Fall 2005

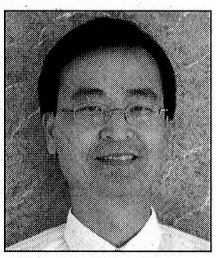
A Survey of Data Warehousing Success Issues

Hongjiang Xu
Butler University, hxu@butler.edu

Mark I. Hwang

A Survey of Data Warehousing Success Issues

Mark I. Hwang and Hongjiang Xu



Mark I. Hwang is in the business information systems department at Central Michigan University.
mark.hwang@cmich.edu

Hongjiang Xu is in the business information systems department at Central Michigan University.
hongjiang.xu@cmich.edu

Data warehousing is an important area of practice and research, yet few studies have assessed its practices in general and critical success factors in particular. Although many guidelines for implementation exist, most are derived from anecdotal evidence. A survey of data warehousing professionals was conducted to gain insight into data warehousing success issues.

The results reveal that data warehousing success is a multi-faceted construct and that improved productivity is the most valued measure for success. The results also put clearly defined business needs/benefits and source data quality at the top of the list of critical success factors. In addition, different success factors were found to be significant for different success measures.

Introduction

Since the early 1990s, the data warehouse has become the foundation of advanced decision-support applications. Using sophisticated business intelligence tools, some corporations are able to use insights gained from their data warehouse to significantly increase sales, reduce costs, and offer new and better products or services. The huge payoff from a well-run data warehouse is routinely cited in the literature. At the same time, stories abound about expensive and failed implementations.

Although many implementation guidelines exist, most are derived from anecdotal evidence, making unclear the precise nature of the critical success factors and their impact. (Mukherjee and D'Souza, 2003) This deficiency takes on new urgency as companies spend more to build ever-bigger data warehouses in pursuit of greater granularity and real-time information. Without a good grasp of the core data warehousing success issues, spending more

money can potentially create bigger problems and result in more expensive failures.

To better understand critical success factors *and* their effects on data warehousing success, this research addresses three interrelated questions:

1. What constitutes data warehousing success?
2. What factors contribute to data warehousing success?
3. Which factors are related to which success measures?

The first question addresses the fundamental issue of how to define the success of data warehousing. Although data warehousing success is likely a multifaceted construct, no consensus exists as to what the underlying measures are and how important each measure is. This study consolidates the multiple measures of data warehousing success discussed in the literature and determines whether any measures are more significant than others.

The second question explores the many factors that may contribute to data warehousing success. This study consolidates the numerous factors discussed in the literature and determines whether some are more important than others. Lastly, different factors are likely to be significant for different success measures.

The final question aims to identify the critical factors for each success measure.

Methodology

Authors have touted numerous benefits and advantages of data warehousing since the early days of its development. (Sakaguchi and Frolick, 1997; Vatanasombut and Gray, 1999; Watson and Haley, 1997; Watson et al., 2001) A review of the literature produced a consolidated list of eight data warehousing success measures:

- ease of use
- speedy information retrieval
- more information
- better-quality information
- improved productivity
- better decisions
- improved business processes
- increased competitive position

The first two measures are related to the quality of the system (data warehouse) itself and the next two are concerned with the quality of the system's outputs. The following two benefits are associated with how these outputs can benefit the individual users, while the last two measures involve how the outputs may produce a firm-level benefit.

There has also long been a keen interest in identifying factors that contribute to the success or failure of data warehousing (Sammon and Finnegan, 2000; Vatanasombut and Gray, 1999; Watson and Haley, 1997). A consolidated list of 11 success factors was similarly generated by a review of the literature. These success factors are related to the four feasibility tests of information systems (IS) development projects:

- Operational: Clearly defined business needs/ benefits, top management support, and user involvement/participation
- Technical: Source-data quality, proper development technology, adequate IS staff and consultants, and project management (teamwork)
- Schedule: Practical implementation schedule and proper planning/scoping
- Economic: Adequate funding and measurable business benefits

A Web-based questionnaire was developed to collect data from data warehousing professionals. A random sample of 6,000 names was selected from a list of over 15,000 data warehousing professionals associated with TDWI. These professionals were offered a $10 gift certificate from Amazon.com as an incentive to participate. A follow-up e-mail was sent three weeks later, and the two rounds of mailings yielded 98 completed questionnaires.

Results

Demographics

Figures 1 through 6 break out the demographic data.

The largest group of respondents (38 percent) was composed of data warehousing specialists, followed by DBAs (21 percent). They worked in a variety of industries, including consulting/professional services (16 percent) and federal government (11 percent).

The largest group of organizations (34 percent) had annual revenue less than $10 million, followed by those with annual revenue between $100 and $500 million (18 percent).

