

CRITICAL FACTORS FOR SUCCESSFUL IMPLEMENTATION OF DATA WAREHOUSES

Ángel Ojeda-Castro, Universidad del Turabo, Gurabo, PR, ut_aojeda@suagm.edu
Mysore Ramaswamy, Southern University, Baton Rouge, LA, mysore@acm.org
Ángel Rivera-Collazo, Universidad del Turabo, Gurabo, PR, arivera@suagm.edu
Ahmad Jumah, Inter American University, San Juan, PR, ahjumah@hotmail.com

ABSTRACT

The literature on business intelligence mentions several critical factors that play an important role in the successful implementation of data warehouses. These factors help determine the proper techniques that guarantee their optimal usage. Based on empirical information, a data warehouse costs in excess of one million dollars and it is estimated that approximately 67% of the efforts at implementation result in failures. The dimensions of these factors that influence the success of data warehouses are still not clear. Reliance on data warehouses is becoming important as organizations are compelled to use their own data to produce reports that may indicate or better estimate tendencies in the market. In this research, we study six different business enterprises in Puerto Rico to develop an effective framework to implement data warehouses. Information technology analysts and managers, data base specialists, and application developers were interviewed on the critical factors in the implementation of data warehouses. The areas of industry represented by those interviewed were: education, finance, manufacturing and health. The results of these interviews and the case studies reveal that factors such as quality of data, range of project, methodology, technological tools and internal team work increase the probability that data warehouse projects are completed on time, within budget and with efficient functionality.

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

INTRODUCTION

The business environment within organizations has changed drastically. Companies compete in a constantly changing global market which necessitates acquiring the knowledge to survive and gain a competitive edge [14, 18]. To gain a competitive edge, organizations need to accelerate decision making timelines in their processes in order to respond rapidly to change [17, 22]. As administrators struggle to create and maintain this competitive edge, one of the most promising resources available is their own data [10, 11]. Organizations use their own data to produce reports that may indicate or better estimate tendencies in the market [2, 4]. Since 1990, techniques of saving data in data warehouses have become a critical field in the field of information technology.

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 customers and organizational markets are critical factors in the success of a given firm [7, 8]. Data based systems have proliferated and have invaded industries in the last few decades. Due to space limitation and high costs, many of the applications for data bases have been modified to suit 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 data bases has increased rapidly since the late 1970s.

In contemporary organizations, data warehousing is gaining popularity in the whole architecture of decision support. Organizations want to benefit from the power to be able to perform multidimensional analysis of historical warehoused data to help companies in their administrative decision making [14, 16, 20]. Data warehousing has evolved to be recognized as a basic requisite in most of the organizations that use business intelligence. It is being

recognized as a must within the critical infrastructure for strategic initiatives, as the relation of the client becomes an important part of the supply chain. It has also become the fundamental basis to support the application of advanced decisions [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]. The data warehouse is organized by subjects, and data is always integrated. Data do not suffer any change in the data warehouse [3, 12]. It is well understood that data warehousing includes data, the tools, the procedures, personnel training and other resources. Data warehousing has as its main objective, the improvement of the executive's intelligence, decision making skills and the performance of the company [9, 15].

In this research, we analyze factors that are critical for successful implementation of data warehouses for enterprises in Puerto Rico by studying six businesses. The rest of this paper is organized as follows. First we provide an overview of the business enterprises we have studied for this research. Research methodology and results are analyzed in the subsequent sections. Concluding remarks form the last section.

BACKGROUND OF ENTERPRISES

In this section, we give a brief overview of the six business enterprises in Puerto Rico we have studied for this research. We are identifying those organizations as Companies A through F due to privacy considerations.

Company A is a private for-profit organization. It is affiliated to a company which is in the healthcare business for more than fifty years. It is also a wholesale supplier of medicines that started its operations in the 90s. It has a highly developed infrastructure for information processing and ample space for warehousing. It is one of the fastest growing companies in Puerto Rico. It has approximately 228,000 square feet of floor space in its offices and warehouses. It has over fifty specialists dedicated to design, develop and maintain information processing systems. The data processing and communication system is valued at more than 4 million dollars.

