DATA WAREHOUSE ARCHITECTURES: SELECTION FACTORS AND SUCCESS EVALUATION

by

THILINI ARIYACHANDRA

(Under the Direction of Hugh J. Watson)

ABSTRACT

There is an ongoing debate about which data warehouse architecture should be used to implement a data warehouse. Using existing theory and interviews with a panel of experts, this research proposed a model that identifies the various selection factors that affect the choice of data warehouse architecture. In addition, the research also evaluated the success of the various architectures. Specific hypotheses were created to investigate how organizational factors influence architecture selection and how data warehouse architectures influence architecture success.

An empirical investigation involving an online field survey was conducted to further validate and test the proposed research models on architecture selection and success. Data collected for the main data warehouse architectures were used in the statistical analyses. The results suggest that various combinations of organizational factors influence the selection of a data warehouse architecture. Further, one architecture was identified as the least successful data warehouse solution while another emerged as the most costly to develop when compared to the rest. Through a carefully designed empirical investigation of data warehouse architectures, this research sheds light on the data warehouse design selection in organizations by examining the impact organizational factors on data warehouse architecture selection and by evaluating data warehouse architecture success.

INDEX WORDS: Data warehouse, Architecture, IT infrastructure, IS success

DATA WAREHOUSE ARCHITECTURES: SELECTION FACTORS AND SUCCESS EVALUATION

by

THILINI ARIYACHANDRA

B.Sc., University of South Alabama, 1998

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

DATA WAREHOUSE ARCHITECTURES: SELECTION FACTORS AND SUCCESS EVALUATION

by

THILINI ARIYACHANDRA

Major Professor:	Hugh J. Watson
Committee:	Dale L. Goodhue Robert J. Vandenberg Elena Karahanna Marie-Claude Boudreau

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2004

To

Ammi

Whose love and courage gives me strength

ACKNOWLEDGEMENTS

This work would not have been possible with out the help, advice, and support of many individuals. First of all, Hugh, Dale, Bob, Elena, and Maric, thank you for working with me on this research. I feel truly blessed to be privy to the advice and guidance of my committee. Dale, your enthusiasm for research is so inspiring. I believe that my passion for research stems from your influence throughout my doctoral experience. Bob, your guidance on methodology was invaluable. Thank you for your patience with me. Elena, your words of encouragement on both research and career development always motivated me to work harder. Maric, my confidant in academia, thank you for your advice and moral support. Hugh, chair of my committee, mentor, father figure, and friend, you willingly took on so many roles in my life when I needed help and guided me through this entire journey. Your influence, kindness, and guidance have had and still continue to have a tremendous influence on my life. I am grateful to you for being part of it.

My sincere thanks also go out to the experts within the data warehousing community who helped me refine my research model and questionnaire, and promoted the survey to potential respondents. Your interest and enthusiasm in this research was pivotal in shaping this study. Especially, thanks to Ron Swift, Julie Kimball, Margy Ross, Karolyn Duncan, and Wayne Eckerson for spending your invaluable time providing input on this research.

I also owe thanks to fellow students, other faculty, and staff in the MIS department. Umashanger, I am sincerely thankful for your kindness and willingness to help me during the last few months of this process. Rick, I really appreciated your guidance and advice during the former years of my time at UGA. Liddell, Daniel, Saurabh, and Rui, we have all been through so much together in the past few years from working in the basement of Brooks Hall to celebrating marriages and new arrivals. Thank you for your friendship.

Friends and family all across the world also provided much needed social and moral support throughout this process. Especially, Ayomi, Nina, Cem, and Tams, I will cherish your friendship.

Finally, my family, I owe so much to you. Thaththi, your life and your goodness is my inspiration in life. It is your love for knowledge that motivated me to pursue this doctorate. Ammi, you encouraged me to realize my goals and unselfishly sacrificed much of your happiness to make my life special. Thank you. Malli, my one and only little brother, thank you for caring, for your loyalty, and for your love. And last, but certainly not least, Kushala, my better half and my pillar of strength, I owe you the deepest and warmest thanks. Your optimism and smile kept me going through the final stages of this process. Without you I would have never completed this work on time.

When embarking on this journey, I had no idea that it would be a life altering experience. It has allowed me to stretch my mind and build character. I am eternally grateful to all my loved ones, friends, and my family at the MIS department in Athens, Georgia for being my support network and for making this journey rich, interesting, challenging, and wonderful.

TABLE OF CONTENTS

Page

CHAPTER 1 – INTRODUCTION

1.1 Introduction

The beginnings of the information age date back to the late 1950's when the number of workers that handled information in organizations exceeded the number of industrial workers (Mukherji 2002). Since then, the role of information in organizations has gradually gained significance as a vital resource necessary to monitor and respond to competitive market conditions (Edelman 1981; Goodhue et al. 1988). Along with the rise in information intensity in business, organizations' capabilities to handle information have also increased (Fedorowicz et al. 1992). Currently, organizations use a variety of information systems dispersed across the organization to support their information needs.

Concurrent with this proliferation of information systems, there has been an overabundance of data captured in organizations, which has created an information crisis. Extracting data from dispersed sources, and combining it to satisfy organizational information needs, has become a challenging task (Niederman et al. 1991). Thus, most companies are faced with an information crisis not because of the lack of data, but because their information technology (IT) capabilities are deficient in effectively managing organizational data resources (Nimmer 1990). The current IT infrastructure of many organizations does not provide them with the ability to combine relevant data housed in dispersed data silos to satisfy business needs. Surveys of information system (IS) professionals have identified the need for a responsive IT infrastructure that can effectively manage the data resource as a major issue (Brancheau et al. 1996; Niederman et al. 1991).

The data warehouse, an IT infrastructure popularized in the 1990's, provides a unique opportunity to improve the effective management of data resources (Agosta 2001; Nemati et al. 2002; Watson et al. 2002; Whiting 2003). A data warehouse addresses data management, integration, and access issues by capturing data from diverse sources and creating a repository of quality data that can be manipulated to meet varying information needs.

Since its introduction as a source of decision support data, the data warehouse is recognized as a mandatory component of the critical infrastructure required for strategic organizational initiatives like customer relationship management (Dobbs et al. 2002; Goodhue et al. 2002), performance management (White 2003), and supply chain integration (Watson et al. 2001). It is also seen as a powerful IT change agent for organizational transformation (Cooper et al. 2000). In the global arena, data warehouses in leading companies are recognized as an important strategic resource used to drive new business opportunities (Watson et al. 2002).

In addition, the data warehouse has evolved to being considered an essential component in most organizations' business intelligence (BI) environment (Strange 2001). As a result, the objective of data warehouse initiatives is shifting from a focus on data transformation into information to a focus on data transformation into intelligence (Moncla 2000). Consequently, as the BI market grows to a projected 12 billion by 2006 (Darrow 2003), data warehouse technology will increasingly become an important part of the IT infrastructure requirements in organizations (Cohen 2003).

Despite the growing recognition and importance of the data warehouse, the data warehouse industry is currently wrought with debate and confusion (Koch 1999). There is no industry consensus on the best architecture to use for this complex infrastructure venture. This has left novice data warehouse developers with no clear direction on how to embark on these high-cost investments (Eckerson 2002). Furthermore, it has hindered efforts to create industry-wide best practices in data warehouse implementation.

Currently, the need to identify the most suitable data warehouse architecture has gained industry experts' attention as a key issue. (Hackney 2000; Joshi et al. 1999; Wells et al. 2002). A study of existing data warehouse initiatives by the Meta Group revealed architecture as one of the prevailing essential factors influencing the success of data warehouse initiatives (Laney 2000). A recent Gartner Group report on BI and data warehousing strategic initiatives named data warehouse architecture as one of the five problem areas facing data warehouse projects today (Strange 2003).

This dissertation explored this problem by investigating how a data warehouse architecture is selected by companies and by evaluating the success of the various architectures. This chapter first introduces the data warehouse and data warehouse architecture. Next, the research questions are presented and the methodology used to conduct the dissertation research is discussed. Finally, the contributions of the dissertation to both academia and industry and an overview of the remaining chapters are described.

1.2 The data warehouse

While the term "data warehouse" was first coined and popularized by Bill Inman (1992), the concept of a data warehouse technology was first described by Devlin and Murphy (1988) at IBM. A data warehouse is a specially prepared data repository designed to support decision making (See Figure 1.1). Raw data extracted from multiple disparate data sources is transformed and stored in data stores such as data marts and/or data warehouses. It is then made available for end user manipulation using SQL, analytical tools (e.g., managed query environment), or decision-support applications (Gray et al. 1998). The process of activities from data extraction through end user data manipulation using analytical tools is called data warehousing. A data warehouse is the repository of data while data warehousing is broader in scope.

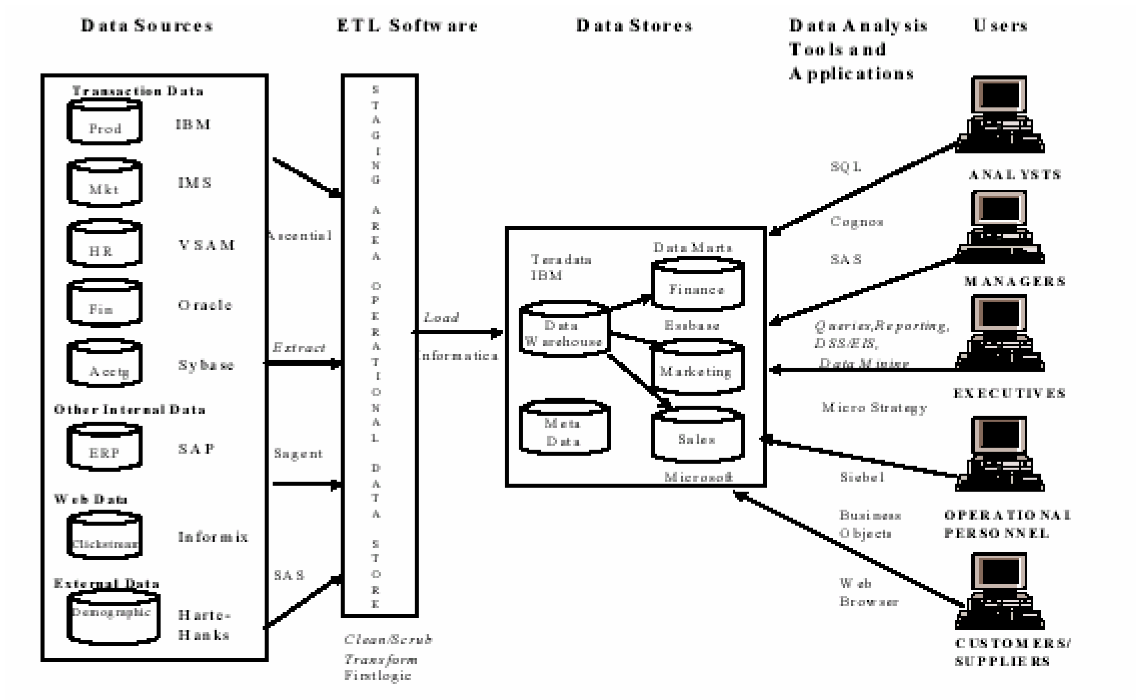


Figure 1.1: The data warehouse (adopted from Watson 2001)

The data warehouse is an IT infrastructure that has become a critical part of the overall IT infrastructure for the entire organization. It is a decision-support infrastructure, as it supports the creation of IT applications that enhance decision making. It is a complex system that can serve as a foundation for the development of new applications to support strategic initiatives (Devlin 1997). However, implementing a data warehouse is very challenging and the cost of the average data warehouse runs close to 2.2 million (Desai 1999). The design and development of a data warehouse requires an overall blueprint (i.e., a suitable data warehouse architecture) to guide the implementation effort.

1.3 Data warehouse architecture

The data warehouse literature reveals many discussions describing the debate on the best way to build a data warehouse (Hackney 2000; Koch 1999; Sen et al. 1998). However, the literature often does not clarify whether the differences in implementing a data warehouse are due to variations in data warehouse architecture or methodology (Wells 2003). Over the years, the differences between the various data warehouse implementation methodologies have diminished (Eckerson 2002), revealing that the true distinctions arise due to variations in the architecture (Hackney 1998).

Data warehousing vendors and experts have promoted a variety of architecture solutions to guide data warehouse implementation efforts. At present, there are two main competing architectures advocated for data warehouse development: the enterprise data warehouse architecture (EDW), and the data mart bus architecture (DBA). In addition, independent data marts (IDM) and the federated architecture (FED) are two other architectures that are also used in some warehouse implementations. None of the four data warehouse architectures has been associated with a 100 percent success or failure rate. Thus, contrary to the claims of experts of different warehouse architecture camps, there is no one architectural design that has been shown to succeed above the rest.

According to Joshi and Curtis (1999), the best way to build a data warehouse is dependent on the idiosyncratic organizational circumstances of each data warehouse initiative. A multitude of organizational factors affect the selection of a data warehouse architecture (e.g., Koch 1999, Hackney 2000, Wells 2003). They range from business requirements (Joshi and Curtis 1999), such as the need for greater data integration, to aspects of the organization's political environment (Hackney 2000), such as the influence of powerful business units. Some of these organizational factors can be explained using information processing theories (e.g., Galbraith 1973) while others can be understood using social /political theories (e.g., Pfeffer 1981). These theories help understand how selection factors, that describe attributes of organizational circumstances, can affect the choice of a data warehouse architecture.

The selection of a data warehouse architecture has implications for the data warehouse that is constructed. The architecture potentially affects specific aspects of the data warehouse, such as system and information quality. As a result, it may impact the net benefits from the data warehouse as well. These attributes reflect the success of the data warehouse architecture selected for implementation. As such, they can be used to evaluate the success of the different architectures.

1.4 Research questions

Although industry experts have offered several prescriptive analyses of the impact of organizational factors on data warehouse architecture, no academic research has been conducted to identify how organizational setting impacts data warehouse architectural design. Furthermore, no hard evidence exists to suggest that any one data warehouse architecture is the most successful. Given the growing demand for data warehouses (Cohen 2003), the tremendous dollar investment required for their development (Desai 1999), and the high rate of initial failures (Inman 2001), there is a need to better understand the data warehouse architecture selection process as well as to evaluate the success of the various architectures. Therefore, this dissertation investigated two research questions.

First, in order to discover and examine how organizational factors influence the selection of a particular architecture, the following research question was investigated:

What is the relative importance of organizational factors in the selection of a particular data warehouse architecture?

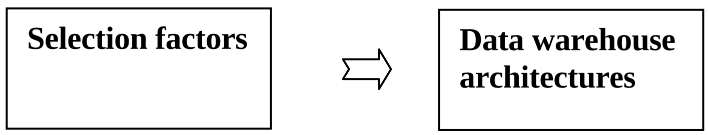


Figure 1.2: Conceptual model for a particular data warehouse architecture selection

Next, in order to evaluate the success of the different architectures, the following question was posed:

How successful are the alternative data warehouse architectures?

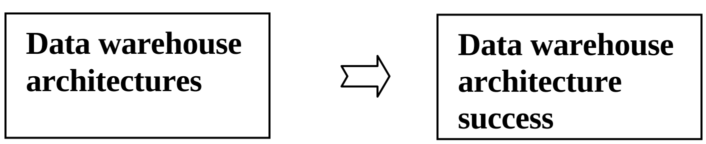


Figure 1.3: Conceptual model for data warehouse architecture success evaluation

While the conceptual model shown in Figure 1.2 represents the research question linked to architecture selection, the conceptual model portrayed by Figure 1.3 represents the research question linked to architecture success. These conceptual models are described in detail in Chapter 3.

1.5 Research methodology

Choosing the most appropriate research methods to conduct IS research has been the cause for much concern in the IS field. While some researchers stress the need for uniformity in IS research, others encourage diversity of methods as a means of increasing the IS knowledge base (Mingers 2001). According to Robey (1996), the choice of research design is dependent on the practicality and applicability of available research methods to accomplish a research aim. Furthermore, use of more than one research method results in a stronger research design that takes advantage of the strengths of different methods (Lee 1991). Finally, descriptive research objectives[1] prescribe the use of methods that enable the creation of a rich understanding of a complex phenomenon (Punch 1998).