Over one-third of the data warehouses (32 percent) took from six to 12 months to develop; 26 percent took 12 to 24 months.

The largest group of data warehouses (22 percent) was deployed two years ago; 20 percent were deployed three years ago.

One-third of data warehouses (33 percent) were less than 100 gigabytes in size. Data warehouses of 100 to 500 gigabytes were found in 16 percent of organizations, the same percent as had data warehouses ranging from 500 gigabytes to one terabyte.

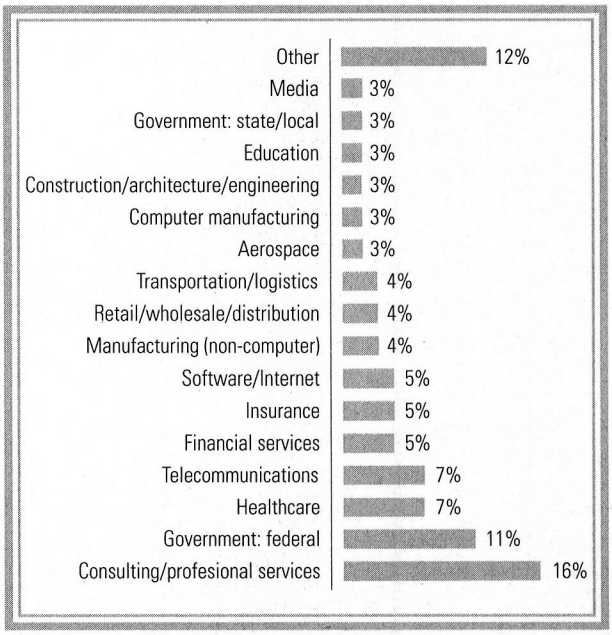


Figure 2. *Respondents' industry*

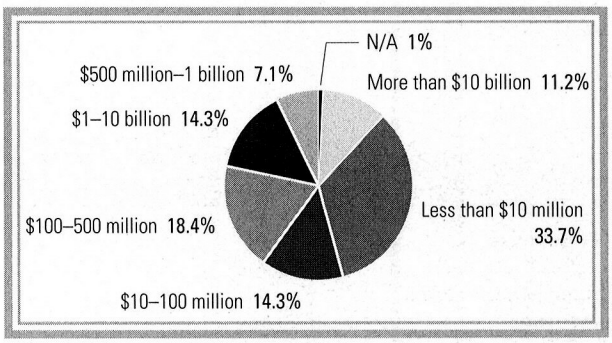


Figure 3. *Annual revenue*

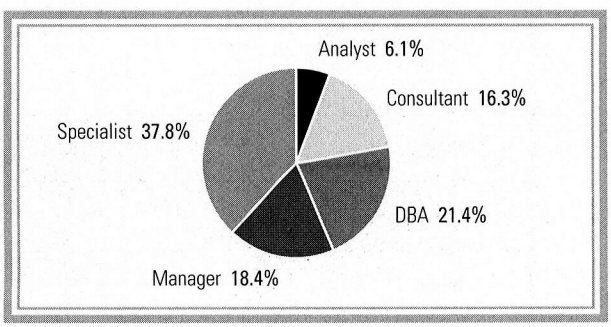


Figure 1. *Respondents' position*

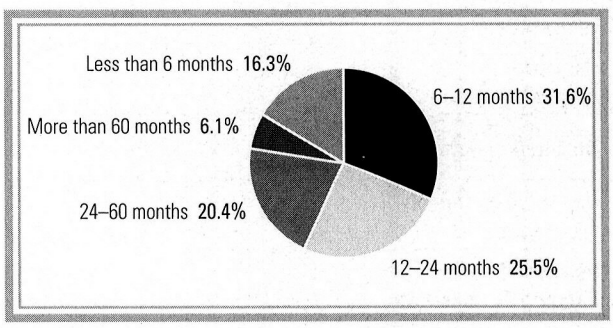


Figure 4. *Development time*

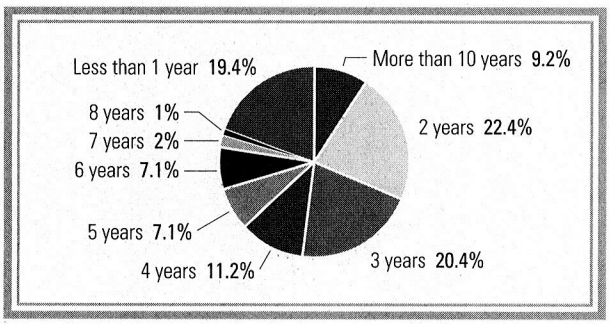


Figure 5. Age of data warehouse

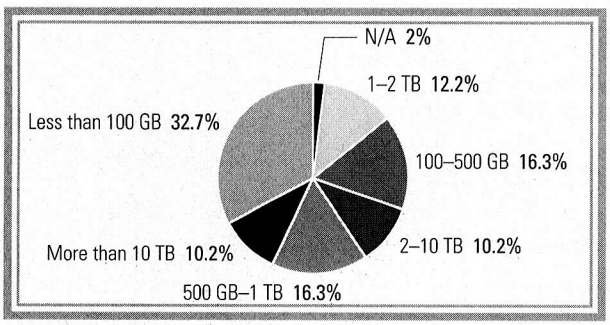


Figure 6. Size of data warehouse

Data Warehousing Success

Table 1 shows the breakdown of responses to the question: How significant is each benefit to your warehouse? The most popular responses are in bold. It appears that respondents considered their data warehouses successful as most rated each benefit as either significant or very significant.

Using a scale of 1 (not significant) to 5 (extremely significant), the mean responses were calculated and shown in the first column under each benefit. The mean responses ranged from 3.56 to 4.02, which indicates how significant each benefit might be considered in data warehousing. For example, respondents valued improved productivity (4.02) higher than more information (3.56) from a data warehouse.

We further investigated whether those perception differences were statistically significant. An ANOVA (analysis of variance) test confirmed the significance ($F = 3.893$, $\alpha = 0.000$) and subsequent multiple comparisons indicated the specific means that were significantly different. Those means are shaded in Table 1.

	Mean (1–5 scale)	Not Significant	Somewhat Significant	Significant	Very Significant	Extremely Significant
Easy to use	3.71	2.0%	7.1%	27.6%	**43.9%**	19.4%
Speedy information retrieval	3.98	0.0%	6.1%	19.4%	**44.9%**	29.6%
More information	3.56	0.0%	6.1%	**43.9%**	37.8%	12.2%
Better quality information	3.99	2.0%	2.0%	20.4%	**45.9%**	29.6%
Improved productivity	4.02	0.0%	5.1%	17.3%	**48.0%**	29.6%
Better decisions	3.94	0.0%	5.1%	28.6%	**33.7%**	32.7%
Improved business processes	3.85	2.0%	5.1%	23.5%	**44.9%**	24.5%
