SOLUTION FOR DATA ORGANIZATION IN THE BANKING DOMAIN

ANA-MARIA RAMONA Stancu[1], MARIAN POMPILIU Cristescu[2], LIVIU CONSTANTIN Stoica[3]
[1]"Dimitrie Cantemir" Christian University
[2]"Lucian Blaga"University of Sibiu
[3]Academy of Economic Studies, Bucharest, Romania

Abstract

In this article I will present "Solutions for organizing data in banking". In the first part I will design a multidimensional model and explain the basics of it and observe it for banking data model will be of "Snowflake". In the second part realize data warehouse architecture and introduce steps to achieve it. In the third part realize dimensions and cubes of data in banking, and in the last part I made the ETL process.

Key words: *algorithm, cluster, distance, iteration, group*
JEL classification: *O31, O32*

1. Introduction

In the first part I design a multidimensional banking model. It is known that multidimensional modeling is a logical design technique used to visualize data models as a set of key variables for the analyzed activity.

Multidimensional patterns are seen as restricted forms of ER models, resulting in almost direct mapping between them. So multidimensional models in data repositories can be derived directly from entity-relationship schemes of operational source systems.

[1] e-mail: ana_maria_ramona@yahoo.com
[2] e-mail: marian.cristescu@ulbsibiu.ro
[3] e-mail: stoica.liviu.constantin@gmail.com

In the second part we made the architecture of the data warehouse in the banking field.

In the third part we made the dimensions and the data cubes, in order to achieve them we needed to create a data warehouse

To create it, I left a questionnaire I gave the respondents, centralized the data in Excel and imported them into Oracle. I mention, the number of respondents was 2500 people in Bucharest and Ilfov, and the questions in the questionnaire were the basis for the database and date of the deposit. The database tables are: Employees, Questionnaire, Bank_Bank, Account, Operations, Branch, and Time.

2.DESIGN OF A MULTIDIMENSIONAL MODEL (Stancu Ana-Maria Ramona, 2017)

Multidimensional modeling is a logical design technique used to visualize patterns as a set of variables for analysis activity. A multidimensional model can be implemented in multidimensional databases or data warehouses.

Basics of model projects are:

a) attributes / members

b) collections of facts and dimensions may be played in a relational database as a table.

b1. Dimensions also show some important features to remind: include statistical data, provide information about variables, are useful in analysis, etc. A dimension contains more members, and each member is used to determine the position of the date element. For example, the Time dimension contains: Date, Month, Semester and Year..

b2. Collection of facts is all related data consisting of variable and context data.

c) The degree data dispersion in the model is an important component which influences the size of the database;

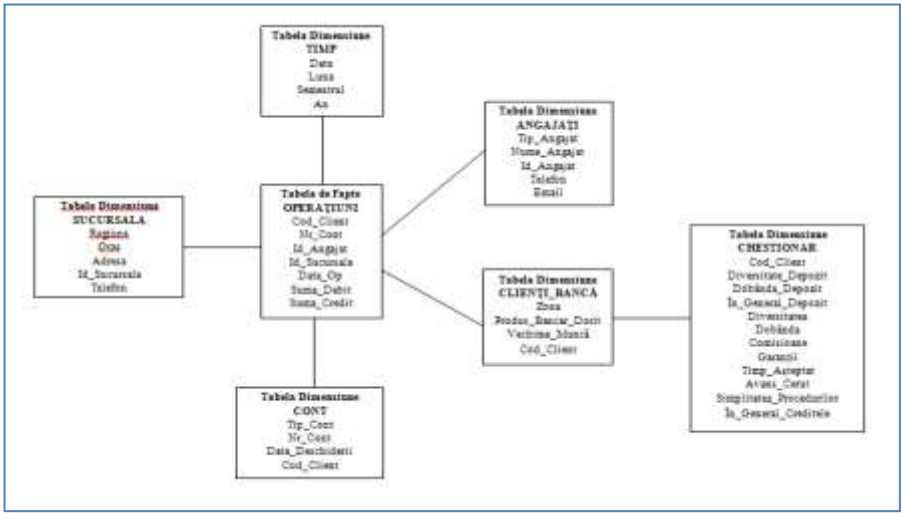
d) hierarchy;

e) Indicators / variables. An attribute of an element from the collection of works is a performance indicator by which we can analyze the performance or non-performance of the banking modeled activity. Indicators are as follows:

