
BEST PRACTICES FOR SUCCESSFUL DEVELOPMENT OF DATA WAREHOUSES FOR SMALL BUSINESSES

*Ángel Ojeda-Castro, Universidad del Turabo, Gurabo, PR, ut_aojeda@suagm.edu
Mysore Ramaswamy, Southern University, Baton Rouge, LA, mysore@acm.org*

ABSTRACT

The movement to real-time is the latest development in business intelligence and data warehousing. In general terms, a data warehouse costs in excess of one million dollars and it is estimated that approximately 67% of the efforts fail. This is a significant factor when data warehouses are being developed for small businesses. Organizations are compelled to use their own data to produce reports that may indicate or better estimate tendencies in the market. This study examines six small businesses to determine best practices for creating data warehouses. Companies included in this analysis are in the fields of health care, manufacturing, education, and financial institutions. In this study, IT professionals and management were interviewed to determine which practices were highly valued in the development of data warehouses. In this paper, best practices for successful use of data warehouses are analyzed by examining six companies to determine the design, development and implementation.

Keywords: Critical Success Factors (CSF), Data Warehousing, Key Performance Indicators (KPI).

INTRODUCTION

In the current corporate environment, a company's success or failure may well be determined by the opportunity to access information. The capacity to identify and increase relations with costumers and organizational markets are critical factors in the success of a given firm [7, 8]. Knowledge based systems have proliferated and have invaded industries in the last few decades. Nevertheless, due to space limitation and high costs, many of the applications for data bases have been adapted to the needs for information of people who handle these operations on a short term basis. Thanks to devices with higher capacities, price abatement for these devices and an increasing growth of Internet technology, it is possible today to send historical data on-line to serve companies in their decision making process. The challenge now for these companies is to convert data files into an integrated source of knowledge. In the majority of the organizations, including health companies, education and government [12, 13] the demand for technology of databases has increased rapidly since the late 1970s.

During the past decades, organizational environments have undergone drastic changes. Today, the ever evolving global markets drive companies to seek a competitive edge over their competition. Thus, companies are constantly seeking knowledge and innovations to give them an advantage [14, 18]. One aspect of gaining a competitive edge over the competition is to shorten decision making timelines when responding to changes in the organization [17, 22]. Another aspect that organizational decision makers have found to give them a competitive edge is maintain and manage their own data [10, 11]. Today, organizations find it more useful to manage their own data resources in order to generate reports that predict marketplace tendencies [2, 4]. The use of information technology and data management has become increasing important practice for collecting data and analyzing information to make informed business decisions.

Previously, the data were primarily employed to understand what had already happened and to predict what would happen in the future. Its use for influencing real-time decisions and current operations was limited. With real-time data, current decisions and critical decision processes such as customer relationship and supply chain applications can be significantly improved. Enterprises in the telecommunications, retailing, and financial services industries have built data warehouses to store vast amounts of customer and sales related data. These organizations are leaders in terms of the size of their data warehouses and how the warehouses are used. Data warehouses for small businesses have to be developed with specific goals in order to contain the cost of development and maintenance. The main objective of the use of a data warehouse is to enhance the executive's intelligence, decision making skills, and the company's performance [9, 15].

With new contemporary companies, in the whole architecture of decision support, data warehousing is gaining popularity. More often, companies seek to warehouse and control their own data in order to utilize it for multidimensional analysis and company decision making [14, 16, 20]. Data warehousing has become a widespread practice in business intelligence for companies who want a competitive advantage. In fact, most businesses understand that it is paramount in strategic planning and initiatives and must absolutely be included in the infrastructure. Data warehousing has become an irreplaceable and essential tool for critical decision making and planning [3, 21]. Using a sophisticated on-line analytical procedure (OLAP) and tools of data mining, some companies are able to exploit their knowledge of data warehousing to significantly increase sales [1, 5], reduce costs [19, 21] and offer new and better products or services [5, 6]. When the data warehouse is used properly, it is organized by subjects and allows for instantaneously integrated and precise that is available whenever desired. In a data warehouse, data can be maintained free from changes [3, 12]. It is well understood that data warehousing includes data, the tools, the procedures, personnel training and other resources.