Increased competitive position	3.61	4.1%	11.2%	**30.6%**	27.6%	26.5%

Table 1. Responses on the significance of data warehousing benefits

	Mean (1–5 scale)	Not Important	Somewhat Important	Important	Very Important	Extremely Important
Clearly defined business needs/benefits	4.01	1.0%	6.1%	18.4%	**38.8%**	35.7%
Top management support	3.79	0.0%	10.2%	25.5%	**39.8%**	24.5%
User involvement/ participation	3.78	3.1%	2.0%	31.6%	**40.8%**	22.4%
Source-data quality	4.01	3.1%	3.1%	19.4%	**38.8%**	35.7%
Proper development technology	3.84	0.0%	5.1%	28.6%	**43.9%**	22.4%
Adequate IS staff and consultants	3.71	1.0%	3.1%	**41.8%**	31.6%	22.4%
Project management (teamwork)	3.73	0.0%	6.1%	33.7%	**40.8%**	19.4%
Practical implementation schedule	3.61	0.0%	8.2%	**41.8%**	34.7%	15.3%
Proper planning/ scoping of project	3.57	1.0%	2.0%	35.7%	**40.8%**	20.4%
Adequate funding	3.78	0.0%	8.2%	20.4%	**42.9%**	28.6%
Measurable business benefits	3.86	1.0%	5.7%	**34.7%**	25.5%	33.7%

Table 2. *Responses on the importance of success factors*

Critical Success Factors

Table 2 shows the breakdown of responses to the question: How important is each factor to the success of your warehouse? The most popular responses are indicated in the table in bold. Respondents considered those factors important as most rated each factor as either important or very important.

We used the same 1-to-5 scale; the mean responses are shown in the first column under each success factor. The mean responses ranged from 3.57 to 4.01, indicating the importance of each factor. Respondents considered "clearly defined business needs/benefits" (4.01) more important than "practical implementation schedule" (3.57) to the success of data warehousing. An ANOVA test confirmed the significant difference ($F = 2.029$, $\alpha = 0.028$) and subsequent multiple comparisons indicated that three means were significantly different. Those means are highlighted in Table 2.

As expected, "clearly defined business needs/benefits" was rated significantly more important than "practical implementation schedule," as was "source-data quality."

Impact of Critical Success Factors

Results presented in the previous two sections are useful to corporations looking to measure what is important in data warehousing success. Moreover, the results help companies focus on the factors most critical to data warehousing success. Further insight comes from examining

the relationship between the critical factors and success measures. Table 3 shows the correlation coefficients of critical success factors and each benefit.