The research design for the dissertation involved two phases: (1) using qualitative semi-structured interviews and (2) a field survey that provided a fuller understanding of data warehouse architecture selection and data warehouse architecture success. Semi-structured interviewing is often used in IS research in the initial exploratory phase of the investigation to gain a clear understanding of the phenomenon under study (Avison 1997). Survey research, a popular research design in the IS field (Pinsonneault et al. 1993), is often used for descriptive purposes and is considered a good method to examine relationships among variables (Babbie 2000).

The initial phase of the dissertation entailed interviewing a panel of ten experts to further enrich the existing literature on selection factors and success measures. The expert interviews and existing literature led to the creation of a comprehensive research model that described the salient selection factors that influence data warehouse architecture. It also identified the success

[1] The main purpose of research may either be exploration, description, or explanation. The purpose of descriptive research it to describe the way things are or answer the “what” question in conducting research. The various objectives of research are described in Chapter 5.

factors influenced by architecture implementation. The second phase, the main task of study, involved conducting a field survey of data warehousing organizations. The survey instrument was created using a literature and interview data, and was reviewed by data warehousing experts for content validity. The field survey, targeting the person most familiar with the design, development and implementation of a data warehouse as survey respondents, established empirical support for the research model created in phase one. The unit of analysis for the study was the data warehouse architecture in the companies surveyed.

After survey questionnaire validation and data collection, the data analysis involved a series of statistical analyses. First, the overall measurement model was assessed using confirmatory factor analysis to establish the validity and reliability of measures. Next, the analysis of data collected on the different selection factors and the architecture selected using logistical regression techniques revealed how several combinations of selection factors influence the selection of a particular data warehouse architecture. Finally, an assessment of architecture success measures using analysis of variance techniques and structural equation modeling helped determine the variation in success of the different architectures.

1.6 Importance of the research

The findings of this study contribute to the enrichment of academic research in information systems as well as provide actionable insight to the world of practice.

1.6.1 To academia

This study contributes to academic research in several ways. First, the IS field has been criticized for not addressing topics relevant to practice (Lee 1999, Benbasat and Zmud 1999, Lyytinen 1999, Ives 1992). Benbasat and Zmud (1999) make a strong plea for more relevance and specify actions to make research more relevant. For instance, they propose that researchers identify topics from IS practice that are current, interesting, and applicable; thus, investigate topics that have the potential to be useful by practitioners. This research was primarily motivated by the current industry debate on data warehouse architectures. Thus, it focuses on a timely and relevant topic to the data warehouse industry that adds to the current sparse empirical research in data warehousing.

Second, specifically in data warehousing research, there is a wide gap between theory and practice (Jeusfeld 2001). Most early academic research in data warehousing considered the technical aspects of the data warehouse, such as the efficient management of materialized views. Meanwhile organizations have showed an interest in more business-oriented issues. Thus, there is a need for more business-oriented research in data warehousing. This research is a first exploratory look at how context affects the choice of a data warehouse architecture. In so doing, it helps extend the scope of data warehouse research in academia by supplementing the few research efforts that examine the business-oriented aspects of data warehousing.

Third, understanding how organizational phenomena affect the development and use of technologies and how technologies shape organizations is central to the agenda of the IS field (Orlikowski et al. 2001). This study looks at how organizational context affects the design of a data warehouse. It supplements the field’s understanding of organizational impacts on IT design by examining it in the context of an IT infrastructure.

Finally, the IS field is rich with research that has observed the design process for IT applications (Benbasat et al. 1978; Chen et al. 1997). In contrast, relatively few studies have examined the architectural design of IT infrastructure. Though the importance of infrastructure flexibility has been highlighted recently (Duncan 1995, Bryd and Turner 2000, Lewis and Bryd 2003), no research has explored the role of infrastructure architecture in creating infrastructure flexibility. The few research efforts that have looked at infrastructure design provide either anecdotal advice or descriptive case illustrations of various aspects of the design process (Fan et al. 2000; Watson et al. 2001). The empirical investigation in this study is one of the first empirical studies in IT infrastructure design. Thus, by exploring the impact of contextual variables on architectural design, this study provides insight into an important aspect of IT infrastructure, which has received less attention in the past. In so doing, this dissertation contributes to the development of the IT infrastructure research stream, which is still in its infancy.

1.6.2 To practitioners

The current practitioner literature is rich with prescriptive anecdotal advice on data warehouse architecture based on personal experience, opinions, and conjecture (e.g., Joshi and Curtis 1999; Hackney 2000). Practitioners lack results from carefully conducted research in order to help guide their architecture selection process. This study offers results from a carefully constructed research study.

The outcome of this dissertation describes how organizational selection factors affect the selection of data warehouse architectures as well as indicates the variation in the success of different architectures. These results will help new comers to the data warehouse industry to select an architecture that fits their organization's needs. The results will also assist vendors in refining their product offerings by considering the impact of organizational factors on architecture choice. Consequently, the research findings may lead to more comprehensive benchmarks that consider variations in organizational context (Swan et al. 1999).

According to the Gartner Group, most organizations are concentrating on redoing their existing data warehouses or looking for ways to gain greater benefits from their data warehouses (Strange 2003). Implementing an architecture that does not fit the organization can lead to high development and maintenance costs (Joshi and Curtis 1999). The results of this study may help organizations to gain cost savings by implementing an architecture that fits the organization. Furthermore, the results also help prevent the need to redo data warehouse initiatives in the future.

Over the past few years, the importance of business intelligence (BI) applications and the infrastructure that support BI has been growing (Cohen 2003). AMR research estimates that the current BI market is at six billion with a projected growth to reach 12 billion by 2006 (Darrow 2003). BI is seen as a way to increase business knowledge, and to react to market and competitive forces in real time (Robinson 2002). The data warehouse is a critical component of the BI infrastructure. In order to effectively use BI applications, it is essential that the underlying decision-support infrastructure be compatible with the underlying organizational structure. To accomplish this task, it is necessary to select a data warehouse architecture that is suited for the organization. This study provides the necessary knowledge to implement a data warehouse architecture that provides a sound infrastructure to support BI applications.

Gaining business agility to counter hypo competition has become the latest aspiration of contemporary organizations (Sambamurthy et al 2003). A flexible IT infrastructure is seen as a key resource necessary to attain business agility. The results of this study provide practitioners with the advice necessary to build a flexible decision-support infrastructure that increases business agility in the competitive market.

1.7 Overview of the dissertation

Chapter 1 describes the data warehouse and its architecture. It presents the research questions that were investigated as well as the contributions of the research to both academe and practice. Finally, the chapter discusses the methodology of the study.

Chapter 2 discusses the concepts of architecture and data warehouse architectures. Next it presents a review of the data warehouse architectures by describing the characteristics of each of the four main architectures.

Chapter 3 presents the results of the expert interviews and the literature review relevant to this dissertation. It describes the exploratory phase of the study involving interviews with experts. It also includes the literature and theory relevant to selection factors, and success measures. Whenever possible, the chapter also presents hypotheses describing relationships among selection factors, data warehouse architectures, and success measures.

Chapter 4 describes the two overall research hypotheses that were tested and the specific hypotheses for relationships between specific selection factors, architectures, and success measures.

Chapter 5 discusses the methodology that was used to conduct this study. It presents details related to the research design, sampling plan, and the research process utilized in the study.

Chapter 6 presents the operationalization of the variables studied and the validation process used.

Chapter 7 describes the results of the statistical data analyses conducted for architecture selection and architecture success.

Chapter 8 discusses the overall findings on architecture selection and success. The specific hypotheses that were not support by the data are also presented and discussed.

Finally, Chapter 9 summarizes the research findings from the study. It also explains how both academe and practitioners can gain from the findings from the current research. Lastly, it presents the limitations of this dissertation and provides suggestions for future research.

1.8 Conclusion

This chapter first introduced the concepts of the data warehouse and data warehouse architecture. Next, the research questions were presented and the methodology used to conduct the dissertation research was discussed. Finally, the contributions of the dissertation to both academia and industry and an overview of the remaining chapters were described. The next chapter reviews the four main data warehouse architectures.

CHAPTER 2 – DATA WAREHOUSE ARCHITECTURE REVIEW

2.1 Introduction

This chapter discusses the concept of data warehouses, IT infrastructure, and data warehouse architecture. This is followed by a discussion of each of the data warehouse architectures.

2.2 The data warehouse

A data warehouse is an information delivery system built to support decision making (Gray et al. 1998). It makes decision-support applications possible without hindering operational systems and presents a flexible and interactive source of strategic information (Devlin 1997).

A data warehouse contains time variant data consolidated from operational databases (Sperley 1999). The data in a data warehouse exists in several forms: current detailed data, older detailed data, lightly summarized data and highly summarized data (Inmon et al. 1994). While the data warehouse is a repository of data, data warehousing has a wider scope, encompassing a broad range of activities, all the way from extracting data from source systems to the use of the data for decision-making purposes (Watson 2001). Specifically, it includes data extraction, transformation, and loading; the access of the data by end users; and applications.

Data marts and operational data stores (ODS) are two related types of data storage that are often discussed in conjunction with data warehouses (Inmon et al. 2001). A data mart is similar to a data warehouse, except a data mart stores data for a limited number of subject areas, such as marketing or sales data (Humphries 1999). As it is smaller in scope than a data warehouse, it also tends to contain fewer subject areas, less data, and supports fewer users and applications. A data mart can be thought of as a logically or physically partitioned subset of a data warehouse (Hackney 1997). It is usually constructed to serve the needs of a particular user community or business function (Kelly 1997).

While similar in many ways to a data warehouse, an ODS contains only current or nearly current data. An ODS is typical used by clerical and operational personnel to respond to daily needs of a business and customers, and it does not need to have the historical data that a data warehouse must store for use by strategic decision makers (Devlin 1997).

In defining a data warehouse, numerous industry experts have described its key distinguishing features (Gray et al. 1998). Inmon's (1992) definition of a data warehouse, which includes four main distinguishing characteristics, is probably the most widely known. He defines a data warehouse as a subject oriented, integrated, time variant, and non-volatile collection of data in support of decision making. *Subject oriented* means that data in a data warehouse is stored by business subjects such as sales or product (Ma et al. 2000). Unlike operational systems, it does not contain data sets organized around operational applications such as accounts receivable. When data is moved into the data warehouse, data inconsistencies are removed and data from diverse operational applications is *integrated* around a common identifier (Calvanese 2001). *Non-volatile* means that data in a data warehouse is updated at scheduled intervals and users cannot update the data (Singh 1999). The final distinguishing characteristic, *time variant*, means that the data warehouse contains both historical data and data considered current as of the last warehouse data load (Gardner 1998).

As the preceding description indicates, a data warehouse is vastly different from operational systems. It serves a different purpose from operational systems. A data warehouse is a decision-support infrastructure, which is a component of the overall IT infrastructure that supports basic decision-making needs such as integrated data and data access capabilities (Wixom et al. 2001). These data-related capabilities are considered part of the most important features of the overall IT infrastructure (Rockart et al. 1996). As a result, they require special attention in the design of the IT infrastructure.

2.2.1 Information technology infrastructure

In the past decade, the importance of information technology infrastructure (ITI) to attain strategic business objectives has gained considerable attention in the information systems (IS) field (e.g., Brancheau et al 1996; Lewis et al 2003). IT infrastructure is perceived as a potential weapon to combat dynamic competitive pressures resulting from ubiquitous connectivity, deregulation, the convergence of industries and technologies, and globalization. Organizations have realized that its ability to rapidly respond to changing market conditions is predicated upon the IT infrastructure's capability to be responsive and flexible to strategic business needs (Byrd 2000). For instance, companies are recognizing that the rapid development and growth of strategic initiatives such as customer relationship management (CRM) require a decision-support infrastructure that can combine relevant data dispersed across the organization (Swift 2001).

However, designing and building IT infrastructure is more complex, time consuming, and costly than IT applications (Chatterjee et al. 2002). Furthermore, most existing IT infrastructure in organizations is more a barrier than an enabler of strategic objectives. In fact, a recent study involving over 500 senior executives indicated that legacy infrastructure composed of incompatible databases and applications, poor data quality, and limited scalability constrained the development of organizations' strategic business initiatives (Prahalad et al. 2002). As a result, there is a need to discover how to plan, design, and build IT infrastructure (e.g. Brancheau et al 1996).

2.2.2 Designing IT infrastructure

Developing the complex set of shared IT resources that form an IT infrastructure requires creating an overall blueprint of the organization's technology needs (Duncan 1995; Rockart et al. 1996). Thus, much like drawing architectural blueprints and constructing a house based on a client's needs, developing an overall IT infrastructure involves information systems planning to create an enterprise IT architecture. Once created, the IT infrastructure architecture provides a framework for detailed analysis, design, and construction of the IT infrastructure (Maier et al. 2000). It is considered the long-range plan for the organization of IT resources in order to best meet the firm's business strategy.

Thus, a company's strategic business objectives affect the information systems planning process required to create an effective enterprise IT architecture. Information systems planning is a high-level planning process that integrates information systems considerations to the corporate planning process and business goals (Raghunathan et al. 1994). According to Ross (2003), there are three steps to creating an enterprise IT architecture: (1) defining the firm's strategic objectives, (2) identifying IT capabilities that support the strategic objectives, and finally (3) creating guidelines and selecting technical choices to attain the IT capabilities. The successful achievement of all three steps creates an effective IT architecture that reflects the tight alignment between strategic business goals and information systems goals. Broadbent and Weill (1997) discuss a similar approach to creating IT infrastructures for a specific organization by aligning business maxims and IT maxims.[2]

Completing these seemingly simple IS planning steps is challenging. Furthermore, alignment between business and IS goals can occur in many different ways (Henderson et al 1994). This suggests that IT architecture may not always originate from the specification of organizational strategic goals. The existing literature provides meager prescriptive advice and conjecture when addressing the creation of an IT infrastructure architecture (Byrd 2000; Duncan 1995; Rockart et al. 1996; Ross 2003). As such, while the distinct process leading to the design of IT architecture has not been rigorously examined, past literature suggests that IT infrastructure architecture design occurs as a result of strategic information system planning. Consequently, because a decision-support infrastructure is a component of the overall IT infrastructure, the architecture of a decision-support infrastructure may also be developed as a result of the planning process.

2.3 Architecture

Architecture is the study of the structure of something (Vail 2002). Architecting, the planning and building of structures, is considered “as old as human societies and as modern as the exploration of the solar system” (Richtin 1991). The concept of architecture has expanded to many different fields since classical architecture originated in ancient Egypt with the creation of the pyramids. New domains like aerodynamics, chemistry, and information processing are calling for new architectures to identify their needs for systems characterized by extraordinary complexity and technological capability (Maier et al. 2000). Architectural principles have been expanded to information technology and architectural frameworks have been introduced to be utilized in the construction of complex IT systems (Sowa et al. 1992; Zachman 1987). In the realm of data

[2] Broadbent and Weill (1997) describe maxims as short statements reflecting the mission or strategy statements.

warehousing, it has become conventional wisdom that a data warehouse is developed by first defining its architecture, which defines how source data is extracted, transformed, and stored.

Despite over 4000 years of history, the notion of architecture in buildings as well as other domains remains elusive (Rechtin 1991). Communities involved in architecture in systems, software, hardware, and other domains are still struggling to find a formal definition of architecture as it applies to their domain. In their effort to establish standard architecture descriptions for software intensive complex systems, the IEEE architecture work group described architecture as the highest level concept of a system in its environment (Maier et al. 2001).

The Webster dictionary provides another lens to look at architecture. According to the Webster, architecture can be defined as a method or style of building or as an architectural product. For instance, while the Gothic architecture of a church refers to a style of building, the end product architecture of the church refers to the architectural product. In the context of system design, ‘architecture as a product’ is referred to as the system architecture and ‘architecture as a style or method’ as the reference architecture (Wynns et al. 1997).

A system architecture abstracts the description of complex dynamic systems by providing a high-level conceptual blueprint(Maier et al. 2000). This abstract description of a particular system may specify the functions of components, the links between components, and constraints. This specification forms the basis for the detailed design and implementation stages of systems development. As a result, the systems architecture helps the designer define and control the integration of the system components within a specific systems development context (Zachman 1987).

Alternatively, the reference architecture forms the basis for designing the system architecture for a particular system (Maier et al. 2000). A reference architecture is the generalized architecture of a type of end systems. For instance, the reference architecture for a data warehouse would define the components that are common to data warehouses and provide guidance for the development of a specific system architecture for a particular data warehouse implementation project.