In this paper, best practices for successful use of data warehouses were studied by examining six companies to determine the design, development and implementation. The study provides an overview of six different companies from the industries of health, education, and manufacturing. It determined which relevant aspects were common for the successful development and implementation of data warehouses. This study also analyzed the unique requirements for small businesses' successfully use of data warehouses.

BACKGROUND OF BUSSINESSES

This section provides an overview of the six enterprises that were studied. To maintain confidentiality, the companies will be identified as Company A, Company B, Company C, Company D, Company E and Company F.

Company A is a private for-profit organization that is a wholesale supplier of medication. Company A has participated in the health market for more than fifty years. This company has a large, intact infrastructure and sufficient space for warehousing information. The company has report continual growth and has over fifty IT specialists to design, develop, support and produce information systems and analyze data.

Company B is a prescription health plan that includes over 4,000 prescribed drugs, which it distributes to more than 900 pharmacies. Company B is in charge of transaction processes, medicine discounts, claims and intranet service for its 900 pharmacies.

Company C is a health care provider that caters to health professionals and patients who have complex treatments, such as: hemophilia, immune diseases, renal diseases, cancer, traumas and other conditions. This company is also involved with medical devices, pharmaceutical products and biotechnology. It employs approximately 4,000 people and aims to improve the health of its patients.

Company D is pharmaceutical company that manufactures drugs to treat diabetes, mental disorders, erectile dysfunction and colorectal cancer and has been operating for approximate 50 years. There are 1,100 employees and 220 of them work directly in the marketing office.

Company E is government health care agency which employees approximately 7,000 employee. Company E has an IT Department with seven divisions. There are 37 people in the IT Department. Additionally, there is Statistics Department, which supports the IT Department. The Statistics Department generates reports, and documents. They are used to generate health and social statistics of a population. They also establish tendencies for managerial decision making. Their Data Warehouse was developed by a outside company. Company E's technology permits citizen or institution to access their information via the Internet.

Company F is a public entity that has the responsibility for overseeing educational institutions. Company F's Data Warehouse was developed by a contracted company. They utilize their Data Warehouse to This organization hired the services of a private company to develop their Data Warehouse. Their intention was to produce federal and state statistics and reports to be shared with other governmental agencies.

METHODOLOGY

The purpose of this study was to identify the best practices necessary for the successful design, development and implementation data warehouses for businesses. This research analyzed the best practices used by companies in the creation of a cost effective data warehouse that was developed in a short span of time.

This research utilized both quantitative and qualitative instruments to ascertain the best practices used by six companies in their design, development and implementation of their data warehouses. The second component of this study included case studies of each of the businesses. Qualitative data that would relate to the factors in best practices phase in the design, development and implementation of a data warehouse were analyzed.

The six companies studies, while different, related to either the health or education industries. Each of the six companies represented participated in interviews that included more than one representatives of the organization. The personnel interviewed in each company were from the areas of IT professionals, general management, top management, and the project managers involved in the data warehouse implementation.

During the interviews, the representatives from each company were asked to respond open-ended and close-ended questions related to the company's technological infrastructure and the decision making process they undertook integrate a Data Warehouse. They were also asked to identify benefits and best practices of the design, development and implementation of the data warehouse. In addition to the interviews, the researchers observed behavioral activities of the participants and a data collection plan and a sampling strategy were used [6].

DATA ANALYSIS

In this section, the results and recommendations of the study area analyzed and discussed related to the best practices for the design, development and implementation of data warehouses in businesses. Each company's successful implementation of their data warehouse was contingent upon several factors uncovered in this study. The data analysis revealed that there were seven dominant factors that yielded successful data warehouse implementation. Essentially, these seven factors were categorized in two different groups: organizational practices and technological practices.

The organizational practices observed were: Scope, Management, Infrastructure, and Data Warehouse Specialists. The technological practices observed were: Data Quality, Methodology utilized to develop the Data Warehouse, and Data Warehouse Development Software.

