

# DATA WAREHOUSE – STRATEGIC ADVANTAGE

Sudesh M. Duggal, Northern Kentucky University, [duggals@nku.edu](mailto:duggals@nku.edu)  
Inna Pylyayeva, Northern Kentucky University, [pylyayevai@nku.edu](mailto:pylyayevai@nku.edu)

## ABSTRACT

*The data warehouse is an increasingly popular tool, which has been considered, developed, and implemented by a growing number of organizations in a vast array of industries. A summary of design and implementation considerations for any individual or group that might be performing analysis or early stage development of a data warehouse for their organization is presented. Some of the implementation challenges and problems associated with successful implementation are also presented, and its use for the strategic advantage for an organization is stressed.*

**Keywords:** Data warehouse, strategic advantage, design considerations, implementation considerations

## INTRODUCTION

A little over a decade ago, businesses used to work with data in megabyte and gigabytes, which is at present the size of a PC hard drive. According to Stephen, a number of corporations today must manage data in measures of terabytes and pentabytes (17). In this increasingly competitive business environment, there is an ever-ending need to acquire any data, anywhere, anytime, and as a result there is a need for an architecture that can include data from different sources and in different formats. Today's businesses require the ability to access and combine data from a variety of data stores, perform complex data analysis across these data stores, and create multidimensional views of data that represent the business analyst's perspective of the data (1). This can be achieved through corporate knowledge repository, also called data warehouse. Data warehouses are being developed by professionals in the field of Information Technology and through the use of commercially available tools and products. The development and maintenance of a data warehouse requires a significant commitment of resources, however if done right, that commitment can pay off many times its cost.

### What is a Data warehouse?

Data warehouse is the hottest tool in Information Technology area, which is defined by William H. Inmon (11) - the father of data warehousing, as "a subject oriented, integrated, time variant, nonvolatile collection of data in support of management's decision making process" (17). Data Warehouse organizes data around subject areas, where a subject area might include items such as customers, employees, financial management, products, or any components that is vital to measuring performance within a particular organization or entity. The data within the data warehouse is integrated, because of its ability to capture data from multiple Online Transaction Processing systems - transforming raw data from these diverse systems into a cohesive set of information, which enables an analyst to explore an idea, identify a trend, develop a course of action, serve a customer, or meet the demand for a product. Data Warehouse data should be accurate as of a certain date and time. A measure of time is typically included in each data

record in the database through an element, which is an implicit part of the key to data warehouse tables, and serves to give the warehouse time variant characteristics. Data in the data warehouse is nonvolatile because it is rarely changed and the changes to the data are normally limited to instances where a material issue of accuracy has been identified. Consequently, when an analyst extracts sales data for the month of January, the results will be the same regardless of when the information was extracted (17).

### **Data warehouse Objectives**

There can be a large list of objectives for a data warehouse, but the main objectives as given by Kimball are:

- To provide access to corporate or organizational data for decision support;
- To ensure that the data are consistent;
- To separate the data by appropriate means to measure performance and analyze the business;
- To provide a set of tools to query, analyze, present, and publish information (13)

## **DESIGN AND IMPLEMENTATION CONSIDERATIONS**

Data warehouse is not a product that you can purchase in the market, but is developed using design and implementation process, and consists of several components that will change and/or evolve over time. As such, it should be designed to accommodate future changes by focusing on flexibility from the start. Effective design and planning of the data warehouse can reduce the cost of implementation and the risk associated with changing existing processes and technologies (9).

### **Design Considerations**

Even though experts have different opinions about the design and implementation; however, there are some common strategies that should be used in the development of a data warehouse.

- Business needs
- Existing management software
- Evaluation of the database engine
- Existing tools for warehouse development
- An open operating environment
- Consideration of storage management issues (13).

Companies should develop unique approaches for particular data warehouse projects consistent with companies' business climates and rules. Since the users are critical to every aspect of design and development of the data warehouse, they should be involved in the process of implementing and designing the data warehouse. A proper balance of involvement from Information Technology and the user community is critical (1). Three elements are crucial to having a complete data warehouse that will truly optimize business analysis: back end (accessing and organizing data easily from disparate sources), preparing data for analysis (querying, searching, and governing the data), and front end (providing means for effective analysis of the information). In the back end, operational data for decision support can be accessed in two

primary ways: the data can be loaded into another database for information access, or the data can be accessed where they reside on the legacy or operational database. However, both methods are often necessary. In the second element, data must be transformed into a more useful resource with the following steps: translation, summarizing, packaging, distributing, and garbage collection of the data. Two key activities that make up the front end are user retrieval and proactive delivery (12).

