# Research Areas in Data Warehouse Systems' Audit

Prof. Traian SURCEL, PhD, Lect. Marian STOICA, PhD
Economic Informatics Department, Academy of Economic Studies, Bucharest

*The data warehouse is the core of decision support systems. The audit of data warehouse systems aims to certify the conformity, credibility and superior performance of the system. This paper presents some aspects of data warehouse audit, investigating the following areas: methodological and architectural problems, data quality, system maintenance, documentation and user training. The ultimate purpose is to ensure the continuity, efficiency and consolidation of businesses based on this kind of decision support systems.*
***Keywords:*** *audit, data warehouse, methodology, change management, architecture, data quality, maintenance.*

## 1 Research areas

The general purpose of auditing a data warehouse system is linked to the evaluation of its daily functioning, of the implementation, maintenance and development processes aiming to identify directions of consolidation and improvement of using it. Standards and best practices direct the audit towards the following areas:
- ✖ methodology to create the data warehouse;
- ✖ change management;
- ✖ system architecture;
- ✖ data quality;
- ✖ maintenance and performance;
- ✖ documentation.

For each of these areas, operative auditing work must follow these directions:
- ✖ analyze the current situation and gather audit samples;
- ✖ evaluate conformity with theoretical approaches and best practices;
- ✖ elaborate recommendations.

## 2. Methodology to create a data warehouse

The methodology to create a data warehouse is the first aspect the auditor must analyze. This requirement is reflected by the presence and the content of the design project and implementation plan for the data warehouse.

Because the data warehouse must integrate with the complex informational environment of the company, the design of the data warehouse must be a component of the design of the informational system of the company, which, in turn, is a component of the Enterprise Consolidation Plan (EPC).

Recommendations made by the auditor about the methodological aspects will deal with the necessity of long term goals and, on short term, the existence of a strategy to achieve those goals and a plan to redesign applications that modify the original design.

These modifications appear as a result of an "impact analysis" which must be carried out periodically, every month. This analysis deals with the impact of using the data warehouse on the company management and functioning. From a methodological point of view, nothing should be forgotten: goals, strategy, impact analysis, change plan (figure 1).

## 3. Changes

Changes are driven by the specific elements of data warehouse design, elements the auditor must know and deal with. The following elements stand out:
- ✖ dynamics of unpredictable requirements, representing those new informational requirements from the beneficiaries, as they realize what can be done with a data warehouse;
- ✖ non-standard informational environment as opposed to ERP systems that use a relatively standardized environment;
- ✖ cross organizational approach of applications, which means a series of applications deal with problems that belong to several departments;
- ✖ post implementation problems.

There must be a change management to coordinate and control their implementation, first inside the development environment and, after