Scope is defined as all components which include all of the fundamental information needed to delineate the all aspects of the company's operation. Management is defined the backing that upper management allows for the development and implementation of its data warehouse project. Infrastructure is defined as the having quality computers at the company's disposal and additional technological equipment, such as servers and telecommunication equipment to make the implement a data warehouse possible. A Data Warehouse Specialist is defined as the personnel with the expertise and the prior successful experience necessary to implement a data warehouse. Data Quality is defined as the levels of extraction and granularity, along with trustworthiness, accuracy and completeness of the the names and meanings of the fields. Methodology used to develop the Data Warehouse is defined the prototypes and models that can have been successfully replicated previously and can be explained and discussed with the business leadership and top management. Data Warehouse Development Software is defined as the technological tools and software used to develop and implement a data warehouse taking into consideration efficiency and cost effectively.

The analysis of the seven dominant factors for the successful design, development and implementation of the data warehouse are presented below:

Scope: This study determined that in companies A (Wholesale supplier of medication), B (Prescription health plan provider) and C (Healthcare provider) incorporated almost all of the fundamental information concerning the operation of the company. The top management of Companies B (Prescription health plan provider), E (Government healthcare agency) and F (Government education department) were not precise when communicating the needs of the business or need and use of the data warehouse.

Management: This study concluded that companies B (Prescription health plan provider), C (Healthcare provider) and D (Pharmaceutical Company) top management of these companies were fully committed to designing, developing and implementing their data warehouses. However, companies A (Wholesale supplier of medication), E (Government healthcare agency) and F (Government education department) did not receive full commitment from their top management in this process.

Infrastructure: The infrastructure of Companies B (Prescription health plan provider) and C (Healthcare provider) was completely intact and possessed adequate computers, servers and telecommunication equipment. However, Companies A (Wholesale supplier of medication), D (Pharmaceutical Company) and E (Government healthcare agency) had a most of the necessary infrastructure, but lacked some infrastructure capacity. This study concluded that Company F (Government education department) had no knowledge was not even aware of the infrastructure that it had. It could not determine if the infrastructure was appropriate.

Specialist: Company A (Wholesale supplier of medication) was fully aware that the designated Data Warehouse Specialist possessed the expertise and knowledge before being appointed to the project. Companies B (Prescription health plan provider), C (Healthcare provider) and D (Pharmaceutical Company) was cognizant of the fact that their Data Warehouse Specialist had some experience, but lacked expertise in some areas. The study revealed that Company F (Government education department) was unaware of the exact experience of their Data Warehouse Specialist and Company expressed that their Data Warehouse Specialist had little experience.

Data Quality: Upon examination of the each company's data quality, it was discovered that Companies E (Government healthcare agency) and F (Government education department) had encountered data quality problems because there were errors in the data base source. Also, companies B (Prescription health plan provider) and C (Healthcare provider) encountered difficulty in interpreting data from the data base source. The only companies that did not encounter a problem with their data quality were Companies A (Wholesale supplier of medication) and D (Pharmaceutical Company). Companies A (Wholesale supplier of medication) and D (Pharmaceutical Company) were proactive and anticipated that they had to have a validation system to ensure the data quality.

Methodology: The study concluded that Companies B (Prescription health plan provider) and E (Government healthcare agency) did not have a model or prototype available before the project. Additionally, they were not aware of the frequency in the evolutionary developments of the versions for the data warehouse during its development. However, Companies A (Wholesale supplier of medication), B (Prescription health plan provider), C (Healthcare provider), and D (Pharmaceutical Company) had models and prototypes available and regularly discussed and utilized them in their design and implementation of their data warehouse.

Data Warehouse Development Software: Companies E (Government healthcare agency) and F (Government education department) contracted external resources to develop their Data Warehouse. Companies A (Wholesale supplier of medication), B (Prescription health plan provider), C (Healthcare provider), and D (Pharmaceutical Company) were aware of the cost-effectiveness, advantages and disadvantages, of comparable software.

Benefits: All companies in this study affirmed that there were some concrete benefits of a Data Warehouse as it relates cost reduction, marketing strategies, generating sales, efficiency and generating reports, decision making based on multidimensional information. Additionally, all companies mentioned that the Data Warehouse allowed them to have other services they did not have before, it saved a time, and there was a definite return of investment.

The average time the companies required to develop and implement a data warehouse was eight months. The factors that help the most in reducing the time necessary to develop and implement a data warehouse are: scope, specialists, data quality and the technological tool. We found that the average cost of a data warehouse was \$381,000 including the technological infrastructure such as computers and telecommunication equipment. Factors such as the specialists and the technological tools are the ones that help the most in optimizing the utility of a data warehouse.