- Basic indicators (quantity sold, sales volume, etc.);
- Additive indicators (are obtained by combining the basic indicators);
- Derivative indicators (total after all dimensions may be);
- Semi-additive indicators (can total as certain dimensions).

In multidimensional modeling for banking data, the analysis suggests using a snowflake scheme. The proposed "snowflake" model is a version of the "star" version and is the consequence of decomposing a dimension (or several dimensions) that has hierarchies.

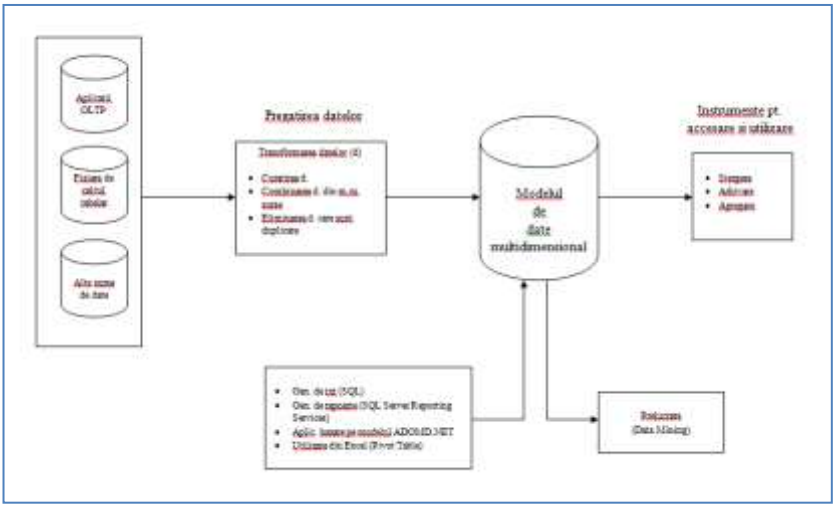
Figure 1 - Multidimensional data model of "Snowflake"



3. DATA WAREHOUSE ARCHITECTURE (Stancu Ana-Maria Ramona, 2017**)**

For organizing banking data I propose the data warehouse architecture presented in Figure 2

Figure 2 - Data warehouse architecture banking



The steps for making the data warehouse are:

- extracting data from the data source followed by copying them;
- preparation and upload of data into the data warehouse (metadata describing the data that is in my data warehouse and how they are obtained and stored);
- data mining processing is based on model selection and aggregation of data is in the data store. The main goals of this in practical terms are:
    - a prediction requires the use of variables or fields from the database with the aim of achieving future estimates;
    - a description guides finding patterns that describe the data that is interpretable in terms of The User.
- It can be noticed that in architecture the tools for access and use can be used by analysts and decision-makers. These tools can be::
    - a tool that is useful to users (e.g., query language SQL, generators of reports);
    - a tool used to assist decision: data mining and OLAP tools. OLAP tools are based on multi-dimensional representation of data (data cube) and allow their rapid analysis by means of roll-

up type operations, drill-up, slice, etc., and the user can obtain timely results.

4. MAKING DIMENSIONS AND DATA CUBES (Stancu Ana-Maria Ramona, 2017**)**

The dimensions of this application are: employer questionnaire EMPLOYEES, QUESTIONNAIRE, CONSUMER_BANK, ACCOUNT, OPERATIONS, BRANCH, and TIME.
Each dimension is characterized by a hierarchy which in turn is characterized by several levels.