### **Implementation Considerations**

There are two basic strategies of data warehouse implementation: “top-down” and “bottom-up” approaches. Although the “top-down” strategy was favored in early initial projects, and it is more elegant design approach, the high rates of failure led the majority of the following projects to the “bottom-up” approach (8). Anyway, there is a need for a central data warehouse that is populated with data automatically and provides effective retrieval and use of information (12). A number of factors that help bring about a successful data warehouse implementation are described below. It is important to know your audience and to understand the market, to which data “retailing” is directed. Who will be using the data warehouse, and what are the subject areas that should be used to measure performance? What are the skill sets of the individuals who will be using the warehouse? Also, there is a need to understand the nature of the information to be handled (2).

Operational cost reduction can be rarely a primary justification for data warehouse implementation. The best-cost objective is to achieve a permanent reduction in the full life cycle cost of delivering information. The objective of data warehouse implementation is to initiate a data acquisition and delivery process that offers lower marginal cost with each new use over time, but the best overall objective is to align the goals explicitly with a strategic business initiative. In other words, data warehouse should be linked to the strategic plan of the entire enterprise (9).

Before implementing a data warehouse, a set of questions should be answered, and if the answer “No” or “Not sure”, the need to implement a data warehouse is dubious. What are the strategic business objectives that the warehouse is supposed to achieve? What are the specific measurements that will be used to evaluate the ROI in meeting the company’s business objectives? Are the key users of the warehouse identified and committed to the success of the project? Is the company trying to build “do-all” warehouse on its first iteration? Does a project have support from executive management? Does the company have a clear understanding of the concept tools involved in data warehousing? Does the company have experience in building warehouses? Does the company have experienced consultants (16)?

### **Building a Data warehouse**

Building a data warehouse is a process of matching decision support needs to the realities of available data, and the success of a data warehouse depends critically on the accuracy and completeness of process and data specifications identified in the requirement phase (15). To build a successful data warehouse, during the implementation phase, there are factors to be considered: maintaining data quality in an ongoing joint user/builder responsibility; training

users one step at a time; implementing a user accessible automated directory to information stored in the warehouse; determining the plan to test the integrity of the data in the warehouse; and coordinating system roll-out with network administration personnel.

## **IMPLEMENTATION CHALLENGES AND PROBLEMS**

### **Implementation Challenges**

At the early stage of data warehouse development, many projects experienced difficulty because managers and designers failed to perceive the size, scope, and complexity of the challenge they make (1). The following are some of the challenges that may arise when implementing a data warehouse project:

- Time required to extract, clean, and load data;
- Problems with systems feeding the data warehouse;
- The need to store data not being captured by any existing system;
- The need to validate data not being validated by TPS;
- The lack of detail in TPS;
- Conflicting business rules;
- Overhead (aggregate tables and indexes to them) can eat up great amounts of disk space;
- Security issues, especially for Web-accessible warehouse;
- Access to information must be simple enough for staff to understand, but most data warehouse access software is complex. It usually requires professional assistance or extensive training (3);
- Data misinterpretation by end users, and the more cross-functional the data, the greater the likelihood for misinterpretation of data requirements (4);
- The challenge of consistency in data usage;
- The challenge of properly building SQL queries;
- The challenge of expanding access to the data warehouse to people who might be unfamiliar about the precise business interpretation of a specific data element;
- Coping with the demand of rapidly inventing new queries and reports as business requirements change;
- The challenge of understanding the warehouse's data properly from a business perspective (7);
- A number of political issues can arise around the data warehouse systems. They can be within the IS organization, between IS and users, and between users. These issues include control, responsibility, cooperation, access, and management of the data warehouse systems.