Company B deals in executing pharmacy plans which include medicines prescribed through Medicare Section D. This company is hired by Medicare and authorized by the Office of the Insurance Commissioner of Puerto Rico to provide benefits. It covers close to 4,000 prescribed drugs in a network of 900 drug stores in Puerto Rico and thousands in the continental USA. It also covers medicine discounts, and claims in a user-friendly network.

Company C is dedicated to provide health care. It provides health professionals and covers their patients' complex treatments, such as hemophilia, immune diseases, renal diseases, cancer, traumas and other conditions. This company relies on its experience in medical devices, pharmaceutical products and biotechnology to make a real difference in the health of the patients. It started its operations in Puerto Rico by the end of the 50s. It now has three units located in Jayuya, Aibonito and Guayama and hires approximately 4,000 employees. It also has sales offices in Guaynabo and a logistics center in Cataño.

Company D specializes in providing health care. It manufactures drugs to treat diabetes, mental disorders, erectile dysfunction and colorectal cancer. It started operations by the mid 60s. It now has three units located in Carolina and Mayagüez and hires close to 1,100 employees. It also has marketing offices in San Juan that operate with 220 employees.

Company E is a public organization that oversees health care needs of the citizens of Puerto Rico. It has approximately 7,000 employees who work in different functional areas. The information technology department has seven units and employs thirty seven people. The statistics department, with the support of the information technology department, prepares the necessary reports and liaises with different groups so that the managers may take appropriate operational decisions. This organization had hired the services of a private company to develop their data warehouse. One of the objectives was to keep the institutions and citizens informed of health related matters. Today any citizen or institution may access this information through the Internet. This also helps in generating statistics relevant to public health.

Company F is a public organization that has the responsibility to grant approval to educational institutions in Puerto Rico. This organization had hired the services of a private company to develop their data warehouse. Their intention was to generate reports and documentation to federal and state agencies in the country.

RESEARCH METHODOLOGY

The purpose of this study is to identify various factors that are critical in successfully implementing data warehouses in businesses located in Puerto Rico. These factors are used in developing an implementation methodology which allows the creation of a cost effective data warehouse with a shorter development time.

Quantitative and qualitative criteria were used to determine if organizations would benefit from the implementation of a data warehouse. In the first phase of this research, 36 questionnaires were distributed to a diverse group that included developers, architects, or data warehouse project administrators. The opinions of the target subjects were focused on organizational factors, techniques, efficiency and benefits of data warehousing. In the second part of this research, six cases were studied. In each case, qualitative data that would relate to the factors mentioned in the first phase were analyzed. The characteristics of the companies studied were sectors related to education and health.

Personnel surveyed, thirty six in all, held positions as developers, architects, project managers or managers of data warehouses. In order to participate in this survey, the requirement was to have participated or helped in the development or implementation of a data warehouse project. Companies that participated in interviews were the health and education sectors. The interviews (six) were conducted with one or more representatives of the organization. The interview also indicated whether each representative was a user, administrator or project manager of a data warehouse.

The interviews were conducted to determine if the establishment of a data warehouse had provided any benefits to the companies. The questionnaire for participants was designed to elicit their overall opinion about data warehouses and determine if the project was successful or not. Contact by email was made to manufacturers and business partners engaged in the development of data warehouse to schedule the day and place of each meeting. Then the participants were invited to answer the questionnaire. For case studies, participants with data warehouses were contacted.

Participants were asked to respond to the questions regarding the technical infrastructure and business processes to determine if there was a need to implement a data warehouse and measure the critical success factors involved. Other tools, such as interviews, were used to develop a data gathering plan and a sampling strategy while behavioral activities were also under observation [6]. The procedure used was a questionnaire and administered to thirty six specialists who either managed or developed a data warehouse, and in the second stage case studies were conducted that involved six companies. Participants were specialists, users and project managers of data warehouses.

SUMMARY OF RESULTS

As indicated in Appendices I through IV, the success of data warehouses in the six business enterprises we studied, vary. The success in the implementation of a data warehouse was influenced by the result of several factors which are elicited by the participants of the surveys and by those who responded to questionnaires in this study. The results of this research show that there were seven factors that heavily impacted in the success of implementation. These seven factors were classified in two groups: organizational factors and technological factors.