From the perspective of small businesses, the cost of developing data warehouses is the most important concern. There are four major cost components: Consulting Costs, Storage Costs, ETL (Extract, Transform, and Load) Costs, and Database Management System (DBMS) Costs. As consulting costs tend to be high, small businesses should try

to develop systems without external consultants. To reduce ETL costs, the scope of the system has to be clearly defined making sure that the data warehouse adequately answers the user queries.

CONCLUSIONS

This study helps to identify elements of best practices in implementing effective data warehouses for a variety of industries. As exhibited by the various industries in the study, the scope and management yielded the highest best practices for data warehouse implementation. The data quality and methodology, which should be increased gradually, followed as lesser importance to best practices. The infrastructure and the technological tools fared as the least important in best practice. However, incorporating the data warehouse specialist as a best practice depends on whether or not the specialist was internal or outsourced. Additionally, companies must make data warehouses an integral part of its strategic planning if implementation and the return of investment (ROI) is to be successful.



FIGURE 1.

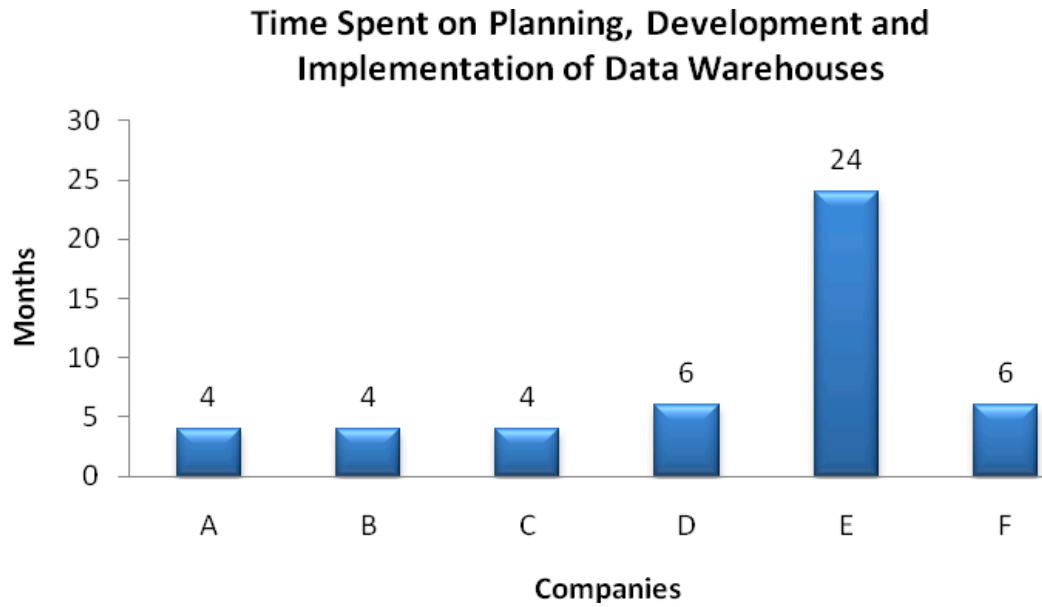


FIGURE 2.