Thus, a reference architecture helps form the design of a system architecture by identifying standard components, describing the responsibilities of components, giving example system architectures, and even defining a development methodology (Lloyd et al. 1999). When designing a system according to an architectural style (i.e., reference architecture) the designer can customize the standard components in ways that are appropriate to create the desired system architecture.

The four main data warehouse architectures currently recognized in the industry are reference architectures. Each reference architecture offers a basis to create a system architecture for a specific data warehouse initiative. The circumstances surrounding an organizational setting determine the data warehouse reference architecture most suited for the data warehouse development initiative. In this study, the main focus was on the data warehouse reference architectures. It identified how contextual factors influence the selection of a data warehouse reference architecture in order to create the specific system architecture for a data warehouse initiative.

The Zachman framework is an architecture framework that emphasizes organizational aspects of selecting an information technology architecture. It is a widely known architecture framework in information technology that establishes standards for describing architectures from the viewpoints of different stakeholders in the organization (Maier et al. 2000). In order to understand where data warehouse reference architectures fit in a framework of architecture descriptions, the Zachman framework is described next.

2.3.1 The Zachman framework

The current dynamic business environment has broadened the scope of information systems design and increased the complexity of IS implementations. For instance, numerous details and relationships must be considered simultaneously when building complex systems and IT infrastructure. As a result, the use of a logical structure to define and control the development and integration of all components of information systems has become a necessity.

The Zachman framework for information systems architecture (ISA), first proposed in 1987 (Zachman 1987) and later extended in 1992 (Sowa et al. 1992), provides a classification schema to describe and design information system architectures. It is a widely used framework to design complex organizational systems, including the design of data warehouses (Sinha 2002). The value of the framework results from its ability to focus on selected aspects of the system without losing a sense of the overall complex system. Its purpose is to provide a basic structure which supports the organization and development of a set of architectural representations of the organization's information systems (Vail 2002). The framework is neutral with regard to the tools or procedures used to obtain the descriptions of the architectural representations (i.e., design artifacts). According to Zachman (1987), an architectural representation is created from the perspectives of different participants developing the structure and is intended to capture the development of business strategy and its linkage to information system strategies.

Drawing upon the discipline of classical architecture and manufacturing, the Zachman framework describes a classification schema based on six dimensions that describe six fundamental questions about the IS systems (i.e., who, what, when, where, why and how) from the perspectives of the planner, owner, designer, builder, and subcontractor of the system (Inmon et al.

1997) (See Table 2.1). The vertical axis describes five stakeholder perspectives, and the horizontal axis provides a classification of the various dimensions.

Each cell in the framework, which constitutes the intersection between a perspective and a dimension, represents a design artifact(s) of the system. Each design artifact becomes a specific blueprint that guides the development of an information system. For instance, consider the cell in the framework, which describes the design artifact(s) constituting the intersection between the location dimension and the designer perspective (See shaded cell in Table 2.1). The location dimension (i.e., the 'where' question) describes a focus on the flows or connections between various components of the system. The designer perspective indicates a conceptual representation of the system components independent of a specific technology (Inmon et al. 1997; Zachman 1987). As a result, the architectural representation described by the intersection between location and the designer perspective indicates how the connections between components would be specified at a conceptual level.

In the context of this study, both the data warehouse reference architecture and the data warehouse system architecture indicate how system components are linked at a conceptual level. As such, both the reference architecture and the system architecture represent the location dimension, which represents the connections or links between system components from different perspectives.

Next, the analyses of the designer perspectives in terms of the locations dimension indicate that the designer view focuses on links between conceptual level system components irrespective of the underlying technology. Since conceptual level links between system components represent the data warehouse architectures of interest to this dissertation, the designer view represents the appropriate perspective on data warehouse architecture for this study.

The reference architecture specifies standard components and relationships between components that form the basis for the data warehouse system architecture. The data warehouse system architecture is the specific architectural representation that results by considering the locations dimension from a designer’s perspective. Because the main focus of this dissertation was on the selection of a data warehouse reference architecture, for clarity purposes, from this point on, the data warehouse reference architecture will be called the data warehouse architecture.

2.4 Data warehouse architecture

The structure that brings all the components of a data warehouse together is known as the data warehouse architecture. The general purpose of the architecture is to provide an overall framework for detailed design and development of a data warehouse (Devlin 1997). Similar to the way that the architecture of a building starts with high-level site drawings and gradually becomes further detailed, the architecture of the data warehouse starts with the high-level architecture and works its way down to levels of greater detail (Inmon et al. 1997). The starting point for data warehouse architecture is the conceptual architecture, which is then expanded and clarified through the logical and physical architecture. Each layer of the warehouse architecture provides an increasing level of detail (Humphries 1999). At the conceptual level, the architecture specifies the components and relationships between components. It shows the major features of the data warehouse without showing the details.

Table 2.1: The Zachman framework

	Data What	**Function** How	**Location** Where	**People** Who	**Time** When	**Motivation** Why
Scope *Planner view*	List of things important to the enterprise	List of Processes the enterprise performs	List of locations where the enterprise operate	List of organizational units	List of business events	List of Business goals /strategies
Enterprise Model *Owner view*	e.g. Semantic model	e.g. Business process model	e.g. Logistics network	e.g. Work flow model	e.g. Business master schedule	e.g. Business plan
System Model *Designer view*	e.g. Data model	e.g. Application architecture	e.g. Distributed system architecture	e.g. Human interface architecture	e.g. Processing structure	e.g. Business rule model
Technology Model *Builder view*	e.g. Physical data model	e.g. System design	e.g. Technical Architecture	e.g. Presentation architecture	e.g. Control structure	e.g. Rule design
Detailed Represen- -tations *Sub-contractor view*	e.g. Data definition	e.g. Detailed Program Design	e.g. Network architecture	e.g. Security architecture	e.g. Timing definitions	e.g. Rule specification
Functioning enterprise	e.g. DATA	e.g. FUNCTION	e.g. NETWORK	e.g. ORGANIZATION	e.g. SCHEDULE	e.g. STRATEGY

2.4.1 Components of data warehouse architecture

At a conceptual level, the typical data warehouse is composed of three major generic components: (1) data sources, (2) data storage, and (3) end user access (See Figure 2.1). The shaded region represents the typical data warehouse architecture at a conceptual level. The components of the shaded region are described next.

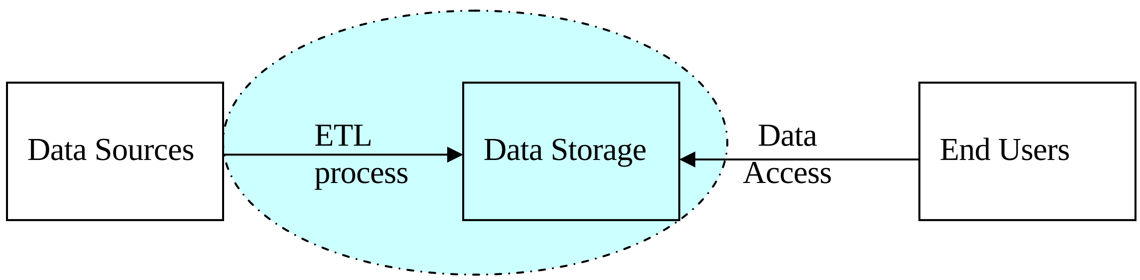


Figure 2.1: The main components of a data warehouse

2.4.1.1 The ETL process

Once data is extracted from various operational systems and external data sources, the data is transformed to a format that is suitable for data analysis. This occurs in a data staging component (Singh 1999). Overall, the data staging component provides an area where the data is cleaned, converted, and integrated to prepare it for storage and use in the data warehouse (Humphries 1999). The data is extracted, transformed, and loaded into the data warehouse. As a result of the ETL process, data in operational data sources are prepared for the warehouse (Singh 1999).

***2.4.1.1.1** Data extraction*

This process involves extracting data from different data sources. The source data is often in legacy systems which utilize a variety of data models (e.g., hierarchical, network, relational) (Finkelstein 1999). While there are vendor tools available for data extraction in the marketplace, some data sources may require in-house programs for data extraction (Richman 1997).

2.4.1.1.2 *Data transformation*

This process involves converting data prior to loading into the data warehouse. Several tasks together form the overall data transformation function. First, the extracted data is cleaned. Cleaning may involve the correction of misspellings or elimination of duplicates in data gathered from multiple data sources. Standardization of data elements forms a large part of data transformation. This task may involve the standardization of data types, field lengths for the same data elements, or semantic standardization (Steinacher 2000). Next, cleaned, standardized data elements are integrated by combining them around a common identifier. A collection of integrated data that is cleaned, standardized, and summarized is created as a result of the data transformation function.

2.4.1.1.3 *Data loading*

The data loading function is composed of two distinct tasks. The first is the initial loading of data into the data warehouse, which involves moving large volumes of data. Once the data warehouse is operational, incremental data feeds occur on an ongoing basis.

2.4.1.2 The data storage component

The data storage for the data warehouse is a separate data repository composed of high quality data. Depending on the data warehouse architecture, the storage component may be composed of a centralized data store, data marts, or both (Hackney 1998). Most data repositories employ relational database management systems while others use multidimensional database management systems.

2.4.2 Differentiating data warehouse architecture and methodology

At present, the data warehouse industry is plagued with debate over the best way to build a data warehouse. A review of the data warehouse practitioner literature indicates that many efforts have been made to resolve this debate (Hackney 2000; Hwang et al. 2002; Koch 1999; Sen et al. 1998). However, the majority of literature does not clarify if the differences are due to variations in data warehouse architecture or methodology (Wells 2003). In the context of system implementation, the term architecture describes the components of a system or provides guidelines on what to build, while methodology covers the procedure of building a system or how to build. The data warehouse practitioner literature indicates that these terms are often used interchangeably in articles on data warehouse design. David Wells (2003) a consultant and director of education at The Data Warehousing Institute (TDWI), described this situation as follows:

> “One of the most frustrating aspects of data warehousing today is architecture and methodology wars. Conference, trade shows, and technical literature are filled with mixed signals. It is difficult to figure out what is the right approach…Also,…sometimes the differences are characterized as differences in architecture…and in other discussions, the debates are seen as differences in methodology…”

As the data warehouse field matured, differences between implementation methodologies have considerably diminished (Eckerson 2002). At present, most major approaches prescribe the use of an incremental development methodology when building a data warehouse. Consequently, a closer examination of the different approaches reveal that actual distinctions emerge through the variations in data warehouse architecture rather than methodology (Hackney 1998). The focus of this research was on data warehouse architecture rather than methodology.

2.5 The major types of data warehouse architectures

A review of the data warehouse literature reveals four widely recognized data warehouse architectures. The following sections describe the major features of each. A summary of the key characteristics of each of the architectures is presented in Table 2.2.

2.5.1 The independent data mart architecture

An independent data mart provides a point solution to solve an immediate business need. The independent data mart (IDM) architecture involves multiple data marts that were developed and used relatively independent of one another. It can be developed using either multidimensional or relational data structures. Independent data marts contain both detailed data as well as summarized data to satisfy the decision-support needs of either a business unit or business function. Thus, this architecture has a scope limited by the business requirements of a functional unit or work group that independently manages the development of the independent mart (Winsberg 1996). The independent data mart architecture focuses on the short-term needs of the functional unit and presents a seemingly low cost alternative to building a data warehouse. It does not consider the information requirements of the overall organization (Armstrong 1996).

The IDM architecture is seen as an attractive option to solve decision-support needs because it takes less time to implement than an enterprise data warehouse and initially costs less (Hoss 2002). As a result, independent marts are implemented by different functional units and work groups within the organization. However, none of these independent data mart initiatives are coordinated and each independent mart is built with its own data definitions and dimensions. Unlike the DBA architecture (described later), independent data marts are not based on conformed dimensions. As a result, the use of this architecture results in an organizational environment composed of multiple operational systems feeding multiple, nonintegrated data marts that are often overlapping in data content (Kelly 1997). Figure 2.2 presents an illustration of the IDM architecture. In the long run, independent marts can be more costly than other data warehouse solutions as they require duplicated development and maintenance efforts as well as duplicated software and hardware infrastructure (Hoss 2002).

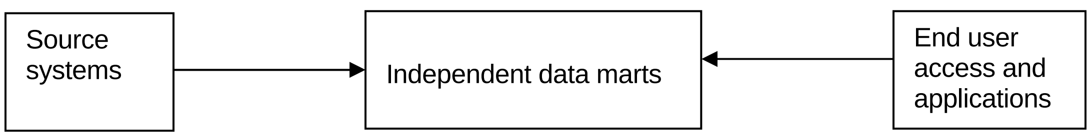


Figure 2.2: Independent data mart architecture

Though the independent data mart architecture provides quick results, ease of implementation and low initial development costs, it eventually yields a disintegrated decision-support environment composed of information silos (Hwang et al. 2002).

2.5.2 Data mart bus architecture

The major proponent of the data mart bus architecture (DBA) is Ralph Kimball, who also refers to this architecture as the data warehouse bus architecture (Eckerson 2002). The DBA architecture is compromised of a staging area and architected data marts. Data extracted from source systems are subjected to transformation in the staging area and fed to an architected data mart. The architected marts contain atomic level data including historical data in addition to summarized data.

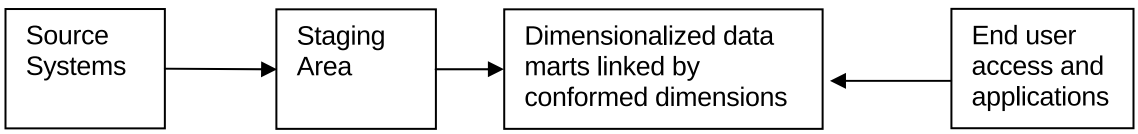


Figure 2.4: Data mart bus architecture

This architecture guides the incremental development of data marts with the view of integrating as they are developed. It is characterized by the step-by-step development of subject area data marts that fit an overall enterprise wide architecture (Hackney 1998). The first mart is built as a proof of concept with a scope limited to several subject areas needed by one or few functional areas. It is built to support a specific business process such as orders, deliveries, customer calls, or billing. Next, additional marts are added according to a master suite of conformed dimensions and standardized fact definitions (Watson 2001) (See Figure 2.4).

The creation of a master suite of common dimensions and fact tables for the entire organization involves conducting an enterprise-level analysis of information requirements. However, the time frame for the enterprise-level requirements analysis is considerably short than required for the EDW architecture (Koch 1999). As a result of this analysis, involving cross-department participation and agreement, consistent data definitions and common dimensions are created. Upon creating the set of conformed dimensions, subject area data marts are prioritized and built on a ‘as needed’ basis using the master suite of dimensions and facts.

Unlike the normalized data structure applied in the EDW architecture, the dominant data structures used in architected marts are dimensional structures using either relational star schema and/or multidimensional data cube. According to Kimball (2003), dimensional structures focus on rapid query performance and enabling users to perform complex ad hoc queries based on a data structure that is intuitive and easy to understand. While relational star schema structures enable the analysis of large volumes of data, multidimensional data cubes enable complex data comparisons and calculations on small data sets. Generally, data structures using star schema based on mature relational database technology is more predominant than the multidimensional data cube (Kimball 2003). In fact, multidimensional data cubes are often created from data existing in relational star schema data structures.

A relational star schema provides a non-normalized dimensional structure, which looks similar to a star that is made up of a fact table in the center of the star and dimension tables as the points of the star. An illustration of a star schema for a grocery store is presented in Figure 2.5. The fact table stores measurements, which are numeric such as dollar sales or unit sales. The dimension tables store descriptions of the context surrounding the measurements, which are textual such as customer information and product information. A dimension is considered a conformed dimension when the same dimensions across subject area data marts are defined exactly the same way in terms of the values of the keys and the attributes (Kimball 1999). For instance, if a customer dimension is unconformed, then various data marts within the organization will have incompatible customer categorizations and labels. As a result, data in the various data marts cannot be combined in queries.