The organizational factors identified were: Scope, Management, Infrastructure, and Availability of Specialists. While the technological factors are identified as: Data Quality, Methodology and Technical Tools.

Scope is defined as that factor which encompasses all the fundamental information needed to describe the operation of all aspects of the company.

Management is defined as all the support the upper management offers to the development and implementation of a data warehouse project.

Infrastructure is defined as the convenience of having quality computers, servers and telecommunication equipments to implement a data warehouse.

Specialist is defined as one who has the previous experience necessary to implement a data warehouse.

Data Quality is defined as that factor which ensures the accuracy of the algorithms used, the levels of extraction and granularity. It also assures that the names and meanings of the fields are trustworthy and complete.

Methodology is defined as the frequency with which the prototypes and models are presented before the data warehouse is completed.

Technological Tools refers to the appropriate selection of technological tools to develop and implement a data warehouse. It should result in lower cost and higher efficiency.

Our analysis on the impact of the above factors on the different businesses are presented below, each factor is studied separately.

Scope: We found that in company A, B and C, they included information that covered almost all aspects concerning the operation of the company. Companies B, E and F showed imprecision regarding the definition and identification of information that should be included in the data warehouse project.

Management: It was observed that companies B, C and D received full support from the upper management to develop and implement a data warehouse, while companies A, E and F received less support.

Infrastructure: Companies B and C had complete infrastructure capacity in the form of computers, servers and telecommunication equipments. Companies A, D and E had almost all the infrastructure capacity. Company F was not sure whether they had the proper infrastructure capacity.

Specialist: Company A understood that the specialists in charge of the development and implementation of the data warehouse had the knowledge and experience before the onset of the project. Companies B, C and D recognized the specialist had almost all the knowledge and experience. Company F argued they were not sure about the knowledge and experience of the specialists. Company E indicated that their specialists had no experience and no knowledge regarding data warehouses.

Data Quality: Companies E and F had problems with the data because the output was incomplete and with errors when retrieved from the data base source. Companies B and C had difficulties to interpret the data from the data base source to represent those in the Data Warehouse. Companies A and D had no problems with the quality of the data because they had a validation system for the data.

Methodology: Companies B and E did not see the prototypes of the projects and there were no iterations in the evolutionary developments of the versions for the data warehouse during the development process. In companies A, B, C, and D the prototypes had been inspected. There was an exchange of ideas on these observations for the evolution models and analysis.

Technological Tool: Companies E and F did not consider informing themselves about the use of one or another technological tool because they hired the services of an external agent to develop and implement their data warehouse. Companies A, B, C, and D were informed about the cost-effective advantages and disadvantages of using one tool over another by comparison.

Benefits: All companies recognized some type of benefit in terms of cost reduction, reach-out to other markets, increase in sales, time saving in amount and preparation of reports, decision making based on multidimensional information obtained, having other services they did not have before and better recovery of their investment.

The specialists that developed and implemented the data warehouses during this study agreed that they interpreted a reduction in costs, reach-out to other markets, increase in sales, time saving in the amount and preparation of reports and more effective decision making bases on multidimensional information obtained as indicators of successful implementation.

The average time the companies required to develop and implement a data warehouse was eight months. But, when we calculated the median we found a value of six months. When we determined the mode, we found a value of four 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.

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. Specialists
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. Specialists
4. Methodology
5. Data Quality
6. Technological Tool
7. Infrastructure

CONCLUSION

This study helps in identifying factors that are critical to the successful implementation of data warehouses used in a variety of businesses. While comparing the success, one of the striking factors that emerge pertains to Company E. Even though Company E spent more money and time for implementation, the level of success of implementation of Data Warehouse was the lowest. It is worthwhile to look at the following facts regarding Company E. In company E, we found four data base sources. Each data base responds to different departments. This organization has to generate more than 40 reports per month. The statistics department is in charge of producing the required reports. This department has seven employees.

