DATA WAREHOUSE MODELS FOR MULTINATIONAL GROUPS

PhD. Student **R.M. BOGZA**
Professor PhD. **Dorin ZAHARIE**
E-mail: romari_61@yahoo.com, zaharied@ase.ro
The Bucharest Academy of Economic Studies

ABSTRACT

Data warehouse models and architectures for multinational groups are presented. Some flows for processes involved in the DWH are also proposed. The paper also presents a possible model and architecture for a local data warehouse, with the layers and processes that are involved.

KEYWORDS: *local data warehouse, group data warehouse, data mart, layers, business intelligence*

1. INTRODUCTION

The paper presents models for a group data ware house that could be a multinational company. Each subsidiary, located within a country or region has its own local data warehouse, and one or more interface layers, through which the data are transferred to the group data warehouse. Each local data warehouse has an OUT layer that will do all processing necessary to translate data from the local DWH to the DWH group.

Each local DWH, depending on the volume of managed data may have several layers. Local DWH is updated daily.

Group data warehouse is updated daily or monthly (some dimensions may be daily, monthly etc.).

Local DWH is used for statutory local reporting and for internal company management.

Group DWH is used for reporting at the group.

For large enterprises with different types of activities, the following format can be used: if activities are big and bulky, and if their specifics vary greatly from one another, then we can create for each type of activity one DWH. But costs can be quite large in this case. Therefore, it is useful keep all data in one data warehouse.

DWH architecture and implemented models are very important to quickly access data required for various reports. Therefore, it is preferably to organize data in DWH in various data marts, each data mart being designed for a specific task. A specific data mart may be designed, built and should be used for assisting managers in that field, but there can be data marts for managing the entire company. The data marts can be used for profitability analysis of each type of activity, over a short, medium or long term.

The data for the group DWH are retrieved from the local DWH through a specific data mart. The local DWH supports the reporting activity for the subsidiary in each country / region. Dynamic reports are obtained from data-mart. The business has no direct access to data from the data store, only the data from data-mart's, or through dynamic reports, or through user interfaces. A group DWH data feed is made through interfaces, acquiring all data from local data stores, processes them to bring the format and level of granularity

desired for group DWH. This interface can be done through an extra level of staging whose role is to prepare data in a format and structure required by the DWH group.

The group DWH supplies the data-marts at group level. These can be accessed by the user or through a graphical interface or on the basis of their reports are obtained.

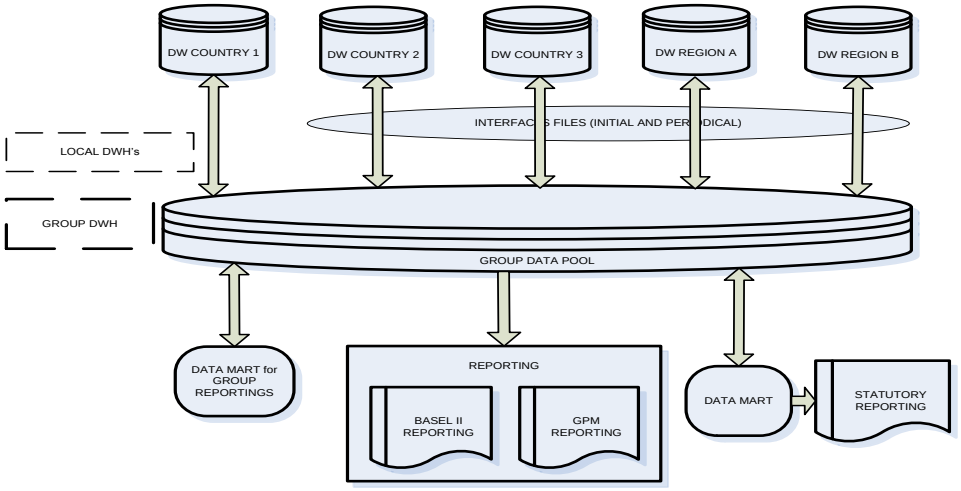


Fig. 1. DWH Model for a group / multinational company

2. PROCESSES AND LAYERS IN THE GROUP DWH

The modules and processes in the group DWH are presented in "Fig. 2".