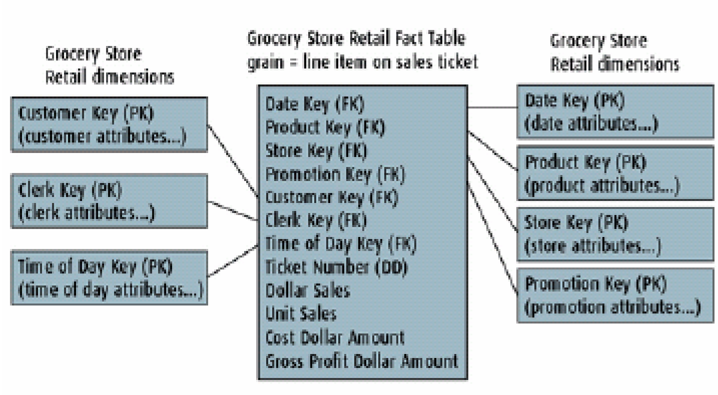


Figure 2.5: A star schema for a grocery store adopted from Kimball (2003)

While this architecture does not create a physically separate data warehouse, each architected mart is seen as part of a data warehouse, similar to a pie wedge that together with other wedges form a whole pie (Ponniah 2002). Thus, a data warehouse is a conformed union of all the architected data marts. As a result, this architecture provides end users with the ability to use simple queries to conduct data analyses that require combining several subject areas housed in different architected data marts.

According to Kimball (2003), the major advantage of this architecture lies in its ability to divide the task of creating a data warehouse into smaller subprojects that result in the gradual creation of the final solution. It enables the autonomous development of marts by loosely coupled data mart teams to build an overall distributed system. However, the major disadvantage of this architecture is that it is challenging to enforce the master suite of conformed dimensions on the various data mart implementation efforts within the organization (Hackney 2000).

2.5.3 The enterprise data warehouse architecture

The enterprise data warehouse architecture (EDW) was first introduced at IBM and by Bill Inmon, who is widely recognized as the 'father of data warehousing.' (Winsberg 1996). This architecture provides an integrated central data store for decision support, either through direct access to the data warehouse or by distributing data to dependent data marts.

In its purest form, the EDW architecture is composed of two main components: (1) a data staging area and/or a ODS, and (2) a central data store. However, as most enterprise data warehouses mature, they are supplemented with dependent data marts built to support the needs of specific functional units or groups (Gardner 1998). The central data store is a database that stores large amounts of detailed data (i.e., individual atomic transactions) and lightly summarized data that spans multiple subject business areas (Inmon 1992). The centralized data repository is fed with data extracted from legacy sources and transformed in the data staging area as well as data from the ODS. Next, the dependent data marts are loaded with data from the data warehouse (See Figure 2.3).

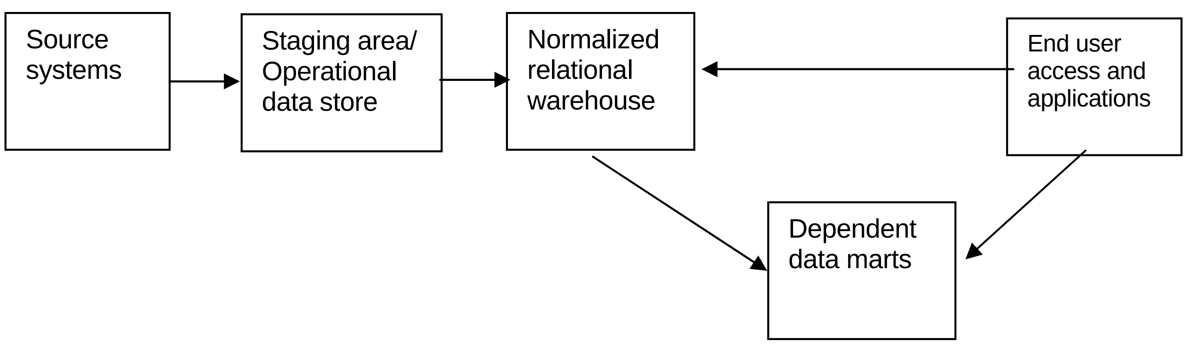


Figure 2.3: Enterprise data warehouse architecture

Implementing an enterprise data warehouse architecture requires gaining an enterprise-oriented view. This enterprise-oriented view of data is gained through an extensive planning and enterprise level requirements analysis involving the cooperation from all affected business areas within the organization. The centralized data store has a relational data structure. The data is stored in third normal form to minimize data redundancy and produce a stable database design that can adapt to changes to business processes (Sperley 1999). The enterprise data warehouse is built in an iterative manner, business area by business area (Gentry 2002). It leads to the gradual creation of a centralized data repository with common data standards and integrated data elements for the entire organization.

Overtime, accessing data directly from the centralized data store becomes increasingly difficult as its design evolves to efficiently integrate and manage large volumes of quality data (Inmon et al. 2001). In response, dependent data marts are created that source their data from the data warehouse. These dependent data marts maintain “the single version of the truth.”

A dependent data mart is considered a collection of data tailored to the decision-support needs of a particular functional business unit. Alternatively, it can be a resource shared by multiple business units, which have common analytical needs (e.g., profitability analysis). Generally, dependent data marts are developed due to user needs for a dimensional view of data and to enable faster query processing. Overall, a dependent data mart is defined as a subset of a data warehouse containing a small amount of detailed data and a generous portion of summarized data. It contains a limited amount of historical data, which is significantly less than the volume supported by the data warehouse. Unlike the normalized data design of the data warehouse, the data in dependent data marts is normally organized in star schema with a dimensional data structure. The flow of data from the data warehouse to the dependent data marts is unidirectional. It occurs on a ‘as needed’ basis as demands are made for atomic or summarized data.

The main advantage of the EDW architecture lies in its ability to provide an integrated enterprise view of the data (Hackney 1998). However, this architecture solution requires a large upfront investment, extensive planning effort, and long-term buy in on the project from the entire organization (Shin 2002).

2.5.4 The federated data warehouse architecture

The federated data warehouse architecture (FED) is an overall systems architecture that provides a framework to integrate data in multiple data warehouses, data marts, and legacy systems. Douglas Hackney, its major proponent, calls the federated data warehouse architecture "an architecture of architectures" (Hackney 1998). The federated architecture presents a global perspective of an organization's information needs and serves to integrate lower-level architectures (Hackney 2002).

The federated architecture is recommended as the only viable solution to deal with the challenges of combining disparate decision-support architectures that already exist within the organization. It is also considered the only alternative that will meet new data silos that will continue to proliferate despite the organization's best efforts to enforce standards and architectural design (Eckerson 2002). Implementing a federated data warehouse is an enterprise-level effort that focuses on creating the best possible solution within the political, cultural, and implementation realities of the organization.

The FED architecture is distinguished by the characteristic of sharing key metrics and measures across multiple data warehouses and/or data marts (Hackney 2000). Creating a federated architecture begins with the identification of the data, dimensions, and metrics that are of highest value to the organization. Upon identifying the critical data elements to be integrated, consensus on protocols and standards that enable data integration is gained from the stakeholders of existing systems. The documentation and the agreed upon semantic and business rules compromise the federated architecture (Hackney 2002).

However, the architecture does not mandate a standard process or generic components that are considered necessary to develop a federated architecture (Eckerson 2002). Instead, the best solution for federation is based on the state of the existing heterogeneous environment and the common data standards determined by the stakeholders of existing systems. For example, creating a federated data warehouse may require a common staging area between data marts to load data as well as a direct feed from a departmental data warehouse to the federated data warehouse (See Figure 2.7). Thus, the political, cultural, and implementation realities of the organization dictate the best way to accomplish integration.

Ideally, the FED architecture can ensure a consistent, unique version of the truth throughout the organization by sharing key metrics and measures across systems. In reality, the extent of data integration achieved depends on the quality of data in existing systems and the difficulty in combining data sources differentiated by data structures as well as hardware and software infrastructure. However, the biggest challenge for this architecture lies in achieving common data standards consensus between stakeholders of the component data warehouse and data mart systems.

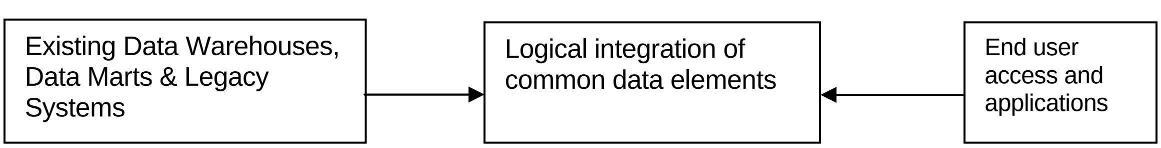


Figure 2.7: The federated data warehouse architecture

Table 2.2: Characteristics of data warehouse architectures

	EDW	DBA	IDM	FED
Components	Common data staging area and/or ODS, central data store, dependent data marts (optional)	Data staging area, architected data marts	Data staging areas, independent data marts	Determinant of the nature of existing systems and consensus between stakeholders
Methodology	Incremental development	Incremental development	Independent marts are created as needed without a structured development methodology.	Incremental development
Data Structure	Centralized data store: Normalized relational data structure Dependent data marts: dimensional data structure	Dimensional data structure	Normalized relational data structure or dimensional data structure	
Data granularity	Centralized data store: atomic level data Dependent data marts: summary data	Atomic level and summarized data	Atomic level and summarized data	Summarized level data
Focus	Centralized data store: Enterprise focus Dependent data marts: function specific	Each architected mart: function specific	Each independent mart: function specific	Enterprise focus

2.5.5 Other Architectures

In addition to the architectures described above, the data warehouse literature mentions several other variations and architecture alternatives. A recent popular alternative described in the literature is the Hybrid approach described by Pieter Mimno (Mimno 2001). However, closer examination of the Hybrid approach reveals that it describes a distinct methodology to build a data warehouse according to the data mart bus architecture (Eckerson 2002). As a result, it was not included in this study.

2.6 Conclusion

This chapter described architecture, IS architecture, IT infrastructure architecture, and data warehouse architecture. Next it discussed each of the four main data warehouse architectures. The next chapter presents the review of theory and literature relevant to the selection factors and success factors.

CHAPTER 3 – LITERATURE REVIEW

3.1 Introduction

This chapter first presents the results of the expert interviews, which identified the most salient factors that affect architecture selection and success. Next, using the results of the interviews as guidance, relevant theory and previous literature are explored to describe factors that potentially affect data warehouse architecture selection. Next, IS success is discussed to identify the success factors most relevant to this study.

Whenever possible, specific hypotheses were developed that describe the influence of selection factors on the selection of a particular architecture based on theory and existing research. Hypotheses were created only when there was strong support for the statements from literature and theory. To the extent that theory and literature did not provide information about a clear relationship between a selection factor and data warehouse architecture, hypotheses were not stated. A similar approach was followed to develop hypotheses for architecture success.

3.2 Interviews with data warehouse experts

In order to ensure that the research models developed for the dissertation matches the realities of the world, a series of interviews was conducted with leading experts in data warehousing (The research design for this first phase is described in Chapter 5 –Methodology). The results of this first phase of the dissertation provided a list of factors that are considered salient to data warehouse architecture selection and architecture success by the panel of data warehouse experts.

Experts were selected on the basis of being recognized authorities and spokespersons for major architectures, leading consultants, highly recognized data warehouse managers, and Fellows of The Data Warehousing Institute. Appendix A describes the panel of experts interviewed in phase one.

Each expert was contacted and asked to participate as part of the panel of experts in the study. Each expert contacted agreed to participate in the study, perhaps suggesting the importance of the study to industry. As a result, ten phone interviews were conducted and each interview typically lasting between 30 minutes to one hour. Each expert was asked two main questions (1) what factors affect the selection of data warehouse architecture? and (2) how do you measure the success of data warehouse architectures? With the permission of the experts, each interview was tape recorded. Each tape recording was analyzed for recurring themes that identified factors that affect architecture selection and architecture success. Content analysis of the interviews surfaced ten selection factors that affect architecture selection and ten success factors that measure data warehouse architecture success.

Next, a document summarizing the factors based on the interviews, literature, and theory were sent to the experts to be reviewed. They were asked to offer their thoughts on the interpretation of the interview data. The experts provided useful feedback, which led to useful additions and changes. The summary of the resulting factors is presented in Table 3.1 and Table 3.2.

Table 3.1: Summary of architecture selection factors from expert interviews

Selection Factors	Rob Armstrong	Karolyn Duncan	Wayne Eckerson	Jane Griffin	Douglas Hackney	Bill Inman	Jim Revak	Don Stoller	Ron Swift	Jim Thomann
Horizontal information interdependence	X		X	X		X			X	
Vertical information flow					X	X			X	
Urgency	X		X	X	X		X	X	X	X
Task routineness		X	X				X			X
Compatibility with existing system		X			X	X		X	X	X
View of the data warehouse				X		X				X
Resource availability	X	X	X	X	X			X		
Perceived ability of the IT staff			X	X	X	X	X	X	X	X
Source of sponsorship	X	X	X		X	X	X	X		X
Expert influence		X	X					X		X

Table 3.2: Summary of architecture success factors from expert interviews

Success Factor	Rob Armstrong	Karolyn Duncan	Wayne Eckerson	Jane Griffin	Douglas Hackney	Bill Inman	Jim Revak	Don Stoller	Ron Swift	Jim Thomann
Flexibility		X	X	X		X		X		X
Scalability				X		X		X	X	X
Integration	X		X		X	X	X	X	X	
Data granularity	X			X		X				
Accuracy	X		X	X		X	X	X		
Completeness		X		X				X		
Consistency	X	X				X		X		X
Individual Impact		X		X					X	
Organizational Impact	X								X	
Development cost	X		X	X	X		X	X	X	X
Development time	X		X	X	X		X	X	X	X

3.3 Organizational selection factors

The discussion of the data warehouse architectures in the previous chapter suggests that different architectures are more appropriate for different organizational situations. The analyses suggest that each of the architectures should be implemented under differing organizational circumstances. For instance, while the enterprise data warehouse architecture provides corporate level data integration, the independent mart architecture provides a point solution for a specific functional area business need. This example illustrates that business need affects the selection of a data warehouse architecture. In addition to the business need, the practitioner literature provides descriptions of other organizational factors that may be relevant to the selection of a data warehouse architecture (Eckerson 2002; Hackney 1998; Joshi et al. 1999). Several streams of academic literature also provide insights into the organizational circumstances surrounding architecture selection.

Having decided to build a data warehouse, the selection of a data warehouse architecture is another critical decision requiring the attention of the organization. Organizational decision-making theories provide a lens to look at the factors affecting the architecture selection decision. The literature on organizational decision making describes two competing theories for effective decision making in organizations (Langley et al. 1995; Pfeffer 1981). These competing theories are rooted in the rational analytical and social political schools of organizational theory.

The proponents of the rational school argue for the use of comprehensive information processing to attain an organizational outcome-maximizing goal (Goodhue et al. 1992; Matheson 1998). Organizational information processing theories (OIPT), part of the rational analytical school of thought, provide some insight into the factors affecting data warehouse architecture selection. Factors that assess the information processing requirements and information processing capacities described in OIPT may describe the characteristics of the organizational context relevant to the selection of a data warehouse architecture (Tushman et al. 1978).

The social political school of organizational theory views decision making as a process of negotiation and coalition building in which multiple ambiguous goals exist (Eisenhardt et al. 1988). This view questions the assumption that an overarching organizational goal accomplishment objective influences organizational action. In the context of data warehouse architecture selection, political theories challenge the notion that an overall company goal determines the architecture selection decision. The rational information processing view and the social political view will be further described later. A review of past IS research indicates that contingency literature and innovation adoption literature may also provide some insight into other factors that affect data warehouse architecture selection.

3.3.1 Contingency research

The theme of contingency literature in information systems is that the best way to organize the information system structure within a given organization is contingent upon internal and external factors specific to that organization. Some of the earliest studies in this area (Ein dor et al. 1978, 1982) examined the influence of organizational context variables such as size, structure, industry dynamism, and psychological view of IS on structural characteristics. More recent research has examined IS distribution and design in multinational organizational contexts as well (Grover et al. 1996; Tractinsky et al. 1995). For instance, Grover and Segars (1996) examined the influence of contextual variables such as firm size, role of IT, and economic sector on the distribution of IS in organizations from three countries.