The upper management had the opportunity of seeing the concept of intelligent business at a conference in the USA. They also saw the application of the tools for data warehouse. The upper management decided to invest one million dollars in the development and implementation of a data warehouse.

The statistics and planning department identified ten necessities and the possible publication in an Internet site of the multidimensional information generated from the data gathered from the main data sources. During this period there was certain resistance to the project and a committee was formed to handle the new way of producing reports and graphics. The upper management decided what was to be published in the Internet site. There were certain problems identifying the initial necessities of the project. An external company was hired with the responsibility to design, develop and implement the data warehouse, taking into account the necessities identified by the organization.

During the development process, there were problems with the data gathered in the data base sources and the interpretation of the data to formulate the equivalent of a data warehouse. The hired company took close to one year working in the quality of the data and used the technological tool “SQL Server 2000” and “Analysis Server” to extract, load and transform the data and generate the reports.

There was no intervention by the local company in the development process. Yet, there was some involvement during the design phase. There were almost 20 changes during the evolution process of the data warehouse. The company understood that the specialists in data warehouse did not have much experience. The amount of time it took to develop and implement the data warehouse was approximately two years. Despite all these difficulties in implementation, the preparation of reports through the web site served to inspire other public agencies in Puerto Rico to do the same.

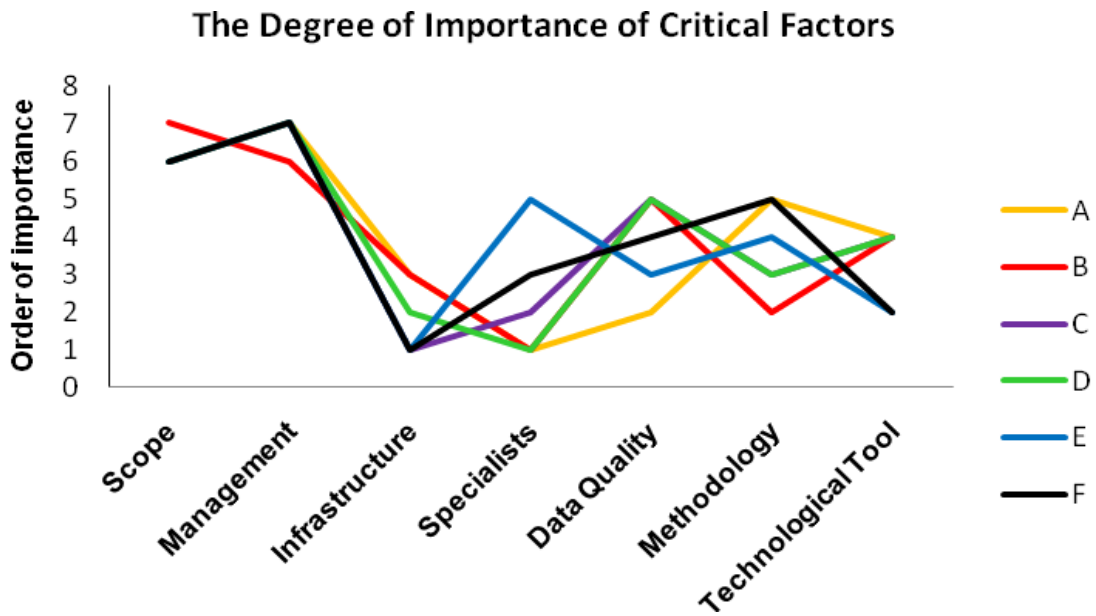
Future research includes comparison among various enterprises and how those affect the cost of implementation and usefulness of the Data Warehouses. We also plan to extend the research focus to businesses beyond Puerto Rico and propose a set of recommendations that will guide companies in their decision making regarding outsourcing of this vital component.