passing the acceptance test, in the real system. For an efficient coordination of the changes, the auditor might recommend and verify a better communication and notification through support procedures in order to document the changes. Small changes may be implemented straight into the real system, but only with a close monitoring after implementation. It is a fast solution, with a small risk, and it is better than the standard procedure mentioned above.
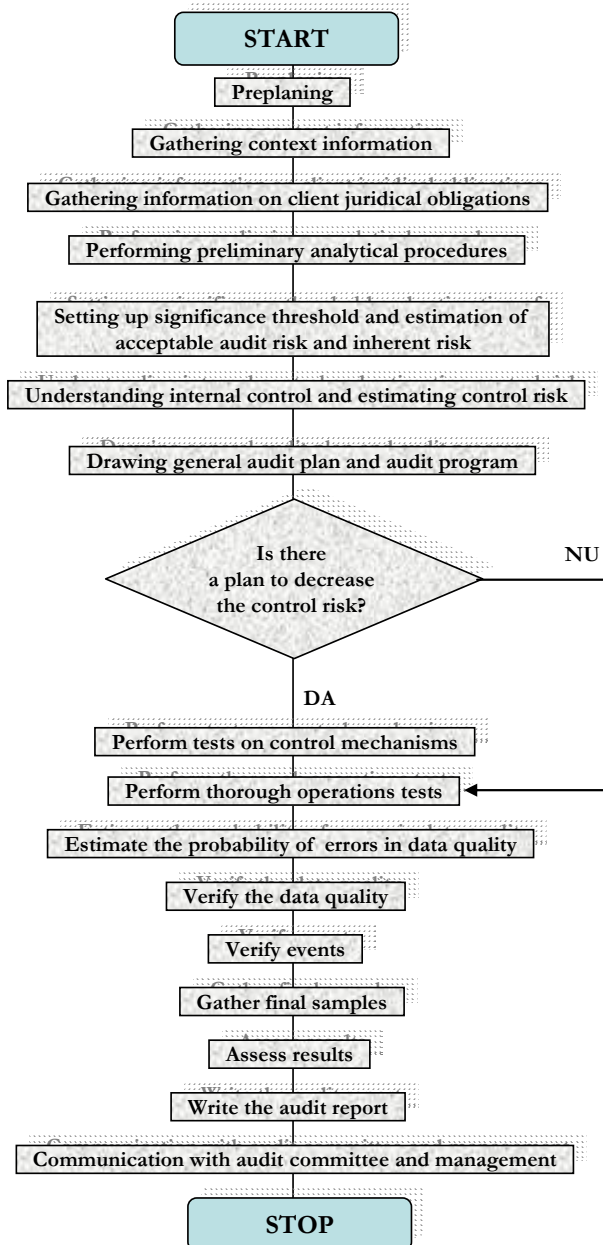


**Fig. 1.** Sinteza procesului de audit
(adaptare după Arens A. şi Loebbecke J.)

## 4. System architecture
Data warehouse architecture is another point of interest for the audit. In order to comply

with the best practices, the architecture must ensure scalability, integrity and flexibility for the system. The auditor must analyze two other elements:
✖ refreshment strategy
✖ operational data fund (ODS), created parallel to the data warehouse.

*Flexibility* means using configurable parameters for the system architecture and the possibility to add new objects to the warehouse. An element that leads to increased flexibility is the use of descriptions instead of codes, which leads to increased performance in creation of cubes and reports. Using the "always include" options in cube creation after upgrading the data series and daily "clean house" also lead to increased flexibility.

## 5. Refreshment strategy
Refreshment strategy is based on refreshment rate. A way to implement the refreshment rate considers the creation date of the record and the increase of refreshment period with a certain increment rate. The disadvantage of this method is that it leads to restoring more records than actually modified.

Another, more efficient, implementation uses a refreshment rate based on the time of the last modification of a record. To restore older transactions, a higher incrementation rate of the recording time marking is recommended.

Indexing increases the performance of loading, extracting and reporting data, more so on large data volumes. Refreshment strategy must be assessed from the point of view of maintaining and restoring indexes. "Update & insert" technique uses existing indexes and is faster, while "delete & append" rebuilds the indexes by "drop and rebuild after refresh" technique, which consumes more time but yields better performance when creating cubes.

ODS is an operational data fund created and maintained in parallel with the data warehouse, with the goal of a better response time and logging the activity of the data warehouse. Data loaded into ODS must be verified before their confirmation in the data warehouse.

Data warehouse audit must asses the feasibility of developing an ODS on the same storage environment as the warehouse it is associated with.

## 6. Data quality

Data quality is a first target of auditing a Data Warehouse, focused on evaluation of data verification and error correction. The audit must identify automated data quality checks, complex checks and cross checks. These checks must signal erroneous data and return them to the source. The audit must asses the checking procedures from the point of view of preventing erroneous data from loading into the data warehouse.

A useful recommendation is designating a process owner or supervisor that will answer to all problems linked to data quality, both for input and reports. He will also be responsible for the operational procedures in the transactional system. This supports the separation and delegation of responsibilities during exploitation of the data warehouse.

## 7. System maintenance

Data warehouse maintenance is not a linear activity; it is more than a suite of technological operations of updating the data and creating reports. Auditing this area must look carefully into:
- ✖ monitoring data warehouse system modifications;
- ✖ preventing resource unavailability;
- ✖ resolving system crashes and recovery of data in case of disaster;
- ✖ help desk system.

