

## DATA WAREHOUSE DEVELOPMENT MANAGEMENT

UDC 004.6 005.8

**Slavoljub Milovanović**

Faculty of Economics, University of Niš, Serbia

**Abstract.** *Data warehousing is a new decision support technology. It offers organization managers and executives modern tools for data access, data analyses, querying and reporting. However, data warehouse building is not an easy task. There are problems in defining information requirements, designing system specifications, choosing hardware and software products and implementation of data warehouse. Purpose of this paper is to present iterative methodology for data warehouse development. The major part of the paper is dedicated to data warehouse development process with its phases: planning, analysis and design, implementation and exploitation.*

**Key Words:** *Data Warehouse, Project Management, System Development, Business Intelligence.*

### 1. INTRODUCTION

Data warehouses capture historical business data allowing business managers to make informed decisions based on the data. Because decision-making is improved when managers analyze similar historical situations, many organizations are defining the information produced by querying a data warehouse as *business intelligence*.

Therefore, data warehouse is a location or a facility for storing facts and information about business events in an organization. A data warehouse system provides a centralization of organization data, provides tools that allow its users to effectively process the data into information without a high degree of technical support and exists in a well-managed environment. Also, it is built on an open and scalable architecture that will handle future expansion of data.

Data warehouse must be able to capture operational data, transform the operational transaction data, store the transformed data in a data warehouse and deliver business intelligence upon receipt of a user data request or query [1, 2].

The reasons for growth in this area are many. With regard to data, most organizations now have access to huge amount of operational data that exists in transactional databases.

With regard to user tools, the technology of user computing has reached a point where organizations can now effectively allow the users to navigate organization databases without a high need for technical support. With regard to organization management, executives are realizing that the only way to sustain and gain a competitive advantage in the market is to leverage information better.

A data warehouse is designed to support decision support queries and applications. A data warehouse assists an organization in analyzing its business over time. Users of data warehouse systems can analyze data to spot trends, determine problems, and compare business techniques in a historical context. The processing that these systems support includes complex queries, ad hoc reporting, and generating standard reports.

The concept of data warehouses is not new. In the past, there were executive information systems (EIS), and decision support systems (DSS), which are antecedents of today's data warehouse systems. However, data warehouse systems have become a rapidly expanding requirement for most information system departments. Responsibility for the development of an effective data warehouse lay on the department and the challenge is very considerable. In order for the department to be successful, it has to adopt iterative and spiral approach to data warehouse development. This paper will just present the iterative methodology of data warehouse development. After basic explanations of the methodology, next parts of the paper are dedicated to phases of the development process: data warehouse system planning, analysis and design, implementation, preparation for exploitation and review.

## 2. ITERATIVE APPROACH TO DATA WAREHOUSE DEVELOPMENT

A data warehouse is a single source of historically significant organization information. All of the entities contained in a data warehouse are interconnected; therefore, the processes that comprise a data warehouse should also be interconnected. The project management of data warehouse relies on these interconnections and the interconnections cause an iterative development methodology. The development process is meant to quickly deliver, in an iterative fashion, the subject-oriented data warehouses required by the target users. The four stages of the methodology are: 1. data warehouse system planning, 2. analysis and design, 3. implementation and 4. preparation for exploitation and review [3].

The spiral development approach assists developers in more rapid delivery and discovery cycles, which allows development staff to avoid lengthy development cycles that miss the requirement. A rapid feedback cycle, in association with a rapid delivery cycle, assists in educating the development staff on how best to meet the data warehouse users' needs. The intent is to utilize an iterative development approach for building an organization data warehouse. Starting small enables development team to lay the groundwork for a successful, ongoing warehouse strategy. This initial project will verify or upgrade the required network computing infrastructure to support a organization-wide implementation [4].

### 2.1 System planning

The system planning phase of a data warehouse development effort allows a project's scope to be set, assisting the project management staff in sizing the time and effort required to deliver a subject-oriented data warehouse, or subsystem of an organization data

warehouse. Within the business scope, it is important to state the business domain that will be covered by your project, namely: the business functions supported; the decision making enabled; and the users or key organizational personnel who have benefit from a data warehouse. The system planning phase provides several key deliverables: a purpose definition, goals and objectives, users identification, critical success factors, and risks and constraints.

**Purpose.** The purpose definition is an articulation of the scope and reason for the subject-oriented data warehouse that is being planned. The project must clearly define a purpose that all developers understand and use as their boundary for development.