STAFF dimensions Employees

To achieve the **dimension** I selected storage type and then we defined the attributes of the dimension (Type_Employee, Name_Employee, Id_Employee, Phone, Email I know that attributes of a dimension are specified by descriptions, names, types of data, and the attribute identifier that defines a level must contain the Business type .After this step we defined levels of the dimension attributes we associate levels, e.g., select each level and each one will indicate attributes that are part of the structure.

Dimension hierarchy is L_TYPE_Employee, L_Employee.

The Staff dimension has attributes: TYPE_Employee, Name_Employee, Id_Employee, Phone, Email, levels: L_TYPE_Employee (Accountains field TYPE_Employee) L_Employee (Accountains fields: Name_Employee, Id_Employee, Phone, Email) and Standard hierarchy (Accountaining levels: L_Type_Employee, L_Employee). This dimension is shown in Figure 3

Cube Questionnaire

In order to create the Rubik cube Survey Questionnaire we selected storage type and after that I chose the dimension of the cube (R_Staff, R_Customers, R_Account, R_Time). After this step we have specified metrics cube (these are Code_Client, Diversity_Deposit, Interest rate_Deposit, Commissions_Deposit, etc.), which are attributes and metrics are used aggregation operations. This cube is shown in Figure 4

Figure 3 - The Employees dimension

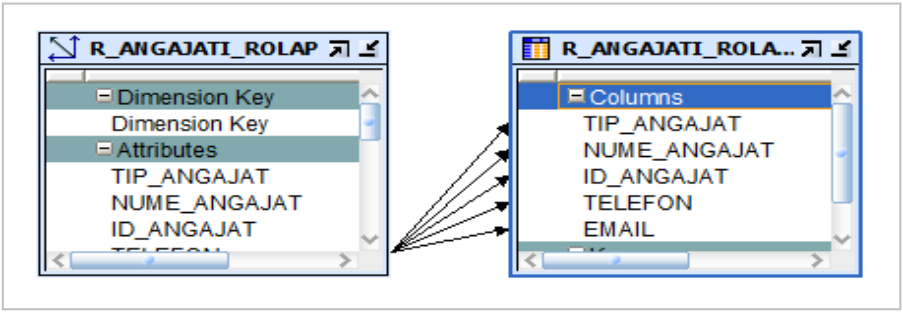
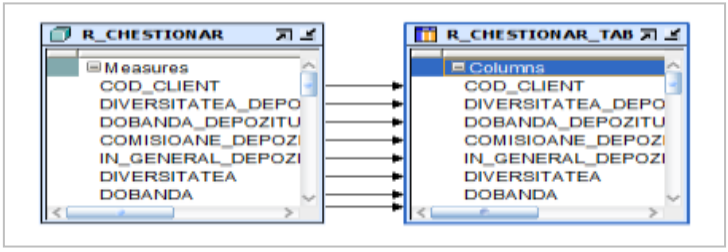


Figure 4 - Cube Questionnaire

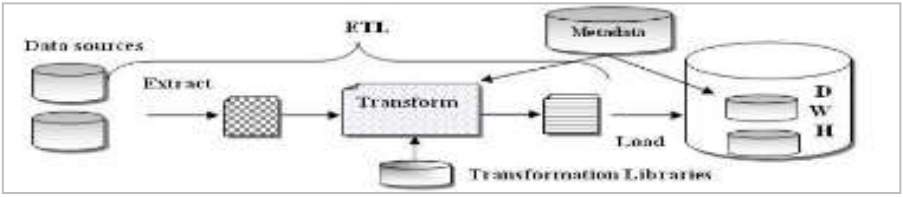


5. ETL PROCESS (Stancu Ana-Maria Ramona, 2017)

Traditional data repositories are static data warehouses that have been deployed and led to the ETL (Extract-Transform-Load) (M. Asif Naeem, Gillian Dobbie, Gerald Weber, 2008).