REFERENCES

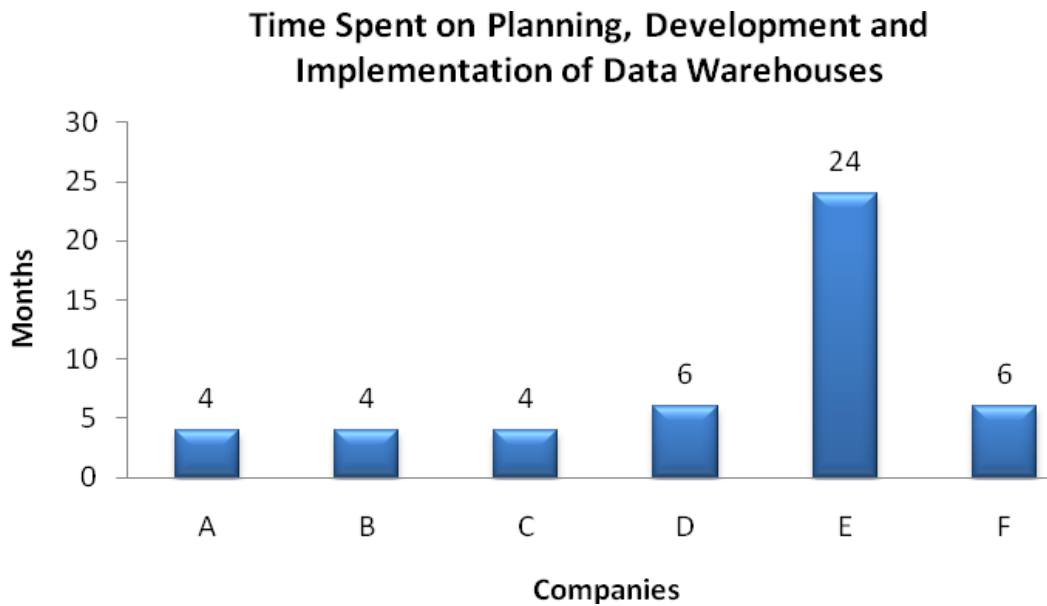
1. Arnott, D. and Pervan, G. (2005). A Critical Analysis of Decision Support Systems Research. *Journal of Information Technology*, 20(2), pp. 67 – 85.
2. Baker, S. & Baker, K. (1999). The Best Little Warehouse in Business. *Journal of Business Strategy*, 20(2), 32-37.
3. Chen, L., Soliman, K., Mao, E., & FrolickM. (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

Appendix I.

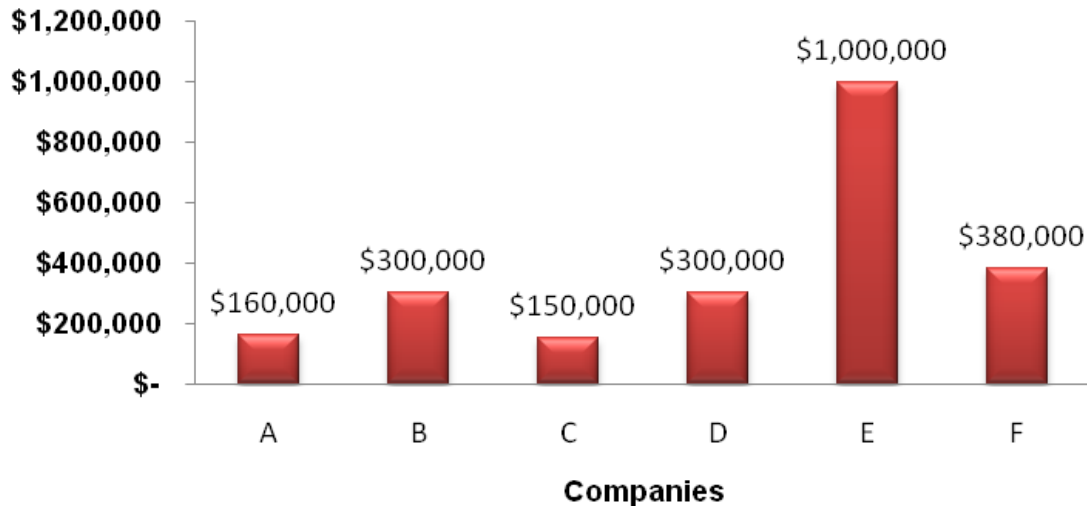


Appendix II



Appendix III

Approximate Cost in the Development, Implementation and Infrastructure of the Data Warehouses



Appendix IV

Level of Success of Each Company