**Goals.** The goals and objectives flow from the purpose statement, and define the specific accomplishments desired from the development effort. Typically, organization must define business-oriented objectives. Technology oriented goals and objectives may also be included within project's purpose statement. These typically include items such as building a subject-oriented data warehouse to verify the planned architecture and verify the architecture's requirements.

**Users Identification.** Users are the individuals who will benefit most or be impacted most by the delivery of the data warehouse. Developers need to gather significant information on them to best determine their needs. This information includes the users' background, interests, job responsibilities, success measures, and all other detail helpful to shaping the information that will be delivered in the data warehouse. This data will be dynamic and change over time. Therefore, the data should allow the development team to identify users and manage any changes during the entire data warehouse development effort.

**Critical Success Factors.** The critical success factors of a data warehouse project should be clearly documented. This benefits the development team and organization management in determining whether the development effort was successful. A data warehousing project typically must meet the following criteria in order to be successful:

- **Integrated view of business.** The data warehouse will provide the organization with the ability to pull together information from a variety of sources at regular intervals to construct an integrated view of business activities.
- **Rapid response to users' requests.** To gain value from this information, rapid response to requests for information is crucial. The tools that access a data warehouse must cross a reporting continuum that includes executive information systems, decision support systems, ad hoc queries, and production reporting.
- **Satisfy users' needs, not technical requirements.** The data and reports that are produced from a data warehouse should be what the users want and need—not what the Information Systems department wants to give the users.
- **Control data warehousing process and system.** In this context, it means controlling the processes of loading, the access methods, the security, and in general the system. The data is a valuable asset of the organization, so specific controls must be placed on this asset so it stays in the hands of the organization, not of the competition.

If these items are successfully implemented, the development team will deliver to users the ability to use and integrate information technology products throughout the data warehouse. This allows the organization and users to choose the best vendor for each warehouse component and to derive full value from its data warehousing investment.

**Risks in data warehouse project.** Each risk area should be completely defined prior to the project's full implementation. Risks should be documented and communicated to users so that the users can assist in controlling the factors that minimize the risk. Risks fall into technical and business areas and usually include: expense exceeds planned budget, development team does not have enough experience with data warehousing, improper division of control over the budget and scheduling of these resources, lack of a user-organization sponsor and lack of a direct executive sponsor, problems in solving "buy versus build" dilemma.

All projects have problems at different points during their life cycle. If these items are known before a project begins, define them as risks and build a contingency plan to avoid any major impact. Within the spiral methodology, risks are identified within the planning phase. The assessment and contingency planning will be completed in the analysis and design phase.

The initial plan is a starting point from which the development team will determine the size of a project. Until development team reaches final agreement on time, resources, and cost, the project plan will not be solidified.

## 2.2 Analysis and design

The analysis and design phase of a data warehouse development better defines the details of a project's scope. The development management team begins to develop detailed answers to *what*, *how*, and *who* questions, such as the following: What information will be contained within the data warehouse? What alternatives are available for delivering the desired information and results to the user? How will the risks identified in the system planning phase be resolved and addressed to minimize their impact on the overall project? Who are the target users? [5, 6]

**Information Domain.** *Information domain* refers to the data that provides users with their required business intelligence. The definition of the information domain provides the development team with information and knowledge that is required for a successful project. To fully deliver the information domain, the development team needs to attend to the following: define overall requirements; document existing systems and systems environment; identify and rank the candidate applications that will utilize the data within the data warehouse; build a transition model that identifies dimensions, facts, and time stamping algorithms for extracting information from operational systems and for placing them in the data warehouse

System analysts will be able to take the previously collected logical models and contacts to begin delivering more detailed information on the source applications. Together with business analysts and user community, the analysts will be able to further define requirements such as the level of data granularity, the level of aggregation, the frequency of data loading, and the number of time periods to maintain.

**Alternative Evaluation.** Every development project has alternative solutions that developers should evaluate. The following questions should be asked at the start of a data warehouse development and continuously throughout the project: Is the purpose already accomplished elsewhere within the organization's information systems? Does a publicly available software product solve the entire problem, or is it a partial solution? Are incomplete solutions available in either of the above places?

**Risk Resolution Planning.** Risk assessment and resolution planning are targeted to formulating a plan to alleviate and avoid any of the previously defined risks. The development management team must assign sufficient priority and resources to complete a risk assessment and contingency plan. If the proper resources and time are not assigned to this effort, the risks may greatly impact the project team's ability to successfully deliver the required software components and applications.

**Target Users Information.** The data warehouse analysts should define the target users and the specifics of information that are important to them. An adequate description of the users' needs and desires should be developed. Such documentation will greatly assist in the overall development process. Developers will continue to add more details of the information domain and users as well as their security and access needs.