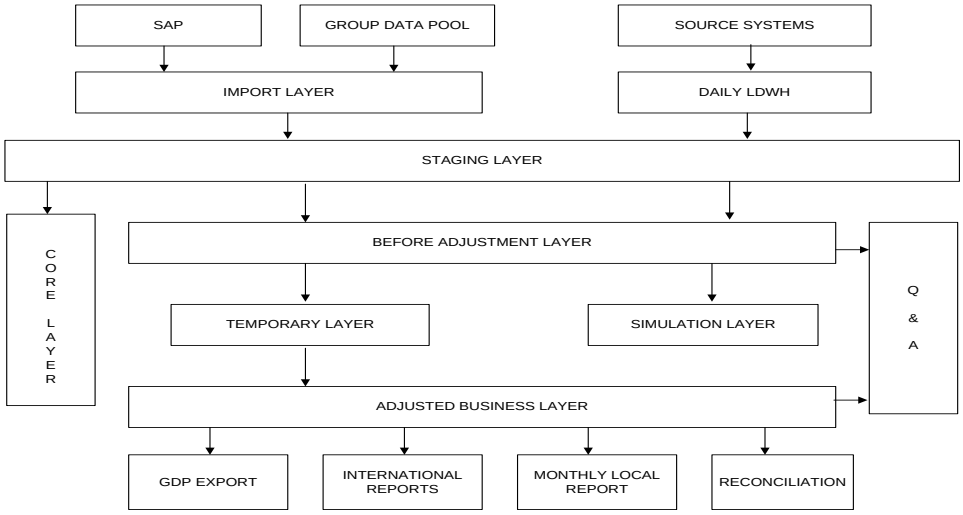


Fig.2. Processes and modules in a group DWH (for a multinational company)

The CORE layer contains data from:

- SAP Tables
- Transfer data from DWH + local transformations

The staging layer

- Has the same structure as import layer, only a few technical columns in addition for audit
- Inspection: Master - Detail and Detail – Master
- Rules for treatment of user errors
- Rules for treatment of data according to the regulations of group

Before adjustment layer

- The data is prepared in the required formats
- Certain rules are defined for loading data

The temporary layer

- Is the layer between DATA BEFORE AND AFTER ADJUSTMENT DATA
- Is used to run standard checks and rules (fixed rules) before checking non-standard rules (user rules)
- User Rules
- Fixed rules
- Group Rules
- Verifying Master-Detail
- Inspection Detail-Master
- Rules for treating errors

Simulation layer

- Is used to simulate user rules
- Data is in the same format as in adjusted for business ADJUSTED FOR BUSINESS LAYER
- Group rules
- User rules for the correction of data

After adjustment layer

- After adjusting the data, a back-up is done on the data and final reporting is done for
- International Bodies
- Internal Group Management
- regulatory bodies in the country where the group is located