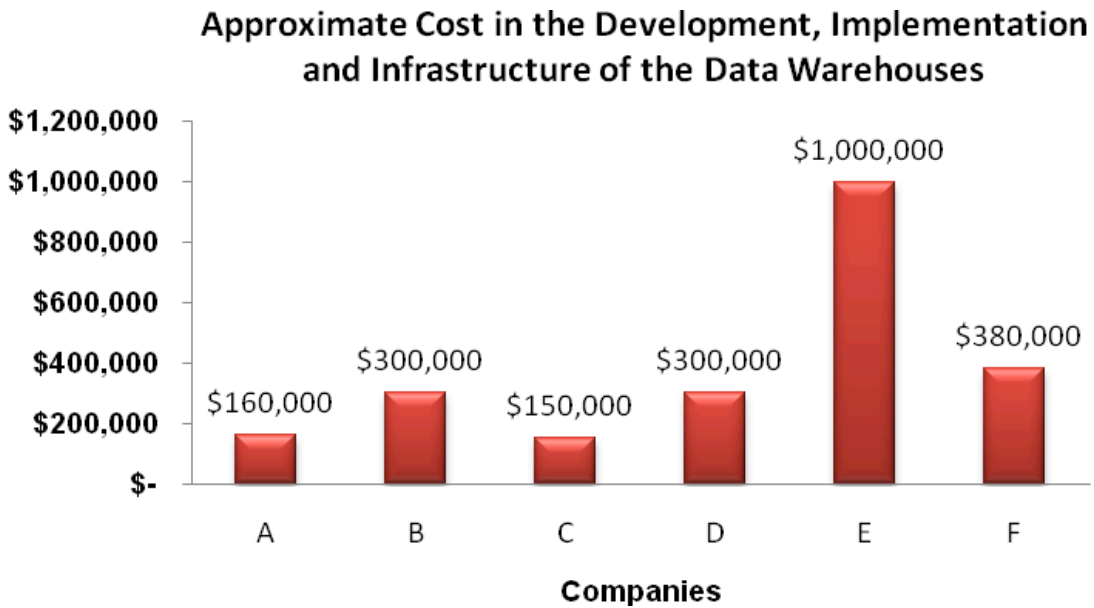


FIGURE 3.

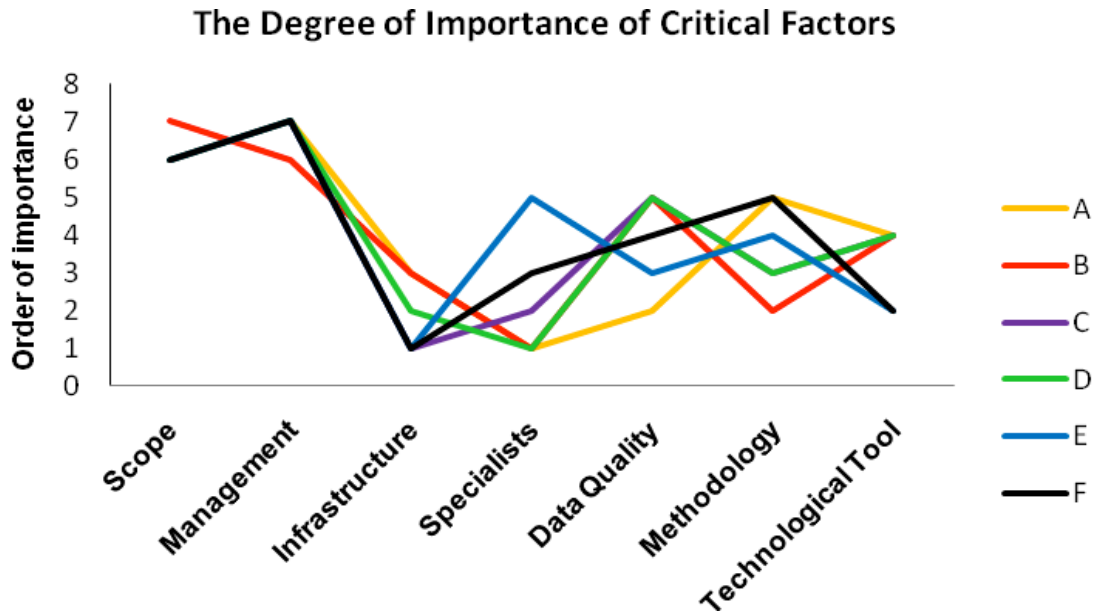


FIGURE 4.

Companies tended to confide in outsourced personnel as having the requisite skills and knowledge to successfully manage the data warehouses. Internal personnel presented more concerns to best practices as the companies have to ensure that they were properly trained and have invested in their professional development. This is exhibited in Company E, which outsourced the implementation of its data warehouse. Despite the fact that Company E spent more time, effort and funds in its implementation, it yielded the lowest level of implementation success. This research determined that there was no intervention by the local company in the development process of the Data Warehouse. However, the company did participate in the design phase. Additionally, there were approximately twenty changes during the implementation of the Data Warehouse. Furthermore, Company E was aware that the data warehouse specialist that they contracted had little experience on in the field. As a result, and implement of the data warehouse took approximately two years.