### **Implementation Problems**

One of the problems with data warehousing is that organizations are rushing to build it without regard to how that will impact existing systems architectures, or how it will be integrated with other applications. Critics also contend that warehouses ignore process and function, essentially sterilizing the data and removing the application context by isolating it in one or more relational database engine (5). According to Lauren, the absence or the lack of the following criteria is the

common causes of data warehouse failures: pre-launch clear objectives or metrics, insider presence on data warehouse projects team, user demand for sophisticated data analysis, and initial involvement of business managers. In addition, other problems can arise: many major systems projects underway simultaneously, the CEO sets budget and deadlines before project team is on the board, and source data availability unconfirmed at the outset. Another potential pitfall involves escalating user demand and unrealistic expectations (14). According to Haisten (9) one of the major causes of failures is that the database product was driving the project, not being driven by it. Others are the lack or the absence of close and effective link between business process analysis and data normalization, multi-national views available with the warehouse context, serious normalization, architecture for an integrated repository of browsable metadata, study of scale and capacity issues, and coordination with legacy application portfolio management.

According to Gupta, errors that data warehouse can contain involve data of four categories: incomplete, incorrect, incomprehensible, and inconsistent. Incomplete errors consist of missing records or missing fields. Incorrect data has wrong codes, calculations, aggregations, information entered into the system, pairing of codes, or duplicate records. Incomprehensibility errors include multiple fields within one field, unknown code, many-to-many relationships that allow multiple patterns, and spreadsheets or word-processing files. Inconsistency errors include inconsistent use of different codes or meaning of a code, overlapping codes, different codes with the same meaning as well as inconsistent names, addresses, business rules, aggregating, timing, and use of an attribute, nulls, spaces, etc. If the data warehouse has incomplete data or values that cannot be transformed properly, it is important for the data warehouse transformation process to use intelligent default values for the missing or corrupt data. It is also important to devise a mechanism for users of the data warehouse to be aware of these default values (6).

### **DATA WAREHOUSING STRATEGIC ADVANTAGE**

Organizations implement data warehouses to achieve strategic advantage and to stay competitive. However, very often, companies undertake data warehouses as a part of corporate strategy to shift from a product focus to a customer focus. A successful data warehouse can help identify new markets, focus resources on profitable customers, improve customer retention, and reduce inventory costs (15). The data warehouse also provides value to the knowledge worker by providing breadth (pulling in data from several sources), cleansing (reconciling differences in semantics, transaction data, currencies, etc.), and depth (consolidating data to higher levels while still supporting queries down to the detailed level and an overall design paralleling an end user's understanding of their business) (17). The future belongs to those who can see it and get there first. However, data warehouses are not enough. Choosing the right tools for Online Analytic Processing Systems provide the leverage needed to realize the benefits of data warehousing.

The business value of data warehouses include more cost effective decision making, better enterprise intelligence, enhanced customer service, enhanced asset/liability management, business process reengineering, and alignment with enterprise "right-sizing" objectives (17). Most organizations are constrained by their ability to integrate and understand available actionable business information. Decision Support Systems improve the decision-making process in an organization by providing timely, reliable information to decision makers. The best

decision support systems include the necessary tools to analyze and visualize the information, and to provide a means to synthesize the information as a result of the analysis. The very best are architected to participate in a smooth and seamless flow of information, providing connectedness between all steps in the decision making process.

Data warehouse application areas may include customer retention and attrition management, cross selling and up selling, campaign management, market, channel, and pricing analysis, and customer segmentation analysis. Data warehouse can help in strategic decision making especially when the work must be done quickly, and it is required with little advanced notice. Another benefit is that data can be aggregated and combined differently, which would provide a different business perspective. In other words, data warehousing can provide business various strategic benefits: immediate information delivery, data integration from across, and even outside the organization, future vision from historical trends, tools for looking at data in new ways, freedom from IS department resource limitations, analysis in various areas (sales, financial, human resource, and other areas analysis), reduced costs and increased or added revenues. Actually, the benefits from data warehouse implementation can be divided into two categories: those that come from cost reduction, and those that come from revenue enhancement. They can include elimination of legacy application maintenance, simplification of the operational environment and reduce of support costs, elimination of non-critical reports, and increasing availability of data ([www.computerwire.com](http://www.computerwire.com)).