Past research on contingency fit describes several factors that appear to have an impact on the design of information systems in organization. These include industry dynamism (e.g., Grover et al. 1996; Bajwa et al 1998), firm size (e.g., Ahituv et al. 1989; Ein-Dor et al. 1982) strategy (e.g., Henderson et al. 1994), locus of decision making (e.g., Ein-Dor et al. 1982, Grover et al. 1996) and business unit autonomy (e.g., King 1983).

The literature review revealed a few research articles that considered the impact of contextual factors on IS architecture. Early research in this area discuses the impact of organizational context on IS design. For instance, in an effort to clarify linkages among information systems designs and organizational context, Leifer (1988) described how to fit IS structure to different organizational configurations. In 1991, Allen and Boynton (1991) discussed two architectural solutions for organizations differentiated by strategy and other organizational characteristics.

More recently, studies have focused on the influence of organizational context on specific technologies. In selecting client server computing architectures to match organizational needs, Anandarajan and Arinze (1998) posited that there was no one appropriate client server architecture alternative that fits all organizational contexts. Instead, they suggested that the choice of different client server architectures depends on the match between the level of information processing capacity of different architectures and the nature of task uncertainty. A study looking at factors that affect EIS architecture found a significant relationship only between organizational external pressures measured as dynamism, heterogeneity, and hostility and EIS architectures (Bajwa et al. 1998). Research has also described the impact of contextual variables such as task complexity and information processing intensity in selecting low cost information system architecture alternatives (Francalanci et al. 1999).

3.3.2 Adoption research

While less applicable than contingency theory, adoption theory also provides insight into the architecture selection process. The IT adoption literature describes many studies that investigate the adoption of IT innovations at the individual and organization levels (Ryan et al. 2001). At the organizational level, the focus is on identifying the circumstances under which organizations decide to invest in an IT innovation (Bajaj 2000). While this research stream may not be directly applicable to architecture selection, it presents a possible source of factors that may affect the adoption of a particular data warehouse architecture.

Past research reveals numerous studies that have looked at the adoption of IS innovations (e.g., Fichman 1992; Kwon et al. 1987). Literature summaries identify several factors that can influence the organizational adoption of an IS innovation, including environmental dynamism, organizational size, and centralization of decision making. Kwon and Zmud (1987) categorized the influencing factors in terms of organizational factors, environmental factors, and individual factors. The majority of organizational IT adoption studies have focused on the adoption of IT applications. Recently, there has been an interest in research that examines factors that affect the adoption of complex IT innovations (Chau et al. 1997; Ryan et al. 2001). They reveal factors that may be applicable to data warehouse architecture selection such as the impact of the past experience of the IT staff, and existing system compatibility on the selection of data warehouse architecture.

3.3.3 The rational perspective and information processing theories

The rational perspective assumes the existence of a single set of objective goals for the organization. Furthermore, it presumes that all stakeholders within the organization are motivated and work towards the accomplishment of these goals (Jasperson et al 2003). Thus, organizations operating under principles of rationality evaluate benefits and costs at the organizational level and its members act in a manner that maximizes organizational-level benefits while minimizing the costs (Goodhue et al 1992).

In the context of IS, the rational view may describe how organizations strive to guide IT decisions in a manner that maximizes organizational benefits (Tractinsky et al. 1995). In addition, the rational perspective promotes the accomplishment of IS design and development tasks in a manner that minimizes implementation costs to the organization (Cavaye et al. 1996). Organizational information processing theories (OIPT), a set of theories that embody the rational perspective, provide a lens to consider the impact of costs and benefits on information system design.

The primary purpose of OIPT is to understand and determine organizational design using the concepts of uncertainty and information processing. It proposes that variations in the degree of uncertainty faced by organizations drive the need for varying levels of information processing. Task uncertainty, the central concept underlying OIPT, is defined as the difference between the amount of information required to perform an organizational task and the amount of information already possessed by the organization (Galbraith 1977, p.36). Tushman and Nadler (1978) described uncertainty at the subunit level of analysis and specified three distinct sources of uncertainty: complex or non-routine tasks, unstable subunit task environment, and interdependence between subunits. Furthermore, as each subunit within an organization has a different task environment and task characteristics, the information processing requirements (i.e., the information that must be acquired and processed) in each subunit may vary. Thus, the most appropriate mechanisms for providing information processing capacity in each unit might be different.

In 1973, Galbraith (1973) described the broad information processing mechanisms available to organizations. According to him, when faced with lower levels of task uncertainty (i.e., low levels of information requirements), organizations could use simple coordination mechanisms such as standard operating procedures and/or hierarchical referral to resolve it. However, when confronted with greater levels of uncertainty that cannot be handled by simple coordination methods, organizations must opt to use some combination of four more complex mechanisms to reduce uncertainty.

Two of these complex mechanisms increase the organization's information processing capacity in order to satisfy information needs. They are facilitating lateral relations or implementing an information system. The other two (e.g., by creating slack or establishing self-contained units) reduce uncertainty by decreasing the need to process information.

In this context, a data warehouse is an information system. As such, according to Galbraith (1973), it may reduce uncertainty by increasing the information processing capacity of the organization. The architectural design of a system ties organizational information requirements to the physical components necessary to satisfy those information processing requirements (Francalanci et al. 1999). As such, organizational information requirements (i.e., sources of uncertainty) affect data warehouse architecture, which influences the information processing capacity of the organization.

Consequently, contrary to Galbraith's conceptualization of an information system as a mechanism to *increase* information processing, depending on the architecture chosen to meet information needs, different data warehouse architectures may *increase or decrease* the information processing capacity. For instance, the choice of an IDM architecture will guide the creation of a system that is limited in scope to a specific business unit or functional unit. Thus, it may facilitate the creation of self-contained units and limit the information processing capacity among subunits. As such, depending on if an organization chooses to reduce sources of uncertainty by either increasing or decreasing information processing, their choice of architecture may differ. As such, OIPT provides a lens to consider factors that affect architecture selection.

Depending on the architecture chosen for implementation, different data warehouse architectures can potentially facilitate the creation of all four complex mechanisms to reduce uncertainty described by Galbraith (1973).

3.3.4 Social political theories

The influence of power on organizations has been of central interest in academia. It is a complex concept that is difficult to define and the subject of much debate (Gandz et al. 1980). Power involves one or more groups having control over resources that others desire (Pfeffer 1981). It is reflected in the group or social actors' capability to use owned resources to overcome resistance and accomplish desired goals.

The majority of research on power takes either a behavioral or a structural perspective (Cavaye et al. 1996). Structural theorists focus on sources of power. They identify factors that enable one or more groups to acquire or possess power. The literature identifies many sources of power such as the ownership of scarce resources vital to the operation of the organization, or the possession of skills necessary for a critical function within an organization (e.g., Pfeffer 1981). Organizational position is also considered a source of power. For instance, power possessed by individuals or groups within the organization can be due to their hierarchical position.

The behavioral perspective studies the actions of individuals to obtain their goals. It focuses on the exercise of power or political activity (Eisenhardt et al. 1988). The use of power or engaging in politics describes ways by which groups influence actions and obtain favorable outcomes by virtue of their power. For instance, by the skillful use of language, individuals may influence organizational decision making and alter existing beliefs (Pfeffer 1981). Some authors recognize conflict as the source of politics in organizations. Other researchers argue that while conflict causes politics, it is not a necessary condition; whereas, power imbalance is considered crucial for political activity (Eisenhardt et al. 1988).

Within IS, power is discussed in terms of political activity that affects different aspects of system implementation. An information system can change the balance of power between groups in organizations. Implementation of an information system can threaten the existing distribution of authority, challenge existing ownership of information, and might be simultaneously perceived as a threat or as an opportunity by different groups within the organization (Markus 1983). Consequently, groups threatened by the system may resist or hinder system development. As such, the process of developing an information system can be investigated as a series of political activities that may shape the final implementation of the information system as well as alter existing organizational structures. Consequently, data warehouse architecture selection can be perceived as a social political change process in which organization members motivated by self interest influence the choice of architecture.

This view undermines the rational perspective of IS implementation which consists of a logical sequence of steps geared towards accomplishment of an overall organizational goal. Organizational politics has been identified as a theory that provides insight into situations that cannot be readily explained by a deterministic rational model (Robey et al. 1999). In the context of IS, it adds a more realistic picture of the organizational circumstances under study that considers the impacts of social interactions on IS development.

3.3.5 Summary of existing theory and research

While past research provides some indication of the factors that impact data warehouse architecture, in order to gain a richer understanding of these factors, the data warehouse literature was extensively reviewed and insights from data warehousing experts were gained. Consequently, the most salient factors that affect data warehouse architecture selection were identified through the past IS literature, the data warehouse literature, and through interviews with data warehouse experts. The following section describes the selection factors that affect the choice of data warehouse architecture.

3.4.1 Information interdependence

The importance of managing interdependence has been widely recognized as an important management issue addressed by information systems in organizations (Rockart et al. 1989). Interdependence refers to the extent to which organizational units need to exchange information or material to accomplish their tasks (Thompson 1967). It occurs where the information or actions taken by one unit affect the actions and work outcomes of other units (Andres et al. 2001).

Organizational information processing theories identify interdependence as a source of uncertainty that affects the information needs in organizations (Galbraith 1974). From an information processing stand point, greater interdependence between units requires greater coordination/communication mechanisms to promote information sharing (Tushman et al. 1978; Goodhue et al. 1992). Implementing an information system is one of two mechanisms that increase the information processing capacity of organizations.

3.4.1.1.1 *Hypothesis derived from literature and interviews*

A data warehouse can increase the information processing capacity of an organization by creating a source of high quality, integrated data for decision support (Devlin 1997). The architecture of a data warehouse influences the information processing capacity provided by the data warehouse. As such, the information processing view suggests that the extent of information interdependence between business units determines the need for different levels of information processing capacity, which in turn determines the data warehouse architecture most appropriate for an organization.

For instance, an organization that has business units that are highly dependent on information from other units will require an organization-wide data warehouse solution to support information sharing across units. An enterprise data warehouse architecture provides guidance to implement a solution with a global or organizational scope (Inmon 2001). As such, an organization with high interdependence would benefit from the implementation of an EDW architecture solution. In contrast, an independent data mart architecture is often implemented to meet decision-support needs in a specific functional unit (Armstrong 1996). An organization which does not require high information sharing may implement an IDM architecture to satisfy local level functional needs.

Some of the experts also provided insight into the relationship between interdependence and data warehouse architecture during the interviews. For instance, one of the experts, Jane Griffin, stated that the need to share information across units within the entire enterprise often leads organizations to select an EDW architecture compared to IDM architecture solutions.

While the relationship between information interdependence and EDW or IDM architecture can be clearly stated, it is more challenging to determine the impact of interdependence on the rest of the architectures. The hypothesis that describes the specific relationship between interdependence and EDW or IDM architectures is:

Hypothesis 1: *Organizations with high interdependence are more likely to select an EDW architecture than an IDM architecture.*

3.4.2 Vertical information flow

Information needs within an organization may require information to be communicated up the organizational hierarchy (McCann et al. 1979). In analyzing case studies on data management practices within firms, Goodhue et al (1988) discovered that making data accessible for senior executive decision making was one of the key reasons for data management practices within organizations. The extent of information made accessible to higher levels of the hierarchy varies according to the degree of centralization in the organization (Ein-Dor et al. 1982). At higher levels of centralization, most strategic decision making takes place at upper levels of the organization (King 1983). As a result, when an organization is highly centralized, upper management requires more information from lower levels of the organization.

In addition, recent government regulations on ethical business practices and reporting contained in the Sarbanes and Oxley (SOX) Act of 2002 has increased the importance and need for accurate information from lower levels of the organization (Badami 2003). The Sox Act has made it necessary for the upper management of publicly traded organizations to have high visibility and access to manage information and reporting practices throughout the organization (Easton 2004) in order to certify the integrity of corporate information to the Securities and Exchange Commission.

In order to access information from lower levels of the organizations, senior management requires an information processing mechanism that facilitates the flow of information from the lower levels to the higher levels of the organization. In describing the option of implementing an information system to increase information processing, Galbraith (1974), introduced “vertical” information systems that provide information vertically up the hierarchy to managers responsible for all organizational units. Volonino et al (1995) described facilitating the vertical flow of information along the organizational hierarchy as one of the three uses of an executive information systems. Tractinsky et al (1995) revealed that the extent of accessibility of business unit level information at the corporate management level was an important factors affecting the IT application deployment decision.

3.4.2.1.1 *Hypothesis derived from literature and interviews*

The information processing perspective suggests that the extent of business unit level information needed by upper management influences the data warehouse architecture implemented to meet an organization’s information processing needs. While certain data warehouse architectures promote higher levels of vertical information flow, others do not (Hackney 2001). For instance, an organization that requires the ability to access a large quantity of integrated data from business units to meet the decision-support needs of upper management, would require an architecture that facilitates high vertical information flow. An EDW architecture guides the implementation of a data warehouse that meets upper management’s need for large amounts of business unit level data (Inmon 2001). In contrast, an IDM architecture is deployed to meet the needs of an individual business unit, and does not facilitate the flow of information from lower levels of the organization to satisfy the needs of managers in higher levels of the organizational hierarchy.

Several experts also discussed the impact of vertical information flow on data warehouse architecture during the interviews. For instance, one of the experts, Wayne Eckerson, stated that organizations with centralized organizational structures implement EDW architectures in order to gain access to local business unit level data.

While the exact relationship between vertical information flow and the rest of the architectures is not as clear, the relationship between vertical information flow and EDW or IDM architecture can be stated in the following hypothesis:

Hypothesis 2: *Organizations with high vertical information flow are more likely to select an EDW architecture than an IDM architecture.*

3.4.3 Urgency

At present, the dynamics of the business environment are changing dramatically and the information required by decision makers has become a critical and urgent need (Bontempo et al. 1998). Information processing theories describe the dynamic environment as a major source of uncertainty that creates the need for mechanisms that increase information processing (Tushman et al. 1978). Implementing information systems is one mechanism that increases the information processing capacity of organizations (Galbraith 1973). Thus, the information processing perspective indicates that the greater the turbulence of the environment that organizations operate in, the more urgent the need for accurate information. Consequently, business environments with greater turbulence will experience greater urgency for information systems that provide accurate information.

Past research indicates that there is a strong need for advanced information technologies that provide accurate information in dynamic environments (e.g., Grover et al 1996). For instance, research in EIS indicates that pressure due to environmental turbulence causes organizations to rapidly implement an EIS to satisfy decision-making needs (Bajwa et al. 1998; Volonino et al. 1995). During the course of studying IT functions in companies across cultures, Rockart et al (1996) discovered that long wait times for systems to become operational is unacceptable in dynamic markets. The researchers indicate that the speed of implementation is a primary issue for organizations in volatile markets. Furthermore, while examining the influence of different organizational structures on information system architectures, Allen et al (1991) claimed that rapid implementation of systems is a necessity in order to be effective in dynamic market environments. Subsequently, they concluded that a dispersed information system architecture rather than a centralized architecture is key to the swift implementation of new systems and rapid deployment of technology.

In the data warehouse arena, speed of project completion has become a necessity. At present, most organizations demand the completion of large projects in small time frames (Griffin 2001). As speed to market has become a necessity in dynamic environments, some data warehouse vendors have begun to stress speed to market as a feature of their new products to customer bases (Hackney 2002).

3.4.3.1.1 *Hypothesis derived from literature and interviews*

In the context of this study, when the environment dictates greater urgency for quality information, it creates an urgent need for a data warehouse solution to meet business needs. Consequently, the extent of urgency for a data warehouse affects the choice of a data warehouse architecture. While certain data warehouse architectures enable rapid implementation of a data warehouse, others take longer to implement. As such, the extent of urgency for a data warehouse limits the choice of data warehouse architectures that can be implemented to satisfy information needs. For instance, Sherwin-Williams urgently needed a data warehouse to meet the needs of business users, and they chose an architecture that could be implemented rapidly and deliver quick returns (Watson et al. 2001). The development team at Sherwin-Williams chose to implement an DBA architecture to facilitate the rapid deployment of data marts. In addition, organizations often implement an IDM architecture as it can be implemented quickly to meet urgent business needs (Hwang et al. 2002). In contrast, the EDW architecture is often described as an architecture which requires a relatively large upfront investment in time to identify the enterprise oriented view of the data (e.g., Hackney 2002a). As such, the selection of an EDW architecture does not facilitate rapid implementation of a data warehouse solution.