As a result of Company E's lack of implementation success, it is recommended that future research compares different industries and how each affect the cost of implementation and usefulness of the Data Warehouses. Additionally, it is recommended to extend the research focus to systems which answer real-time queries and propose a set of recommendations which will aide companies in their decision making regarding outsourcing its Data Warehouses.

We concluded that the order in which the factors determine the success of a data warehouse, without considering external help is:

1. Management
2. Scope
3. Data Quality
4. Methodology
5. Technological Tool
6. Specialist
7. Infrastructure

If an external company is to be hired to develop and implement a data warehouse, the order of the factors to determine success is as follows:

1. Management
2. Scope
3. Specialist
4. Methodology
5. Data Quality
6. Technological Tool
7. Infrastructure

REFERENCES

1. Arnott, D. and Pervan, G. (2005). A Critical Analysis of Decision Support Systems Research. *Journal of Information Technology*, 20(2), pp. 67 – 85.
2. Baker, S. & Baker, K. (1999). The Best Little Warehouse in Business. *Journal of Business Strategy*, 20(2), 32-37.
3. Chen, L., Soliman, K., Mao, E., & Frolick, M. (2000). Measuring User Satisfaction with Data Warehouses: An Exploratory Study. *Information & Management*, 37, 103-110.
4. Conner, D. (2003). Data Warehouse Failures Commonplace. *Network World*, (3), 24.
5. Cooper, B.L., Watson, H.J., Wixom, B.H., and Goodhue, D.L. (2000). Data Warehousing Supports Corporate Strategy at First American Corporation. *MIS Quarterly*, 24(4), 547-567.
6. Cooper, D. R., & Schindler, P. S. (2003). Business Research Methods. In M. Graves (Ed.), *Business Research Methods* (8th ed., pp. 145-189). Boston: McGraw-Hill Irwin.
7. DeLone, W. & McLean, E. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1), 60-95.
8. DeLone, W. & McLean, E. (2003). The DeLone and McLean Model of Information Systems Success: A Ten Year Update. *Journal of Management Information Systems*, 19(4), 9-30.
9. Devlin, B. (1997). *Data Warehouse: From Architecture to Implementation*. Reading, MA: Addison Wesley Longman.
10. Eckerson, W. (2003). Evolution of Data Warehousing: The Trend toward Analytical Applications. *Journal of Data Warehousing*, 25(1), pp. 1-8.
11. Gorla, N. (2003). Features to Consider in a Data Warehousing System. *Communications of the ACM*, 46(11), 111.
12. Hwang, M.I., & Thorn, R. (1999). The Effect of User Engagement on System Success: a Meta-Analytical Integration of Research Findings. *Information & Management*, 35(4), 229-236.
13. Hwang, M.I., & Cappel, J.J. (2002). Data Warehouse Development and Management: Practices of Some Large Companies. *Journal of Computer Information Systems*, 43(1), 3-6.
14. Inmon, W. H. (2002). *Building the Data Warehouse* (3rd Ed.). New York: Wiley.
15. Joshi, K. & Curtis, M. (1999). Issues in Building a Successful Data Warehouse. *Information Strategy*, 15(2), 28-35.
16. Poe, V. (1996). *Building a Data Warehouse for Decision Support*. Upper Saddle River, NJ: Prentice-Hall.
17. Shin, B. (2003). An Exploratory Investigation of System Success Factors in Data Warehousing. *Journal of the Association for Information Systems*, 4, 141-170.
18. Vatanasombut B., & Gray, P. (1999). Factors for Success in Data Warehousing: What the Literature Tells Us. *Journal of Data Warehousing*, 4(3), 25-33.
19. Watson, H., Annino, D., Wixom, B., Avery, K. & Rutherford, M. (2001). Current Practices in Data Warehousing. *Information Systems Management*, 18(1), 47-55.
20. Watson, H.J., Fuller, C., & Ariyachandra, T. (2004). Data Warehouse Governance: Best Practices at Blue Cross and Blue Shield of North Carolina. *Decision Support Systems*, 38, 435-450.
21. Watson, H., & Haley, B. (1998). Managerial Considerations. *Communications of the ACM*, 41(9), 32-37.
22. Wixom, B. H., & Watson, H. J. (2001). An Empirical Investigation of the Success Factors for Data Warehousing. *MIS Quarterly*, 25(1), 17-4