Companies develop data warehouses to leverage existing businesses and generate new growth opportunities because timely and accurate information becomes an integral part of the decision-making process. Users can manage and access large volumes of information in one cohesive framework, and manager can distribute subject-oriented information on a variety of platforms, enabling faster decision-making. Moreover, data warehousing has widespread applicability to consolidate and speed up reporting and analysis. In short, because data in the data warehouse are integrated, have historical perspective, and are stored in summary and detail fields, the data warehouse architecture provides a solid basis for Decision Support System processing. It increases effectiveness and productivity because information inquires is satisfied without disrupting the performance of operational applications (12).

The ability to establish and understand the correlation between activities of different organizational groups within a company is one of the biggest advanced features of the data warehousing systems. The data warehouse systems can serve not only as effective platform to merge data from multiple current applications, but also from multiple versions of the same application. Designed properly, the data warehouse can allow for year-on-year analysis even though the base operational applications have changed. Data warehousing provides a catalyst for aligning what the business needs in an information strategy with what IT can actually produce (1).

## CONCLUSION

Data warehouse functions as a conduit between the users and corporate data resources creating a flexible staging area for decision support applications. It provides a specialized decision support database that manages the flow of information from exiting corporate database and external sources to end user to support strategic decision-making. To gain maximum value, an

organization that relies on information should consider the development of a warehouse in conjunction with other decision support tools, such as Online Analytical Processing Systems, data mining, and Executive Information Systems. The warehouse should serve as the foundation for the vital process of delivering key information to analysts and decision-makers, which are the key users of the information.

The decision to develop of a data warehouse should not be taken lightly. It is a resource intensive process and a considerable amount of time is required to thoroughly investigate business alternatives, determine system requirements, and develop performance measurements. According to Bischoff, data warehousing is a journey rather than a destination, and an understanding of the road map is vital to reach that elusive destination (1). Look toward the end of the journey, but seize the day in terms of the tremendous opportunities that today's technology affords for putting information at the fingertips of knowledge workers and customers anywhere any time. Never lose the renaissance perspective that applies so well to data warehousing. The journey is the reward (10).

## REFERENCES

1. Bischoff, Joyce. (1997). "Introduction to Data Warehousing", Data warehouse: Practical Advice From the Experts. New Jersey: Prentice-Hall, Inc.
2. Butler, M. (February 1998). How Valid is Data Warehousing?, [www.butlergroup.com](http://www.butlergroup.com).
3. Carlson, B. (March 24, 1997). Secrets of Success, [www.computerworld.com](http://www.computerworld.com).
4. Dyche, J. (Fall, 1998). New Rules for Business Discovery, [www.teradatareview.com](http://www.teradatareview.com).
5. Eckerson, Wayne W. (November & December 1995). Building the Legacy Systems of Tomorrow, Editorial. Open Information System.
6. Gupta, V. An Introduction to Data Warehousing. [www.system-services.com](http://www.system-services.com)
7. Hackarthorn. (August, 1997). Data Warehousing Credibility Crisis. Byte Magazine [www.byte.com](http://www.byte.com).
8. Hackney, D. (February, 1998). Data warehouse Delivery: Who Are You? Data Management Review Magazine.
9. Haisten, M. (February, 1995). Planning for a Data warehouse. InfoDB.
10. Hessinger, Paul, (1997). "A Renaissance for Information Technology", Data warehouse: Practical Advice From the experts, Prentice Hall.
11. Inmon, B. (May, 1999). Information Management: Charting the Course. Data Management Review Magazine. [www.dmreview.com](http://www.dmreview.com).
12. Kalakota, R., Whinston, A. (1996). Frontiers of Electronic Commerce. New York: Addison-Wesley Publishing Company, Inc.
13. Kimball, Ralph. (1996). The Data warehouse Toolkit. New York: John Wiley & Sons, Inc.
14. Lauren Gibbons Paul, (November 15, 1997). Anatomy of a Failure. CIO Enterprise Magazine. [www.cio.com](http://www.cio.com).
15. Mannino, Michael. (2001). Database Application, Development, and Design. New York: McGraw-Hill Higher Education.
16. Marco, D. (October, 2000). Everyman's Guide to Data Warehousing. [www.adtmag.com](http://www.adtmag.com).
17. Stephen, Amy. (1996). Introduction to Data Warehousing. University of Nebraska Computing Services Network. <http://www.nulook.uneb.edu>.