During the interviews, several experts indicated that urgency was a critical factor that affected the choice of data warehouse architecture. Specifically, Doug Hackney, one of the experts, claimed that the need to build a data warehouse solution to satisfy an immediate "business pain" leads most organizations to select an IDM architecture compared to an EDW architecture solution.

While the relationship between urgency and FED is less clear, the hypothesis describing the relationship between urgency and the rest of the architectures is described below:

Hypothesis 3: *Organizations with high urgency are more likely to select a DBA or an IDM architecture than an EDW architecture.*

3.4.4 Task routineness

Organizational tasks have been described as actions carried out to turn inputs into outputs (Goodhue 1995). Past IS research has focused on many characteristics of tasks that indicate greater or less need for information systems, such as task complexity (Specht 1986), task analyzability (e.g., Anandarajan et al 1998), and task routineness (Goodhue 1995). Campbell (1988) claims that tasks that are ill structured and ambiguous create a need for high information processing to accomplish tasks.

Information processing theory provides a lens to examine the relationship between information systems and task characteristics. According to Galbraith (1973), tasks differ in their amount of predictability. In other words, tasks vary in the amount of uncertainty that a user must experience during task execution. While routine, structured tasks require less information processing for task accomplishment, tasks that are not well planned and are not well understood are associated with greater uncertainty (Tushman et al. 1978). Goodhue (1995) describes tasks characterized by the analysis of ad hoc situations in new ways as non-routine tasks; whereas, tasks that are structured and repetitive as routine tasks. Information processing theory suggests that the more non-routine a task is, the more information processing capacity is needed to reduce uncertainty associated with non-routineness. Specht (1986) surveyed users in five public organizations and found that unstructured, non-routine tasks require more information processing than simple, routine tasks. In data warehousing, when tasks (i.e., ad hoc query analysis) involve novel unstructured ways of analyzing data, the non-routineness of tasks in data warehousing environments increase (McCarthy et al. 2000).

3.4.4.1.1 *Hypothesis derived from literature and interviews*

When potential data warehouse users require the ability to do tasks characterized by high non-routineness, the information processing demands made on the data warehouse will be high. The information processing capacity provided by data warehouses differ based on their underlying data warehouse architectures. Some architectures are better able to support information processing requirements dictated by task routineness than others.

The greater the prior understanding of the steps and data required for query analyses is, the greater the task routineness. As such, when tasks are more routine, it is easier to identify the key metrics and measures necessary for task accomplishment. If the metrics needed can be easily identified, then a federated architecture would support the organization's needs as the FED architecture solution is characterized by combining key metrics needed to meet decision support needs. However, when the task requirements are unstructured and the data and steps needed for task accomplishment are not well known, users need the ability to run queries in novel ways. The FED architecture only provides limited support for novel and unstructured task requirements as it furnishes only known metrics for analysis purposes. In contrast, the EDW architecture provides data in various levels of detail from data sources across the organization. As such, the EDW architecture is better able support unstructured non-routine tasks.

During the interviews, several experts also indicated that the nature of tasks affected the selection of data warehouse architecture. For instance, Jim Thomann stated that the need to accomplish complex, ad hoc data analysis may require the implementation of an EDW architecture solution. He further indicated that if task requirements were routine, and predictable, organizations often move towards less complex architecture solutions than an EDW architecture solution.

The hypothesis that describes the relationship between task routineness and EDW or FED architecture is:

> ***Hypothesis 4:*** *Organizations with tasks that are relatively more routine are more likely to select a FED architecture than an EDW architecture.*

3.4.5 Compatibility with existing systems

The state of existing systems is one indicator of the existing information processing capacity of organizations (Tushman et al. 1978). When increases in uncertainty mandate the need for greater information processing capacity, organizations may implement an information system to increase the existing information processing capacity of organizations (Galbraith 1973). Goodhue et al's (1992) analysis of the benefits and costs of data integration suggest that organizations will design and develop a new information system if the benefits gained from the increase in information processing capacity outweigh the costs of systems development.

The cost of systems development may be influenced by many factors. The nature of existing systems is one factor that may affect systems development costs. It can reduce costs by easing the implementation of new technologies or it can increase costs by impeding new development. Existing systems may increase the development costs when the existing system technology is incompatible with the technologies proposed for the development of the new system. For instance, when existing vendor technology does not match the technology considered for a new decision support system implementation, the organization will incur high systems development costs in trying to bridge the technology gaps between the existing and the new systems. Furthermore, as existing IT staff expertise reflect knowledge of existing system technologies (Swanson 1994), the organization will also have to invest in acquiring new skills and expertise required to design, implement, and maintain the new systems.

In contrast, if the existing vendor technology was used for the implementation of the decision support system, the organization can minimize development costs. For example, the organization can benefit from the use of staff skills and knowledge on existing technologies to support the implementation of the new technology. Consequently, past research indicates, that some firms concentrate on products offered by existing technology vendors or those compatible with existing technology vendors for new systems development (Huff et al. 1985).

3.4.5.1.1 *Hypothesis derived from literature and interviews*

To minimize the costs of systems development, organizations may choose products and technologies that are compatible with existing technology. Thus, the nature of the existing system products may affect the choice of new system products.

Currently, there are many vendors in the market that offer data warehouse products to potential clients. In developing products, vendors often align themselves with a particular data warehouse architecture that guides the development of a data warehouse. For instance, Teradata, a division of NCR, is a company that offers technology solutions aligned to the EDW architecture while Microsoft offers data storage products optimized for dimensional data structures (O'Donnell et al. 2002). Consequently, if an organization’s existing systems are more compatible with Teradata data warehouse technology offerings, they may choose Teradata products to implement a data warehouse in order to lower data warehouse development costs. Coincidently, the organization will be choosing an EDW architecture to implement a data warehouse solution.

The expert interviews also indicated that existing system compatibility might lead organizations to select a particular data warehouse vendor product. For instance, Ron Swift claimed that organizations with existing data sources that are IBM products may select an IBM data warehouse product in order to gain cost advantages. He indicated that since most IBM data warehouse products embody a FED architecture, the organization acquires a FED architecture as a results of selecting an IBM data warehouse product which is compatible with existing IBM data sources within the organization.

In conclusion, when implementing a data warehouse, an organization may select vendor products and technologies that are compatible with existing source systems as a means to reduce overall development costs. As a result, the technologies and vendor products selected for the data warehouse implementation will determine the data warehouse architecture for the organization.

Hypothesis 5: *Organizations in which existing systems are highly compatible with a particular data warehouse vendor product are more likely to select the data warehouse architecture that is aligned with that particular data warehouse vendor's products.*

3.4.6 View of the data warehouse

Organizations vary in the way they view the role of IT in their organizations. The literature presents several different categorizations of the view or role of IT. McFarlan and McKenney (1983) developed a strategy grid framework that categorizes variations in organizations' perception of IT. Johnston and Carrico's (1988) research indicated that an organization's perception of IT may fall under three different categories: (1) focus on IT to support operations that are not strategically related, (2) focus on IT to support strategy, and finally (3) companies may perceive IT as integral part of strategy.

Information processing theory suggests that organizations may differ in their view of IT depending on the source of uncertainty faced by the organization. For instance, Duncan (1995) asserts that organizations faced with uncertainty from a turbulent environment may perceive IT infrastructure as a mechanism that enables long-term competitive advantage. Consequently, organizations that view IT to be a strategic tool may implement complex technologies to support strategic initiatives. Ryan and Prybutok's (2001) study demonstrated that organizations which viewed IT as critical for its daily functions and future applications were more likely to adopt knowledge management systems to facilitate the creation of long-term competitive advantage.

In their study of IT infrastructure investment decisions in organizations, Broadbent and Weill (1997) noted that depending on their business and IT strategic objectives, organizations may differ in their view of IT infrastructure. For instance, companies may perceive IT infrastructure as an enabler of long-term strategic initiatives or may perceive IT infrastructure as a set of IT services that satisfy a specific business area need. In addition, they may view IT infrastructure as a utility that permits the organization to reduce costs or view IT infrastructure as a set of IT capabilities that is dependent on and supports current strategy. Each of these four views indicates the need for different IT capabilities (Weill et al. 1998). Thus, they provide a high-level view of the specific IT infrastructure desired by an organization.

3.4.6.1.1 *Hypothesis derived from literature and interviews*

Depending on the organization's motivation, organizations may differ in the manner in which they view the implementation of a data warehouse (Chen et al. 2000). For instance, companies may consider the implementation of a data warehouse as the development of decision-support infrastructure that enables long-term strategic initiatives or they may view it as a short-term point solution that addresses pressing business needs. The way they view the warehouse indicates the IT capabilities desired from a data warehouse implementation. As such, it influences the data warehouse architecture that should be selected. While discussing the evolution of data warehousing, Griffin (2001) indicated that expressing the future state or view of the desired data warehouse architecture is necessary in order to select a suitable architecture.

However, certain data warehouse architectures are limited in their capability to support organizational goals. Consequently, some architectures may not result in a data warehouse solution that is aligned with the organization's view of the data warehouse implementation. For instance, an IDM architecture provides decision support capabilities for a functional unit within an organization (Kelly 1997). Thus, if an organization views implementing a data warehouse as a point solution to meet a functional need, an IDM architecture should be selected. In contrast, First American Corporation viewed the data warehouse as a means for bringing organization-wide change, and they implemented an enterprise data warehouse architecture to meet their strategic objectives (Watson et al. 2002). Thus, the manner in which an organization views the data warehouse affects the data warehouse architecture selected for implementation.

During the interviews, some of the experts also provided insight into the relationship between how a data warehouse is viewed prior to implementation and the data warehouse architecture. For instance, Jim Thomann stated that an organization may see the data warehouse as a means of achieving organizational strategies, and might implement a robust data warehouse architecture solution such as an EDW architecture.

The existing literature and expert interviews indicate that the relationship between the view of the data warehouse and the EDW or IDM architectures is relatively clear compared to the rest of the architectures. The hypothesis describing the relationship between the view of the data warehouse and EDW or IDM architecture is:

> ***Hypothesis 6:*** *Organizations that view the implementation of a data warehouse as a short-term point solution rather than a strategic infrastructure project are more likely to select an IDM architecture than an EDW architecture.*

3.4.7 Resource availability

Organizations require slack resources to induce innovation (Miller et al. 1982). The development of complex innovations, in particular, requires resources beyond that needed for the daily operations of an organization. The analysis of innovation adoption research indicates that slack resources are positively related to organizational innovations (Damanpour 1991). They enable organizations to bear the costs of designing, developing, and instituting innovations. As such, an organization's ability to adopt innovations may be constrained by the availability of slack. In IS research, the availability of resources in terms of money, time, and people has been described as a partially controllable variable that affects IS development (Ein-Dor et al. 1978). In a field study that examined systems development, Tait and Vessey (1988) found that resource availability can constrain the design and development of information systems.

In the management literature, slack resources have been described as actual or potential resources that allow an organization to adapt to changes in the business environment (Greenley et al. 1998). Slack allows organizations to generate and hold resources that can enable strategies to help adapt to the dynamic external environment. For instance, slack gives the organization the opportunity to meet its need for increased information processing capacity by allocating resources to the development of information processing mechanisms (Miller et al. 1982). However, the amount of slack resources available for the development of information processing mechanisms may affect the information processing capacity that can be obtained.

3.4.7.1.1 *Hypothesis derived from literature and interviews*

The data warehouse is a mechanism that may enable an organization to meet its information processing requirements. However, depending on the data warehouse architecture chosen for implementation, the amount of slack resources required for the development differs. While some data warehouse architectures require high investment in resources, others require relatively less investment, at least initially. The greater the amount of slack an organization has to allocate to a data warehouse project, the wider the range of architectural options available for implementation. Likewise, the less slack resources an organization has to allocate for a data warehouse implementation, the architecture choices available for the organization become limited.

Thus, the choice of architecture is constraint by the availability of slack resources. For instance, because an IDM architecture requires less resource investment, many organizations with a low resource budget may choose a IDM architecture over an EDW architecture (Armstrong 1996). Kimball (1997) and Mimno (2001) state the ability to obtain a data warehouse solution with relatively less resource investment leads organizations to the implementation of an DBA architecture rather than an EDW architecture. In addition, Hackney (2000a) also suggests that FED architecture is a low cost alternative to EDW architecture.

The expert interviews also revealed that the availability of resources affected the choice of architecture implemented in an organization. For instance, according to Don Stoller, limited resource availability often leads organizations to develop data mart architecture solutions to satisfy their information needs.

The hypothesis stating the relationship between resource availability and the data warehouse architectures is:

Hypothesis 7: *Organizations with low resource availability are more likely to select an IDM, DBA, or FED architecture than an EDW architecture.*

3.4.8 The perceived ability of the in-house IT staff

Previous research in self-efficacy and organizational learning provide insight into how the perceived ability of the in-house IT staff affects choice of architecture.

According to the stream of research in self-efficacy, judgments or perceptions about an individual’s ability to use a computer, is defined as computer self efficacy (Agarwal et al. 2000).

This general conceptualization of computer self-efficacy does not help evaluate an individual's perception of their ability regarding a specific computer application. Specific computer self-efficacy refers to an individual's perception of efficacy in performing specific computer-related tasks (Marakas et al. 1998). In the context of a team, Gully et al (2002) describes efficacy as the team's belief about a task-specific team capability.

The most important antecedent of efficacy is enactive mastery experiences, which are experiences gained through past performance (i.e., either success or failures) in a task domain (Agarwal et al. 2000; Bandura et al. 1989). A person or team's existing specific self-efficacy along with experiences related to the specific application affects the formation of the next subsequent specific self-efficacy (Lindsley et al. 1995; Marakas et al. 1998). For instance, in the context of this study, existing perceptions of their ability to develop a data warehouse and experiences developing a data warehouse contribute to the creation of new perceptions about the IT staff's ability to develop a data warehouse. In addition, efficacy affects the decisions or choice of actions an individual or group may undertake (Gibson 1999). For instance, high team efficacy may lead them to higher performance and pursuits that they would otherwise not undertake.

Organizational learning research suggests that there are many knowledge barriers that impede the adoption, development, and infusion of complex technologies (Ravichandran 2001). Researchers have found that organizations with an IT staff that has experience related to a complex technology that is considered for adoption are more able to successfully adopt and sustain the complex technology (Fichman et al. 1997). Thus, as a result of the existing experience and skills, an IT staff may have low knowledge barriers to the adoption of a complex technology. The prior discussion on efficacy suggests that existing skills and successful prior experiences influence a team's perception of their ability to develop a complex technology. The staff's perceived ability affects their course of action with respect to adopting, developing, and sustain a complex technology. Thus, a team with high efficacy may perceive low knowledge barriers to the adoption of a complex system.

The rational perspective on organizations suggests that organizations will invest in organizational learning as long as the benefits from system implementation outweighs the total costs of implementation, including costs of organizational learning. The greater the perception of low knowledge barriers to adoption, the less investment required in organizational learning. This in turn helps minimize the overall costs of systems implementation.

3.4.8.1.1 *Hypothesis derived from literature and interviews*

A data warehouse is a decision-support infrastructure and a complex technology that may encounter knowledge barriers related to its development. However, some data warehouse architectures are more complex than others, leading to high knowledge barriers to development. As such, organizations that wish to implement a complex architecture will incur higher learning costs than organizations that choose to implement simpler architectures. However, the greater the in-house IT staff experience in implementing data warehouse solutions successfully, as well as the greater the relevant technical knowledge, the greater the perceived ability of the IT staff. Consequently, the greater the perceived ability of the IT staff, the lower their perception of knowledge barriers to implement a complex architecture. The perception of lower knowledge barriers suggests the need for low investment in organizational learning. As a result, when the perceived ability of the IT staff is high, the choice of architectures is not limited to simple architectures. Therefore, the extent of the perceived ability of the in-house IT staff determines the choice of data warehouse architectures.

Upon analyzing the difference between data warehouse architectures, Breslin (2003) concludes that developing an EDW architecture has high knowledge barriers requiring high technical skills and the experience of specialists. In comparison, data warehouse architectures with multi-dimensional data structures have relatively low knowledge barriers. As such, Breslin (2003) suggests that the perceived ability required to develop architecture solutions with dimensional structures (i.e., DBA and IDM) is relatively lower than the perceived ability required to implement an EDW architecture.