3. Integration of a local data warehouse in a group data warehouse

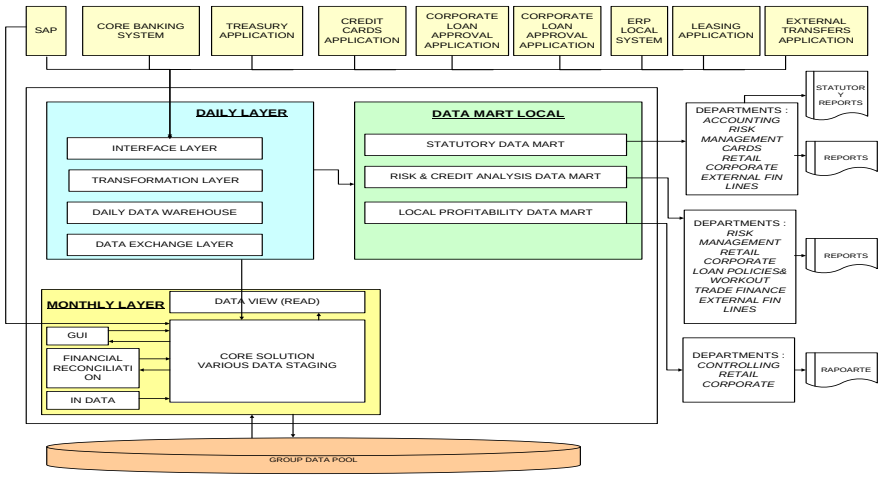


Fig. 4. Integration of a LDWH into a Group DWH (for a multinational company)

4. Model for a local DWH

4.1. Block shema for a local DWH

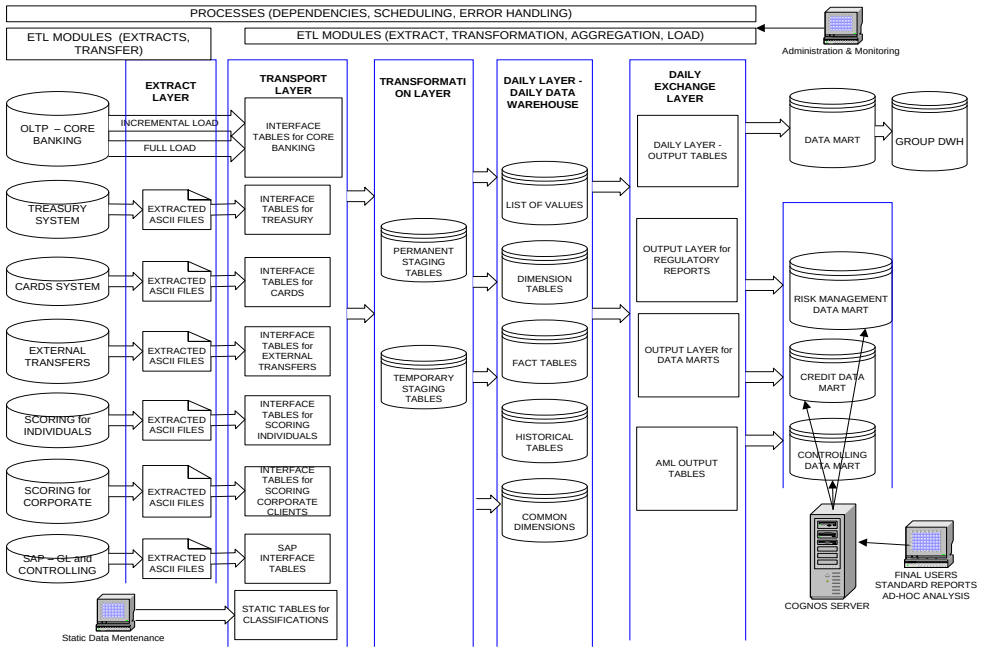


Fig. 5. Block diagram - a model for a local DWH

In a subsidiary there may be more applications, each serving a specific functionality within the company. If we take for example, a financial institution working with clients, such information subsystems are:

- Transactional System (core banking)
- The cards system
- Treasury system
- External transfers system
- The accounting module
- Corporate credit scoring system
- Retail credit scoring system
- Leasing IT system
- Catalogs, classifications

Data is extracted from each source system, and follow an ETL process.

4.2. Extract process – transport layer

Data files are generated in the source systems and transferred in a buffer zone, specifically designated for that. ETL processes responsible for extracting data, load data from the buffer zone in the transport entities, providing log files that are accessible through the environment management and monitoring processes. Through log files, the end users can verify the accuracy of load data. After extracting the correct data from the source system, the files are archived.