**Specification Assembly.** As a data warehousing project proceeds, specifications should be built for each component defined within the information domain. The primary purpose at this point is to provide a valid reference point for detail design activity and acceptance testing criteria.

The specifications refine the goals and objectives in more specific terms. The detailing of these statements includes adding a layer of constraints, business rules, and other requirements. A specification acts as a guidebook for the implementers who create the actual data warehouse components. The specification should fully document the applications, data, and technology components (including tools) to utilize in delivery of the distinct data warehouse parts. The specification should also identify any restrictions based on standards, guidelines, or policies previously defined for the project.

### 2.3 Implementation

A good data warehouse design accounts for all of the previously defined elements - the purpose, goals, objectives, success factors, risks, constraints, users, and information domain - and combines them to produce a plan for properly implementing a data warehouse. Very practical issues are involved in the implementation of a data warehouse, such as how the user will access the information, from which the development team will gain a sense of judgment and experience. This allows the team to begin delivering the ultimate data warehouse [7].

A successful data warehouse implementation guarantees that all elements work in concert and harmony to provide the user with the proper system. The analysis and design process was greatly influenced by the knowledge of what is possible within the data warehouse. The implementation process begins to focus on the best way to physically implement the design.

**Detail Design.** The detail design of data warehouse components involves several tasks, including the following:

- **Crystallize the needs of users and complete the model for the data.** It includes designing the dimensional hierarchies and business measures required to deliver the required business intelligence. These detail designs will specify access mechanisms and further define standardized data structures.
- **Map operational data sources.** These maps and data definitions should be placed in the metadata repository. As the map evolves, so will the transformation logic required to deliver data to the warehouse.

- **Detailing physical requirements.** The rules and configuration management information will be physically defined to the data warehouse metadata repository. These items include factors such as business rules or constraints, distribution or replication requirements, indexes, and partitioning strategies.
- **Develop user applications.** The detailed plan will need to define the applications and tools that will be utilized by the user to access the newly placed data in the warehouse.
- **Determining quality tasks.** It is important to define the data integrity process in detail design. This allows the development process to begin performing data validation so that users get right the resultant information.

**Coding.** This implementation activity signifies that resources will be physically implementing the system. This means the following: the database administrator will build or modify the physical data structures that house the new data; the development resources will build the transformation routines and integration logic to move data from the source to the warehouse; the administrative staff will also begin to automate the extraction, movement, and transformation processes as well as other administrative tasks such as backup and recovery schemes.

The desired outcome of these tasks is to build a component, subject-oriented data warehouse that can be integrated into the overall organization data warehouse architecture and released to the users.

**Testing.** Testing could be unit testing and integration testing. Each of these testing techniques should be utilized to gather data on the quality of all delivered components.

After individual components have been built, the process of testing them commences. Unit testing verifies single programs, stored procedures, and other modules in an isolated test environment. Unit testing is often done by the development person who produced the code, because he or she best understands what the component does.

Integration tests take the individual components and verify their interfaces with other components of the overall data warehouse system and subsystems. Integration testing should be done by a dedicated quality assurance team.

#### 2.4 Preparation for exploitation and review

After a data warehouse is built, development team must guarantee success by formally defining a deployment strategy. This includes the concepts of training users, obtaining their acceptance and feedback, and generally promoting the availability of the data warehouse throughout the organization.

**Training.** Training users how to use delivered components is extremely important to the overall acceptance and usage of a data warehouse. The training should focus on the following points: an introduction to data warehouse concepts, users' view of the data model, how users access the data warehouse, a clear definition of the tools and the type of analysis each tool provides and how to utilize the tools provided in the architecture.

Training should be provided by seasoned trainers who have a way of communicating with and understanding the needs of their students. The course should be conducted close to the users' environment and should include exercises that demonstrate how the data warehouse delivers the users' requirements.

**Acceptance.** Acceptance testing validates that the system or a component of the data warehouse system matches the users' requirements. In an acceptance test, development team should validate that the software can be installed and operated effectively within the users' environment. It is also advisable to conduct usability tests with the users. Acceptance tests clearly define any shortcomings that exist in the delivered components as well as provide valuable feedback for the development staff in the areas of usability, content improvement, accuracy of data, and required alternate views into the data warehouse.