Several experts also described the impact of the perceived ability of the IT staff on the data warehouse architecture chosen for implementation during the interviews. For instance, Bill Inmon indicated that specialized IT skills and successful past experiences implementing data warehouses would influence a development team to choose an EDW architecture.

The impact of the perceived ability of the in-house IT staff in selecting a FED data warehouse architecture is difficult to determine from the existing literature. The hypothesis that reflects the relationship between the perceived ability of the in-house IT staff and EDW, IDM, or DBA architecture is:

> ***Hypothesis 8:*** *Organizations that have an in-house IT staff with low perceived ability are more likely to select an IDM or a DBA architecture than an EDW architecture.*

3.4.9 Source of sponsorship

A sponsor is an individual or group that allocates resources for a systems development project, fights political resistance, and promotes the benefits of a systems development project within the organization (Willcocks et al. 2000). Generally, a sponsor controls a scarce resource desired by other organizational members. Social political theories indicate this as a source of power for the sponsor (Bradshaw-Camball et al. 1991). It gives the sponsor the capability to influence decisions related to system implementation as the sponsor controls resource allocations for the project (Jasperson et al. 2003). It gives the sponsor more negotiating power over other stakeholders on project-related decisions. Consequently, the sponsor can gain desired outcomes from the system implementation project.

In a study examining IT deployment in both domestic and global organizational contexts, project managers stated that local level IT application design decisions were influenced by corporate management's motivations when upper management sponsored the application development project (Tractinsky et al. 1995). The study results indicated that upper management may impose centralized IT application solutions on business units when sponsoring IT applications in order to gain more control of business units. While analyzing the EIS development practices in industry, Watson et al (1995) discovered that the executive sponsor influenced decision making during various phases of the EIS development initiative.

Research indicates that sponsorship can come from many different sources. A survey of data warehousing managers revealed that 48 percent of respondents indicated that their warehouses were sponsored by the IT function, and 41 percent reported that a senior executive or functional area manager sponsored the project (Watson et al. 1998).

3.4.9.1.1 *Hypothesis derived from literature and interviews*

Depending on the source of sponsorship, the sponsor may impose a system design that is most beneficial to the organizational level represented by the sponsor, or meets the sponsor's personal goals for the data warehouse[3]. For instance, a social political view suggests that upper

---

[3] Information processing theories suggest that the influence of sponsorship on architecture decisions may be due to rational reasons. For instance, the upper management of an organization characterized by high interdependence between organizational units, may sponsor the development of a data warehouse to ensure that the data warehouse solution satisfies the need for greater information sharing across the entire organization. However, as social political

management may sponsor a data warehouse project to gain more control of the organization. As such, they will mandate the selection of an architecture that gives them more control, such as an EDW solution. In contrast, when sponsorship for the data warehouse implementation comes from specific functional units, such as at Sherwin Williams, a DBA architecture may be selected (Watson et al. 2001). Furthermore, the practitioner literature suggests that project sponsorship for IDM architecture implementations often arise from the functional unit that desires a data warehouse solution (Hackney 2000a).

The impact of the source of sponsorship on architecture selection was discussed by several experts during the interviews. For instance, Rob Armstrong suggested that the sponsor may use political influence to affect the choice of data warehouse architecture implemented. He claimed that when sponsorship comes from a particular department or functional area, the sponsor may select an architecture that specifically accommodates that department or functional (e.g., an IDM architecture).

The effect of sponsorship on the selection of FED is less clear. The following hypothesis describes the relationship between source of sponsorship and EDW, DBA, or IDM architecture.

Hypothesis 9: *Organizations with upper management sponsorship is more likely to select an EDW architecture than a DBA or an IDM architecture.*

3.4.10 Expert influence

Organizations face various knowledge barriers in the adoption of complex technologies (Chau et al. 1997; Ravichandran 2001). The management literature suggests that traditionally

theory suggests, the upper management may sponsor a project purely to accomplish desired personal objectives. As the impact of interdependence on architecture was previously described as a rational factor that affect the selection of a data warehouse, this factor will look at the influence of the source of the sponsor from a social political perspective.

organizations look to consulting firms, mass media, and business schools to reduce IT related uncertainty (Abrahamson 1991). Organizations perceive such sources of information as knowledge experts that help reduce the knowledge barriers that affect the development of information systems in organizations (Bloomfield et al. 1992). Most external organizations such as IT vendors and consultants utilize avenues like IT forums, mass media, and direct contact with client organizations to provide their expert advice and make recommendations on what technology, designs, and deployment approaches best fits a client organization. Additionally, they may use these client organization interactions to disseminate information regarding their IT product(s), its underlying design, development methodology, and services.

Social political theory suggests that these external organizations perceive their intimate knowledge about a technology as a source of power that can help them influence client decisions regarding the technology (Bloomfield et al. 1992). As such, in addition to being a technical knowledge sharing activity, the external organization's knowledge dissemination activities and each interaction with client organizations may be perceived as a social political activity which may influence clients to consider adopting the product(s), design, or service offered by the expert source (Bloomfield et al. 1995). Considering the impacts of purely social political activities of experts, Willcocks (2000) cautions organizations of the influence of outside consultants on ERP implementations. The author suggests that consultants or vendors may affect architectural planning of an ERP implementation by recommending the implementation of an incompatible architecture that promotes the use of products and services that they offer. However, it is difficult to discern the extent to which external organizations' interactions with client firms are motivated by social political goals versus technical knowledge dissemination objectives. In general, an external organization that engages in a high degree of political activity will soon loose credibility among clients for recommending products and services that does not benefit organizations.

Thus, while external organizations may interact with client organizations for rational reasons (i.e., to disseminate technical knowledge and provide an appropriate solution to the organization), they may also engage in interactions for social political reasons. While recognizing that rational reasons could drive activities by external organizations, this study considers the influence of social political activities of external organizations on architecture selection.

3.4.10.1.1 *Hypothesis derived from literature and interviews*

In the data warehousing realm, currently there are many external organizations serving the needs of client organizations. They provide technical expertise to embark on these complex decision-support infrastructure projects, recommend products, methodologies, architectures, and strategies to data warehouse implementation. Theory suggests that client organizations may choose to adopt a technology, design, or service based on the influences of expert sources (Abrahamson 1991).

The expert interviews also indicated that influences by consultants, trade seminars that are attended, and the data warehouse literature may influence an organization’s choice of data warehouse architecture. For instance, Wayne Eckerson stated that the consultants an organization employs to develop the data warehouse may prescribe the use of a data warehouse architecture that the consultants are most skilled at implementing.

As such, in the context of data warehouse architecture selection, the exposure to expert influences may affect the choice of architecture in a particular organization.

> ***Hypothesis 10****: Organizations with exposure to expert(s) are more likely to select the architecture aligned to the knowledge and services offered by the expert(s).*

3.4.11 Summary of data warehouse architecture selection

The following research model (Figure 3.1) indicates the most salient factors that affect data warehouse architecture selection based on the analysis of relevant literature and interviews with experts.

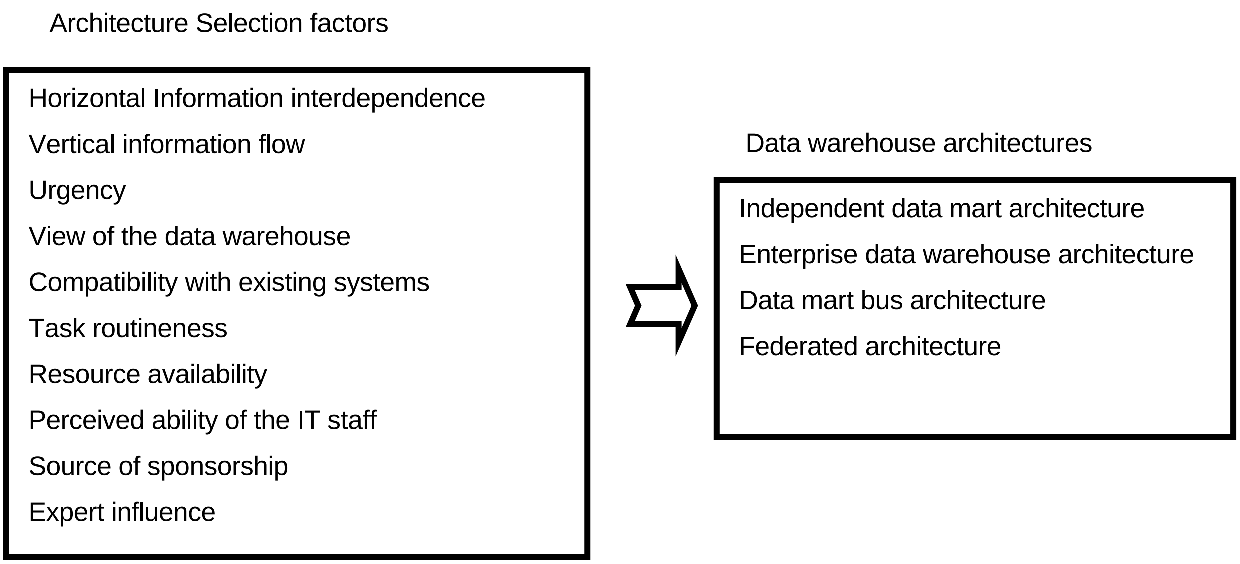


Figure 3.1: Data warehouse architecture selection

3.5 IS success

Understanding information system success is critical to the management of IS and evaluating the value of IS investments (Raymond 1990). Over the years, researchers have investigated information systems success in numerous ways. As early as 1973, Powers et al (1973) investigated MIS success in terms of use, user satisfaction, time, and budget. Recently IS researchers have explored other surrogates for IS success such as service quality (Pitt et al. 1995) and individual impacts (Goodhue et al. 1995). Currently, there is a vast body of literature on IS success that has examined various factors as surrogates for system success.

A review of this success literature reveals that system use and user information satisfaction is the most predominantly used surrogates for success in the literature (Tait et al. 1988). System use and user information satisfaction focus on the success of the system product. Timo (1996) states that in addition to the success of the system product, development process success is an important element of success that requires assessment in order to gain a comprehensive understanding of IS success.

3.5.1 Data warehouse architecture success

As described in chapter two, a data warehouse architecture is a blueprint that provides guidance for the development of a data warehouse solution. As such, the assessment of the success of the architecture requires the analysis of the success of the architecture development process as well as the evaluation of the resulting architecture solution. Data warehouse architecture success can be described as the efficiency and effectiveness of the data warehouse architecture used. It takes into account two key types of success: development process performance (i.e., an assessment of the process of implementing a data warehouse architecture solution and the project) and system product performance (i.e., an assessment of the quality of the developed architecture solution; in other words, the product or output of that process). Each type of success requires a separate assessment, as they are not necessarily highly correlated (Wateridge 1995). For example, an over-budget or beyond-schedule project may deliver a high-quality product. Conversely, a within-budget and on-time project may deliver a product of poor quality. Both system product success and development process success are discussed next.

3.5.2 System product success

Delone and McLean's (1992) IS success model (i.e., here after referred as the D&M model), first introduced in 1992, is by far the most influential model in conducting research on information systems product success factors. It organizes the dependent variables used to measure IS success into a taxonomy with six interrelated dimensions: (1) system quality, (2) information quality, (3) system use, (4) user satisfaction, (5) individual impact, and (6) organizational impact. The model recognizes both process and causal links between the six dimensions. Subsequent studies have refined and extended the model (Drury et al. 1998; Pitt et al. 1995), as well as critically analyzed the model (Seddon 1997).

Pitt et al (1995) observed that service quality was an important category of IS success not addressed in the D&M model of information system success. Drury et al (1998) presented a modification to the D&M model by introducing a hierarchical structural model that organized IS success research by two basic types of properties: (1) generic properties shared by all successful systems and (2) specific properties specific to a type or class of systems. Seddon (1997) proposed a respecified model of IS success stating that the D&M model can lead to potentially confusing measures of success as it combines both process and causal interpretations of IS success.

Delone and McLean (2003) revised the model in 2003 to address changes and progress in IS success research since their original study. The revised model is presented in Figure 3.2. The researchers combined individual and organization impacts into a new dimension named net benefits and added service quality as an additional component to measure success. They also distinguished intention to use from system use as an important component that may be more effective to measure success than system use in certain contexts.

Through their investigation of success measures in IS research, Delone and McLean (1992; 2003; 2002) concluded that there is no one success measure that applies to all studies. They suggest that the selection of success dimensions and measure must be contingent on the objectives and the context of the empirical investigation. Specifically, they recommend that researchers should consider the aspect(s) of the information system under study and the independent variables under investigation when selecting IS success measures. In order to gain a better understanding of success for the phenomenon understudy, they also recommend the use of multiple measures in evaluating IS success. Consequently, these recommendations were considered in identifying success measures for this dissertation.

3.5.3 Development process success

According to the project management literature, process success has been the primary measure of success in the assessment of project management success (Wateridge 1998). Specifically, process success determined in terms of the successful accomplishment of cost and time objectives has been the focus of many studies that evaluate IT project success (Baccarini 1999). While the time criterion of process success has been measured in terms of schedule over/underun as a percentage of the initial plan, the cost criterion of process success has been measured in terms of cost over/underun as a percentage of the initial plan (Might et al. 1985).

Wateridge (1995) suggests that the judgment of whether a project has successfully met the objectives of time and costs is a short-term measure of success made on the completion of a project which may seem less significant over time as the organization begins utilizing the functionality of the system. However, the assessment of process success provides a means of determining the efficiency of a particular systems development process that can help evaluate different systems development methodologies, approaches, and architectures.

3.5.4 Choice of success dimensions

Delone and McLean (1992) identified the original dimensions of IS success based on the three levels of communication described by Shannon and Weaver (1949) and later adopted to IS research by Mason (1978). The three levels are the technical, semantic, and effectiveness levels. The technical level, which describes the accuracy and efficiency of the system that produced information, is measured by system quality. The semantic level, which describes the success of information in conveying the intended meaning, is evaluated using information quality. Finally, the remaining success dimensions measure the effectiveness level, which describes the effect of the information on the receiver.

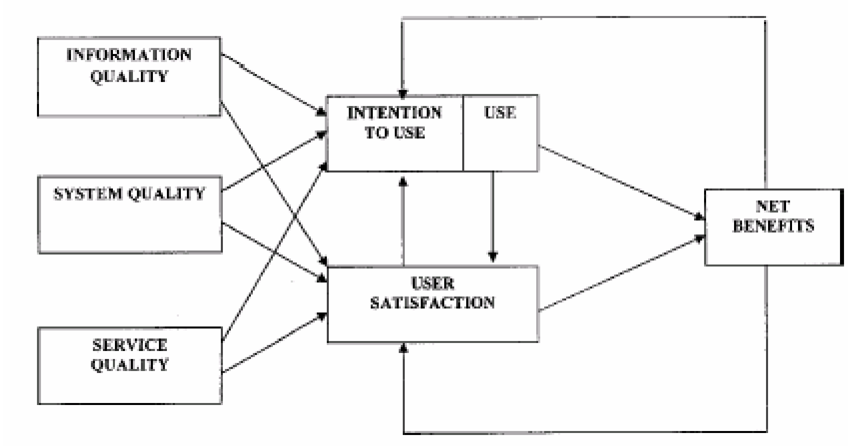


Figure 3.2: The updated D&M Model adopted from (Delone et al. 2003)

The aspect of the information system, which was of interest to this study, was the architecture chosen to implement a data warehouse product. An analysis of the three different levels of success described above suggested that measuring the success of data warehouse architectures required the measurement of the technical and semantic levels of success. A data warehouse architecture presents a blueprint to guide the development of a decision-support infrastructure (i.e., the data warehouse). The architecture does not dictate the development of specific decision-support applications that end users may utilize to access information in the data warehouse. Thus, the effect of information on the data warehouse user as described by the effectiveness level is less applicable to the evaluation of data warehouse architecture success. As such, system quality and information quality, which represents technical and semantic levels of success, were chosen to evaluate data warehouse architecture success.