Data can be loaded from the buffer zone in the transport layer by two ways:

- Incremental Load – the data from input source systems are loaded into the transport layer of data warehouse only if it differs from similar data record of the last extract. It is used for large tables, with more then one million records or for those who rarely change; Examples: balances, deposits, accounts, loans, clients, etc.
- Full Load - the selected data from input sources, are fully extracted from the input sources (all the records and all the attributes), daily or with another chosen frequency; it is used for classifications, catalogs, which are smaller, for entities that are changing frequently, such as: transactions, commissions, interest rates, or for the initial load.

Results of the extractions from source entities are found in tables of layer transport interface. In this phase, the data are not changed, but some auditing columns are added.

4.3. Transformation process – the staging level

From the transport layer, the data are cleaned and processed by the transformation process to be ready for loading in the data warehouse. We call this level: the staging layer, where data are subject of the data leansi- ng process and, structures are mapped on the structures from the next layer, being prepared for loading in to the data warehouse.

On this level:

- Data is transformed according to the model implemented in the data warehouse
- Various filters and transformations are applied on data
- The surrogate (technical) keys are generated
- The data from various source systems are integrated.

4.4. The load process – the daily layer

Data are loaded into the data warehouse, where you can find various types of entities:

- reference tables: lists of possible values for certain attributes / catalogs / classifications
- Dimension tables
- History tables
- Common Facts / Facts tables

4.5. The daily exchange layer

From the daily layer, the data is retrieved in the required format for the next level, DWH OUT, through the process of acquisition and data processing in order to get prepared for:

- interface with group data warehouse
- Data-mart for special processing, ie preventing and combating money laundering
- Data-mart for statutory reports
- Data-mart for reporting financial performance / economic indicators
- Controlling Data-mart
- Risk management data-mart
- Credit data-mart
- Data-mart for treasury

The data are supplied from the DWH Out in the format agreed with the other target systems.

4.6. The interface layer between the daily exchange layer and the group DWH

In this layer, the data is transformed in order to be prepared for a direct load into the group DWH.

4.7. Reporting process – reporting server

The prepared data from data-marts are accessed by the user directly, or by a GUI, or through dynamic reports, which are implemented on a server that allows the OLAP techniques and the data mining techniques. It can be a COGNOS server.

4.8. Administration and monitoring layer

Performing all the processes that allow the data extraction from the source systems, their cleaning and transformation , preparing them for loading in the data warehouse, further processing for interfacing with the group DWH and for populating the data-marts , can be managed through a specific management and monitoring processes.

It allows:

- Planning Processes
- running automatic processes
- re-running processes that didn't finish because of errors
- conditional running of processes

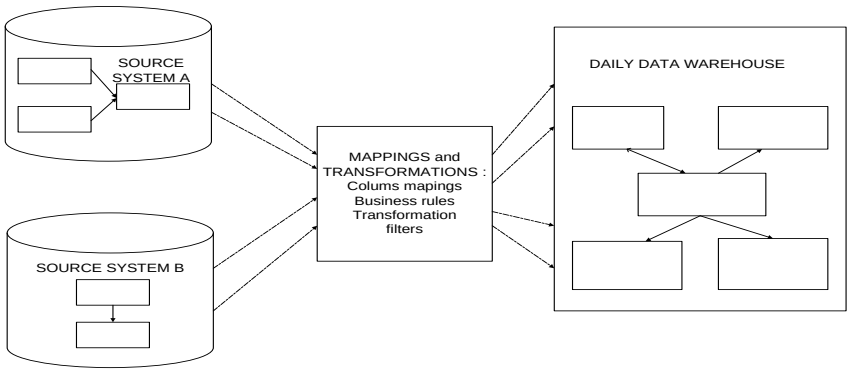
5. Principles for implementation of a data warehouse