**Promotion.** The main goal of promoting the completed components of a data warehouse is to keep users informed about data warehouse content. During the promotion of a new data warehouse, it is important to obtain skills in areas such as public relations and mass communication. At times -primarily during the introduction of new components for a data warehouse - users will experience information overload. So, proper infrastructure should be built to give users an educational continuum and to give developers ongoing feedback. Forums such as Lotus Notes discussion databases, e-mail, Usenet newsgroups, and World Wide Web sites are prime targets for such forums.

**Project Review.** The organization data warehouse team should reconvene after each completed data warehouse development project to determine if the architecture is in fact valid and working. More than likely, a majority of the architecture will be viewed as sound, while several improvements may be noted. This should be further documented within the standard document and enforced in further development efforts. And, importantly, it should be retrofitted to the current implementation of any subject-oriented data warehouse or the initial project to constitute a standard.

### 3. CONCLUSIONS

Competitive advantage in today turbulent business environment follows from information (historical, geographic, across business units, and across product lines). Information on how to respond to a product opportunity, a competitive threat, a trade imbalance, or a political challenge can be mined from the data contained across the various departments within an organization.

Extracting this information from a huge amount of data means redefining the concept of decision support. Managers need to operate in discovery mode, constructing their ad hoc and quick queries. The interactive access to information coupled with the promise of rapid response has become the focus of today's data warehouses.

A data warehouse development is iterative in nature. Business analysts and executive users will discover new information that they need as they become proficient in using a data warehouse. Therefore, it is important to implement a project management in which development team will deliver smaller projects at a more rapid pace than in traditional systems development. The spiral development methodology is recommended for delivery of data warehouse components, or subject-oriented data warehouses, because it achieves such results. Within this methodology, there are following phases: data warehouse systems planning, analysis and design, implementation, deployment.

Within the planning phase, development team defines a project and provides a comprehensive plan on its proposed delivery. During the analysis and design phase, development team further specifies and gains users' reaction on the development effort results.

During the implementation phase, development team builds the components to be placed in the data warehouse system and architecture. Upon completion of development and testing, development team deploys the system to the user community with a lot of training. This phase should guarantee usage of the organization knowledge base.

Following the entire development cycle, the development team should reconvene and perform a review of the architectural blueprint. Because a data warehouse is a dynamic environment based on constantly changing data and business processes, the development process is never complete. Development team will continue to develop additional components for organization data warehouse. In addition, after development team completes first subject-oriented data warehouse, it will be time to proceed to the next.

#### REFERENCES

1. Ponniah, P. (2001), *Data Warehousing Fundamentals - A Comprehensive Guide for IT Professionals*, John Wiley&Sons, Inc., New York.
2. Yao, J. E., Liu, C., Chen, Q. and Lu, J. (2006), Administering and Managing a Data Warehouse, in: *Encyclopedia of Data Warehousing and Mining* (Edited by Wang, J.), Idea Group Inc., Hershey, pp. 17-22.
3. Taylor, J. (2004), *Managing Information Technology Projects - Applying Project Management Strategies to Software, Hardware, and Integration Initiatives*, American Management Association, New York.
4. Inmon, B. (2005), *Building the Data Warehouse*, Wiley Publishing, Inc., Indianapolis.
5. Imhoff, C., Gallemmo, N. and Geiger, J. G., (2002) *Mastering Data Warehouse Design - Relational and Dimensional Techniques*, Wiley Publishing, Inc., Indianapolis.
6. Kimball, R. and Ross, M. (2002), *The Data Warehouse Toolkit - The Complete Guide to Dimensional Modeling*, John Wiley & Sons, Inc., New York.
7. Mundy, J., Thornthwaite, W. and Kimball, R. (2006), *The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset*, Wiley Publishing, Inc., Indianapolis.

## UPRAVLJANJE RAZVOJEM SKLADIŠTA PODATAKA

**Slavoljub Milovanović**

*Skladištenje podataka je nova tehnologija za podršku odlučivanju. Ona nudi rukovodiocima u organizacijama savremene alate za pristup podacima, analizu podataka, postavljanje upita i izveštavanje. Međutim, izgradnja skladišta podataka nije lak zadatak, jer postoje problemi u definisanju informacionih zahteva, dizajniranju specifikacija sistema, izboru hardverskih i softverskih proizvoda i implementaciji skladišta podataka. Osnovni cilj ovog rada je da prezentira iterativnu metodologiju razvoja skladišta podataka koja treba da reši ove probleme. Veći deo rada je posvećen procesu razvoja skladišta podataka sa sledećim fazama: planiranje, analiza i dizajn, implementacija i eksploatacija.*

Ključne reči: *Skladišta podataka, upravljanje projektom, razvoj sistema, poslovna inteligencija*