Data Warehouse *system modifications* fall into two categories: modifications forced by real economy and information environment modifications. Real economy modifications are a result of data, logical structures change, rules for data protection and data violation error.

*Informational environment system modifications* are both of hardware and software nature, but mostly software: patches, performance backup systems, new applications etc.

*Preventing resource unavailability* concerns reliability of hardware, networks, connectivity etc. Also, it deals with software resources locking, crashes of SQL server, parallel processes accessing the same resources.

*Resolving crashes* is essential to ensure system reliability. The auditor must acknowledge that once the scheme of logical error is built and tested, there should be no more hardware and software errors.

It is necessary to verify that there are no applications that keep running after a crash or an error. Procedures that solve errors must seek to minimize the error impact on users. A better coordination of data warehouse refreshment needs a monitoring of the *crash rate* and *success rate*. An efficient approach is refreshing the data on demand; it also implies delegation of control to the end user, increasing the efficiency of the activity.

The auditor must identify and evaluate *disaster recovery procedures*, which add to the crash solving procedures. Basic premise for disaster recovery is the existence of backup copies. Copies should be placed on different machines. Fast recovery will answer the need for *continuous availability* of the system.

*Help Desk* designates support for maintaining and implementing changes. From a procedural point of view, help desk includes all actions for identification of ways and means to improve functioning and development of the data warehouse system. This includes following actions:
- ✖ meetings to find new designs;
- ✖ interviews;
- ✖ research of data and processing rules;
- ✖ impact analyses;
- ✖ analyzing the functioning of new procedures in the new modified context.

Help desk is the way to stimulate designers and users to analyze system functioning and find new ideas to plan the strategy to improve the functioning of the data warehouse. This strategy consists of:
- ✖ control procedures for data quality and data restoring;
- ✖ updating the task list and responsibility delegation;
- ✖ reviewing help files and improving communication and change notification.

## 8. Documentation and training

Documentation and training are the support for knowledge, operating and data interpretation skills transfer. There is a need for a strategy to create the documentation and plan and organize the training of users. This strategy must be part of the general strategy of creating the data warehouse system.

Specifications for informational demands, design specifications and term glossaries must be complete, dated and constantly updated. Documentation must clearly highlight data architecture, process architecture and system technical architecture.

We must highlight again the importance of specifications for coordinating refreshment processes, complete and correct explanations of back step engineering used to capture the last modification, the last updating.

Definitions, economic rules and terms that belong to the business community must be included in a glossary that specialists name "Gognos Glossary and End User Guide". Documentation assessment also evaluates the existence of information concerning creation and maintaining standards and recommendations for the best practices implemented in the audited system.

Training must differentiate competence and responsibility levels of the users. There are three categories: regular users, professional users, help desk users involved in perfecting the system – extending and diversifying applicative functionality.

Auditing also targets continuous training for the new procedures, which means internal seminaries, discussions forums to establish the feed-back needed to stimulate new opinions about current functioning and future evolution of the system. This is an additional element that increases user satisfaction.

User satisfaction is analyzed from the point of view of achieving goals established at design time, as well as system functioning. User satisfaction is reflected by the frequency, nature and structure of additional requirements to the initial system. All these aspects are found in the level of benefits, recording and acknowledging of catching new business opportunities, ensuring continuity and consolidating the business that exploits the data warehouse system.

## References
- Arens A., Loebbecke J., Audit. *O abordare integrată*, Editura ARC, Chişinău 2003
- Champlain Jack, - *Auditing Information Systems,* John Wiley & Sons, New Jersey, 2003
- Ivan I., Noşca G., Capisizu S. - *Auditul sistemelor informatice*, Editura ASE, Bucureşti, 2005
- International Federation of Accountants, - *Audit financiar 2000, Standarde de audit*, Editura Economică, Bucureşti, 2002
- Jacques Renard – *Théorie et practique de l'audit interne*, Editions d'Organisation, Paris, France 2002