- in the local data warehouse , data is generally normalized
- data in data marts are not normalized; they are organized in dimension tables, with or without a hierarchy of dimensions (eg time / organizational structure), fact tables (that contain measured values of the scale)
- the granularity of facts in the data marts is high (that means the fact data are stored at a very detailed level), allowing to get a full detail of the measures across the hierarchy of the dimensions or across various dimensions and hierarchy of the dimensions
- no data is deleted, everything is kept in the history ; for keeping the deleted data, there is a flag, deleted flag='Y'
- the most recent data is kept in 'snapshot' tables, that contain the last image of that type of data; from this 'snapshot' tables, the historical tables are loaded
- the historical data may be kept partitioned after the data, for a more rapid access to the data
- Tables are organized logically:
  - ✓ reference tables: catalogs and lists of values,
  - ✓ primary data tables,
  - ✓ tables of relationship,
  - ✓ tables of facts,
  - ✓ tables and figures combined size (usually those describing events or states),
  - ✓ historical tables;

- for a more rapid and accurate access to the data, there are no null's in dwh (in case of null's , a default data is defined)
- names of the tables is relatively standardized (according to rules established apriori), so that they can be relatively easily identified in phase of analysis for business requirement; example : the dimension tables may have the suffix _DIM, the fact tables may have the suffix : _CF, historical tables , the suffix : _H; the suffix is added to the name of the current table
- column names of each entity are standardized: each table is a prefix, which is found as a prefix to the names of all attributes of that table, the second part of the column name refers to the area of interest of the attribute, the last part of the name attribute may refer to field the values for that attribute, for example:
  - ✓ _KEY - may mean a technical key for that entity
  - ✓ _ID - may appoint an ID from the source systems
  - ✓ _DESC, name - may mean an attribute name
  - ✓ _DATE - date calendar
  - ✓ _FLAG - a possible indicatorhaving the value of Y or N
- all tables are audited, they may have some audit columns:
  - ✓ source system ID
  - ✓ ID registration process that inserted that
  - ✓ Date and time of actual insertion that registration entity
  - ✓ reporting date on which that entry relates
  - ✓ ID ETL process that changed the record that
  - ✓ Datetime of the ETL process that inserted / updated a line in the table
  - ✓ Deleted flag - could have the values Y / N, the default = N
  - ✓ Updated flag - can range Y / N, the default = N
  - ✓ Start date of validity for that record – valid from date
  - ✓ end date validity for that record – valid to date
  - ✓ last loading date
- All audit columns have default values, which are established at the beginning of the technical implementation of the DWH.

6. Data flow in a DWH

The data flow is presented in Fig. 6:



The data flow in the data warehouse is determined by :

- the data stored in DWH
- the volume of data stored in DWH
- the mappings and transformations applied to the data imported from the source systems
- the data model in the DWH

7. Conclusions

The model and the architecture presented above offers the possibility to manage a large amount of data in order to obtain results at various levels of an organization or of a group. Of course, there are also other odels and architectures, that are much more simpler and not so many layers. The layers of extract, transport and ransformation/ staging can be performed in a special application. But also, this application should have inside it, a sibling architecture.

The advantages of such an architecture is that it really offers the possibility to extract data from various systems and to assemble them, with an adequate model of data, in the desired format.

References

1. Peter Baker, Marco Canessa – Warehouse Design: A Structured Approach – *European Journal of Operational Research* 193 (2009) 425-436
2. Rodica Maria Bogza – *Strategic Intelligence with BI Competency Center* - II International science conference "Knowledge Society", Nessebar, Bulgaria
3. Robert S. Kaplan, David P. Norton - *The Strategy focused organization*, 2001, Harvard Business School Publishing Corporation
4. Dorin Zaharie, Felicia Albescu, Veronica Ivancenco, Irina Bojan, Corina Vasilescu – *Sisteme informatice pentru asistarea deciziei*, Editura Dual Tech 2001
5. Robert Reix – *Systèmes d'information et management des organisations*, Vuibert, Edition : 5^{e} edition (2004)