design.pdf>. Accessed on November 22, 2007.