ETL means Extraction-Transformation-Loading. Using terminology Warehouse Builder ETL process is actually a data stream called mapping. Definitions mappings are established in the context of how destination. Mappings are developed in an Oracle database. Following the completion of mapping, Warehouse Builder generates Codee majority PL / SQL. Technology ETL (Extract-Transform-Load) supports relational formats (e.g. CSV files, tables in relational databases) that have found a way to examine objects stored. Figure 4 is the principle of LTE technology.

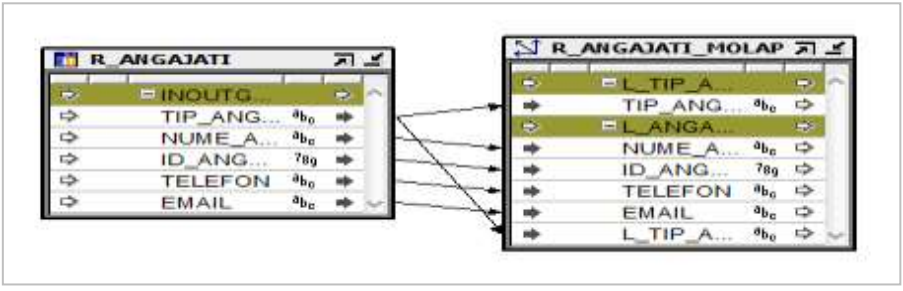
Figure 4 - The principle of LTE technology [Huong Morris, 2008]



Mapping is done both for dimensions and cubes. To achieve mappings from database tables must first indicate the link between the data source (table) and the object in the data store that will be loaded. Objects between that establish the links are brought into the workspace, and the links will be of 1: 1 or 1: n.

Mappings from database are EMPLOYEES, CUTOMERS_BANK, ACCOUNT, BRANCH, TYME, UNITS.

Figure 5 - Mapping Staff



Mapping EMPLOYEES

Employees have to perform mapping of the source module selected (SOURCE _SEMINAR) Table R_Staff and destination module (DESTINATION_SRV) R_Staff_Molap dimension I selected, and then I realized links between the attributes of the two (table and dimension). Staff mapping is shown in Figure 5

6. CONCLUSIONS

Data warehouses provide tools and architectures useful to management by understanding, organizing and using data in decision making.

Data warehouse features are: Subject orientation, integration, historical character, and persistence of data. Also, the data warehouse needs to focus on the enterprise, be change-resistant, be able to load large volumes of data in a short time, and have to be done for the data mining process.

The proposed data warehouse has to meet various purposes, such as: providing increased access to data for users, providing a single version of the truth, accurate record of the past, combined data / data access, separation of analytical and operational processing .

Data warehouses are an important part of Business Intelligence and will evolve as technology develops and new risks emerge. But when building a data repository, it must go from its essence, that is, listen to the users and analyze the issues to be solved.

In the first part we designed the bank data warehouse scheme that sells various banking products to individual clients and we noticed that the data can be uploaded from various external data sources and we identified the steps to accomplish it.

In the second part we designed a multidimensional model in which we used the "snowflake" model, and in the last part of the chapter we made the dimensions (employees, clients, account, branch, time) and the cubes (questionnaire, operations) MDG we have accomplished the ETL process (we have mapped for dimensions and cubes).

7. REFERENCES

- Huong Morris, Bringing Business Objects into Extract-Transform-Load (ETL), IEEE International Conference on e-Business Engineering, 978-0-7695-3395-7/08, 2008, DOI 10.1109/ICEBE.2008.72
- M. Asif Naeem, Gillian Dobbie, Gerald Weber, An Event-Based Near Real-Time Data Integration Architecture, Enterprise Distributed Object Computing Conference Workshops, 2008 12th, pp 401-404, ISBN 978-0-7695-3720-7
- Stancu Ana-Maria Ramona, 2017 “Soluţii de extragere a cunoştinţelor din volume mari de date”. Teză de doctorat, Academia de Studii Economice, Bucureşti