Correlation coefficients indicate how strongly two variables are correlated and can range from -1 to +1. A zero correlation means the two are unrelated. Statistically significant correlations are shown in bold in Table 3, all of which are positive. This suggests that they have a strong positive effect on data warehousing success. These significant correlations offer additional insight into the role that each critical success factor plays in different measures of success. For instance, the last section shows that "clearly defined business needs/benefits" is a more important factor than "practical implementation schedule." However, when the objective is to build a data warehouse that is easy to use, a practical implementation schedule is more important than clearly defined business needs/benefits.

Since different organizations may have different goals or emphases in their data warehousing efforts, Table 3 offers the factors that they should address in their particular environment.

DISCUSSION

This research has enhanced the understanding of data warehousing success and factors that impact success. For organizations looking to identify and measure success, results of this study offer three benefits/success measures to consider: improved productivity, better-quality information, and speedy information retrieval. Interestingly, each of these three benefits measures a different aspect of success. The most highly rated benefit of a data warehouse is its ability to improve the productivity of its users. The quality of the data warehouse as measured by speedy information retrieval and the quality of information it provides are also significant benefits.

For the critical success factors, results of this study are consistent with the literature and suggest that organizations clearly define the business case as the first step in their data warehousing endeavors. Another factor that demands top

	Clearly defined business needs/benefits	Top management support	User involvement/ participation	Source-data quality	Proper development technology	Adequate IS staff and consultants	Project management (teamwork)
Easy to use	.23	.13	**.40**	**.34**	.23	.09	-.01
Speedy information retrieval	13	.14	**.33**	**.32**	**.28**	.09	.08
More information	.08	.18	**.26**	**.30**	.17	**.31**	**.35**
Better-quality information	**.32**	**.31**	**.35**	**.54**	**.35**	**.26**	**.46**
Improved productivity	**.28**	.06	**.44**	**.32**	**.31**	**.31**	.23
Better decisions	**.39**	**.30**	**.32**	**.43**	**.34**	**.39**	**.59**
Improved business processes	.23	.18	.21	**.29**	.21	.21	.24
Increased competitive position	**.37**	.11	**.27**	.21	**.32**	.22	**.37**

Table 3. *Correlations of data warehousing benefits and success factors*

priority is source-data quality, which again has been advocated in various studies. After all, a data warehouse is only as good as the data that goes into it. Other factors, however, do play an important role depending on the specific benefit of data warehousing that is being pursued. Table 3 provides a handy reference for the factors to focus on for various data warehousing benefits.

CONCLUSION

Data warehousing is an important area of practice and research, yet few studies have assessed data warehousing practices in general and critical success factors in particular. Although many guidelines for implementation exist, most of them are derived from anecdotal evidence. This research provides empirical data on data warehousing success and factors that contribute to success.

The results confirm that data warehousing success is a multi-faceted construct and suggest that a data warehouse be built with improving user productivity in mind. Other highly valued benefits include speedy information retrieval and better quality information. The results also indicate that success hinges upon factors of different dimensions. For operational feasibility, clearly defined business needs/benefits stands out; for technical considerations, source-data quality tops the list. Equally important, different factors are significant when different benefits of data warehousing are considered. This contingency perspective helps an organization prioritize its pursuit of data warehousing objectives. ■

Practical implementation schedule	Proper planning/ scoping of project	Adequate funding	Measurable business benefits
.27	.16	**.35**	**.36**
.12	.07	**.30**	.23
.33	.25	**.41**	.22
.16	**.32**	.17	.24
.35	**.31**	**.47**	**.44**
.14	.21	.18	**.35**
.14	.13	**.37**	**.50**
.25	.14	**.35**	**.57**

REFERENCES

Mukherjee, D., and D. D'Souza. "Think Phased Implementation for Successful Data Warehousing," **Information Systems Management**, (2003), 82-90.

Sakaguchi, T., and M. Frolick. "A Review of the Data Warehousing Literature," **Journal of Data Warehousing**, Vol. 2, No. 1 (1997), 34-54.

Sammon, D., and P. Finnegan. "The Ten Commandments of Data Warehousing," **Database for Advances in Information Systems**, Vol. 31, No. 4 (Fall 2000), 82-91.

Vatanasombut, B., and P. Gray. "Factors for Success in Data Warehousing: What the Literature Tells Us," **Journal of Data Warehousing**, Vol. 4, No. 3 (1999), 25-33.

Watson, H., D. Annino, B. Wixom, K. Avery, and M. Rutherford. "Current Practices in Data Warehousing," **Information Systems Management**, Vol. 18, No.1 (2001), 47-55.

Watson, H., and B. Haley. "Data Warehousing: A Framework and Survey of Current Practices," **Journal of Data Warehousing**, Vol. 2, No.1 (1997). 10-17.