While cautioning researchers to choose success dimensions that fit the objectives and context of research, Delone and McLean (2003) recommend Seddon’s (1999) IS effectiveness matrix as a valuable reference for the selection of success measures based on context. Seddon (1999) recommends that measures of IS success should be selected with the type of system and the stakeholders in mind. The author describes the stakeholder as the person for whom the system is being evaluated, such as an independent observer, individual, or group that wants to benefit, or the manager or owner that wants the organization to benefit. The type of system dimension is used to categorize the system that is being evaluated and is composed of six components: (1) an aspect of a system, (2) a single IT application, (3) a type of IT or application, (4) all IT applications across sub-units or the organization, (5) an aspect of a systems development methodology, and (6) the IT function. The resulting two-dimensional framework contains cells, each describing a particular type of system being evaluated from the point of view of a stakeholder (See Table 3.3). To better understand the context and the success measures relevant for this study, the IS effectiveness framework was utilized.

According to the framework, the type of system examined in this dissertation fell into the fifth category, which describes an aspect of systems development methodology. Studies in this category consider the methodology as *the system* and they focus on the effectiveness of various methodologies for developing a system. This study focused on the effectiveness of different data warehouse architectures for implementing data warehouses. As such, it required the point of view of a stakeholder who would focus on the benefits for the organization that planned to implement the data warehouse architecture. The framework suggests that owners or managers focus on the benefits for the organization; therefore, the data warehouse manager was chosen as the stakeholder for this study. This study represents the fourth row, fifth column of the IS effectiveness matrix (See shaded area of Table 3.3).

The point of view of the data warehouse manager suggests that some success dimensions better fit the context of this dissertation. Generally, the data warehouse manager is an individual who directs the development of the data warehouse and currently manages the operational system.

As such, the data warehouse manager is most familiar the nature of the system and its information. However, the data warehouse manager would be less qualified to assess success in terms of users' satisfaction with the system. Consequently, using the point of view of the data warehouse manager suggested the use of information quality and system quality from the D&M model to measure data warehouse architecture success.

However, while academic literature pointed to the selection of information quality and system quality as the appropriate architecture success measures for this study, the expert interviews as well as the expert review phase suggested otherwise. As mentioned previously, the point of view of the data warehouse manager suggests that he/she maybe most well qualified to answer questions on information quality and system quality, and less qualified to accurately answer questions regarding the individual and organizational level impacts and the business value from the data warehouse. Yet, the experts strongly recommended the inclusion of net benefits as an additional measure of success during both the initial interview phase and the feedback/review phase. From a practitioner standpoint, individual and organizational impacts represent the most important and interesting measure of architecture success. Furthermore, a recent study that examined data warehousing success (Wixom et al. 2001) also included questions on net benefits despite using the data warehouse manager as the key respondent of the study. As such, in addition to information quality and system quality, net benefits was also selected as another measure of architecture product success.

While system quality, information quality, and net benefits evaluate the success of the data warehouse solution (i.e., product success) resulting from the implementation of a data warehouse architecture, it does not assess the impact of the architecture on the development process. Timo (1996) claims that the assessment of IS success should extend to the evaluation of the systems development process and not be limited to the assessment of the system product. Systems development projects vary in terms of the quantitative and economic measures such as costs and duration of systems development (Doll 1985). Managers often use these quantitative metrics to assess the benefits of investing in one development project compared to the other alternatives (Timo 1996). The effectiveness of the development process of each architectural solution may differ according to the architecture selected for implementation. The development process performance of each architectural solution may differ in terms of the time schedule and budgetary considerations. As such, in addition to product success measures such as information quality and system quality, development process success was also included as an important dimension of architecture success.

Table 3.3: IS effectiveness matrix based on Seddon et al. 1999

STAKEHOLDER / INTEREST GROUP		(1) An aspect of IT design or use	(2) A single IT application in an organization	(3) A type of IT or IT application	(4) All IT applications used by an organization or sub organization	(5) An aspect of a system development methodology	(6) An IT function
(1)	Independent observer						
(2)	Individual Primary focus: Individual better-offness						
(3)	Group Primary focus: Group better-offness						
(4)	Management or Owners (of a firm) Primary focus: Organizational better-offness					Data warehouse architecture success evaluation	
(5)	A Country Primary focus: Society's better-offness						

The data warehouse literature also indicates the quality of information, the flexibility and scalability of the resulting data warehouse (i.e., system quality), and the costs associated with developing a data warehouse solution (i.e., development process success) as important measures of architecture success that vary among architecture implementations. For instance, Armstrong (1996) discusses information quality in terms of accuracy and completeness as being low in IDM architecture implementations as compared to EDW architecture implementations. Hackney (2002) addresses the system quality of architecture by stating that one of the flaws of the IDM architecture is its inability to be scaled for changes and growth. Kimball et al (2001) describes net benefits to data warehouse users as the single most importance factor that dictates the success or failure of an architecture implementation. Hwang et al (2002) stresses the importance of development process success measures by stating that the implementation costs are a major barrier inhibiting the implementation of an EDW architecture compared to an IDM architecture.

Therefore, the success dimensions chosen for the study were information quality, system quality, net benefits, and development process success.

3.5.5 Information quality

Information quality evaluates the quality of the information available in an IS system. Past IS studies have examined information quality by evaluating a variety of measures of information quality such as accuracy, timeliness, and consistency (Wang et al. 1996). Information quality is also recognized as an important dimension of success in the data warehouse literature. In a survey that assessed industry data warehousing practices, Watson et al (1997) discovered that 79 percent of data warehouse managers stated that the need for information quality as the primary motivation for implementing a data warehouse. Recently, two empirical studies in data warehousing that measured data warehouse success investigated attributes of information quality such as accuracy, consistency, and completeness of data in data warehouses (Shin 2003; Wixom et al. 2001). While certain architectures ease the establishment of information quality characteristics like consistency, others make it more challenging to gain high information quality.

During the expert interviews, the experts consistently indicated that information quality measures were an important criterion of data warehouse architecture success. For instance, Don Stoller stated that quality measures such as accuracy, consistency, and completeness are metrics that provide an estimation of the quality of the information in the data warehouse architecture solution. Another expert, Bill Inmon, suggested the levels of detail of data that can be stored in the data warehouse is another indicator of the information quality of the architecture solution. Next, different characteristics of information quality are described.

3.5.5.1 Accuracy

Past research indicates that information accuracy is one of the ten most important factors related to user information system satisfaction (Bailey et al. 1983). Furthermore, among measures of information quality, accuracy represents one of the most extensively studied attributes of information in the IS research literature (Rai et al. 2002). In data warehousing, accuracy is also considered a necessary attribute of information quality (Devlin 1997). A recent survey that evaluated the information quality of a data warehouse further confirms that accuracy is important to data warehouse users (Shin 2003).

It is challenging to identify the specific relationship between different data warehouse architectures and accuracy. This is due to the fact that accuracy, the ability to represent real world objects correctly in the data warehouse, is considered fundamental to implementing a data warehouse. Nevertheless, it is conceivable that certain architectures may ease the creation of information correctness than others.

3.5.5.2 Completeness

Much like accuracy, information completeness has also been previously examined as a key attribute of information quality in IS research (Delone et al. 1992). It is an important measure that describes information quality issues faced by many organizations. An industry survey on information quality indicated that more than 60 percent of organizations surveyed had problems with information quality attributes, including information completeness (Arnold 1992).

Past researchers have used data completeness to discover if all needed relevant data was stored in the information system. In the context of the data warehouse, Jarke et al (1999) describes completeness as a factor that measures the extent to which all needed business rules, entities, and attributes of data are included in the data warehouse. Similar to accuracy, the exact nature of the relationship between data warehouse architecture and completeness is difficult to discern.

3.5.5.3 Consistency

Wang et al (1996) developed a hierarchical framework of information quality to obtain a complete view of all the attributes that encompass information quality. They recognized data consistency as an important characteristic of information quality, which was not mentioned in Delone and McLean's (1992) analysis of IS success. In data warehousing, consistency measures whether the representation of data in the data warehouse is uniform with no data duplications, overlaps, or confusing definitions (Vaduva et al. 2001). As a result of the survey of data warehouse system success, Shin (2003) discovered that data consistency was a difficult problem to address in data warehouse implementations.

3.5.5.3.1 *Hypothesis derived from literature and interviews*

Depending on the data warehouse architecture chosen for implementation, data consistency may be compromised to a certain extent. For instance, data consistency is relatively challenging to create in an IDM architecture implementation because the development of each data mart occurs independent of the other (Hackney 2000). Thus, compared to the other architectures, IDM architecture may not lead to the development of a data warehouse solution with high data consistency. This relationship is described in the following hypothesis:

> ***Hypothesis IQ1****: An IDM architecture solution is less likely to be associated with high data consistency than a DBA, EDW, or FED architecture solution.*

3.5.6 System quality

System quality focuses on the system product. Flexibility, integration, and resource utilization have been used in previous studies to evaluate system quality (Delone at al. 1992). In the data warehouse literature, system quality is considered one of the most important advantages of data warehousing. While examining the factors that influence data warehousing success, Wixom et al (2001) concludes that architecture design may affect different aspects of data warehouse system quality.

The expert interviews also indicated that the quality of the system was an important measure of data warehouse architecture success. For instance, Jim Thomann suggested that the flexibility and the scalability of the resulting data warehouse architecture solution are significant metrics that highlight the success of a data warehouse architecture. Next, different aspects of system quality are described.

3.5.6.1 Flexibility

Flexibility has received much attention among IS researchers as a factor that can enhance competitive advantage in organizations (Byrd 2000; Duncan 1995). In the context of system architecture, Allen and Boynton (1991) describe flexibility as an emerging issue critical to designing an architecture that can adapt to changes in the dynamic environment. Furthermore, in designing decision support systems, a flexible decision-support architecture is deemed necessary to support different decision styles and different problems (Fazlollahi et al. 1997).

3.5.6.1.1 *Hypothesis derived from literature and interviews*

A review of the practitioner literature suggests that certain data warehouse architectures support the development of data warehouse solutions that provide the capability to adjust (i.e., flexibility) to changes in current and future information needs (e.g., Hackney 2000a). Some authors claim that normalized data structures are better able to handle changes to business rules and information needs than dimensional data structures (Barbusinski et al. 2001; Manning 2000). They argue that architectures with normalized data structures provide the flexibility to meet future information needs. The EDW architecture includes normalized data structures. Therefore, the practitioner literature suggests that EDW architecture support greater flexibility to meet changing information needs than the other architectures. The hypothesis that describes this relationship is:

> ***Hypothesis SQ1****: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than an IDM, DBA, or FED architecture.*

3.5.6.2 Integration

Integration has been the cause of much discussion in the literature as a factor generally considered to be essential for effective decision making (Goodhue et al. 1992). As a measure of system quality, integration indicates the extent to which data in different data sources of the organization are linked together through an information system (Gattiker et al. 2003). Many researchers have focused on integration as an important attribute of system quality. For instance, recently, Gattiker et al (2003) analyzed the impact of ERP systems in creating data integration between subunits in manufacturing plants.

3.5.6.2.1 *Hypothesis derived from literature and interviews*

A data warehouse enables integration of disparate data across multiple data sources (Gray et al. 1998). The extent of data integration varies according to the data warehouse architecture chosen for implementation. An EDW architecture guides the development of a data warehouse solution that integrates data across numerous disparate data sources within the organization (Inmon 1992). In contrast, the IDM architecture focuses on integrating data within a specific functional unit. The propagation of IDM architecture solutions across functional units is often criticized as leading to the creation of disintegrated information silos (Mimno 2002). Thus, an EDW architecture allows greater integration across disparate data sources than an IDM architecture. However, the relationship between the rest of the data warehouse architectures and integration is less clear. The hypothesis that describes the relationship between EDW or IDM architecture and integration is:

> ***Hypothesis SQ2****: An EDW architecture is more likely to lead to the development of a data warehouse solution that enables high data integration than an IDM architecture.*

3.5.6.3 Scalability

The growth of the Internet and the explosive growth in data has created the need for scalable IS solutions (Bowman et al. 1994). Furthermore, advocates of open systems stress the importance of developing scalable information systems that can handle increased demands for system growth (Chau et al. 1997). Therefore, recently, the importance of scalability has received considerable attention in the literature as an essential measure of information systems quality.

3.5.6.3.1 *Hypothesis derived from literature and interviews*

In data warehousing, the rising volume of data stored for decision support needs has become a cause of concern for some organizations. Whiting (2002) claims that as certain data warehouse architectures solutions are not scalable; they are incapable of handling escalating user demands for information. For instance, architecture solutions with multidimensional data structures are restricted in the number of subject areas they can support while the EDW architecture is a scalable architecture solution that enables organizations to meet growth in subject areas due to demands from the dynamic environment (Armstrong 1997). The data warehouse architecture with purely multidimensional data structures is the DBA architecture. The impact of the rest of the architectures on scalability is difficult to determine. The hypothesis describing the relationship between EDW or DBA architecture and scalability is:

> ***Hypothesis SQ3****: An EDW architecture is more likely to be associated with the development of a scalable data warehouse solution than a DBA architecture.*

3.5.7 Net benefits

Net benefits describe the impact or value gained by implementing an information system (Seddon 1997). The impacts from an information system differ along a continuum depending on the entities that can be affected by IS activities including individuals, groups of individuals, organizations, and industry. According to Delone et al. (2003), the choice of which impacts should be measured is dependent on the system being evaluated and the purpose of the system.

In the context of data warehousing, Wixom et al. (2001) assessed perceived net benefits from data warehousing in terms of both individual and organizational impacts. Their study revealed that higher levels of data and system quality are associated with higher levels of net benefits. In addition, several empirical studies in IS literature have assessed and found that these three dimensions of success are significantly related to each other (Fraser and Salter 1995; Seddon and Kiew 1994). As such, the links from system quality and information quality to net benefits was added to the research model for architecture success (See Figure 3.4). However, when compared to both system quality and information quality, net benefits is loosely linked to data warehouse architecture success. There can be many confounding factors influencing the perceived net benefits of an architecture implementation. For example, the application and data access tools used to access the architecture solution may influence the perceived individual and organizational benefits.

Nevertheless, the expert interviews also indicated that individual level impact and organizational level impact resulting from the implementation of a data warehouse architecture are important measures of architecture success. For instance, Margy Ross described the importance of individual impacts by stating that some data warehouse architectures are less likely to result in the development of a data warehouse that enhances the decision-making ability of end users due to the complexity of the architecture’s data structure. Rob Armstrong discussed organizational level impacts of data warehouse architecture implementations by stressing that the return on investment of implementing a data warehouse may differ according to the data warehouse architecture chosen for implementation.

3.5.8 Development time and cost

Research in systems design and development has focused on development time and development costs in order to assess IS development success (Doll 1985). Keil (1995) suggests that monitoring the project duration and resource investment in a systems development effort indicates if a project is headed towards failure. According to Sambamurthy et al (2000), time schedule and budgetary considerations are among the key factors used to evaluate the outcomes of systems development.

In the data warehousing realm, many industry experts have also examined the development time and cost in order to assess the success of data warehouse implementations (Hackney 2000a). Many data warehouse publications suggest that the EDW architecture requires a large upfront monetary investment compared to the other architectures (e.g., Hwang et al. 2002, Joshi et al. 1999). Armstrong (1997) states that although initially IDM architecture solutions may seem less expensive compared to EDW architecture solutions, they require a high investment in the long term. Thus, the impact of architecture on cost is hard to discern from the existing literature. Similarly, it is challenging to identify the exact nature of the relationship between data warehouse architecture and development time.

The expert interviews indicated that development process success is an important measure of architecture success for organizations. The experts suggested that the time and costs incurred to implement the first roll out of the data warehouse architecture solution, as well as for the overall data warehouse project, are key determinants of data warehouse architecture success.

3.5.9 Summary of data warehouse architecture success

The following research model (Figure 3.4) indicates the most appropriate dimensions to measure data warehouse architecture success based on the analysis of relevant literature and interviews with experts.

3.6 Conclusion

This chapter first presented the results of the expert interviews, the relevant research from contingency literature, adoption literature, and rational and social political theories of organizations to identify the ten salient factors that affect architecture selection. In addition, through the review of literature and interview results, hypotheses linking selection factors to certain architectures were created. The review of IS success literature that followed identified the two types of success and their attributes appropriate for this dissertation. An attempt was made to create hypotheses that describe the relationship between certain architectures and success measures. The third chapter built the foundation to develop the research models for data warehouse architecture selection and data warehouse architecture success. Chapter four describes the specific research models, and presents the research hypotheses.

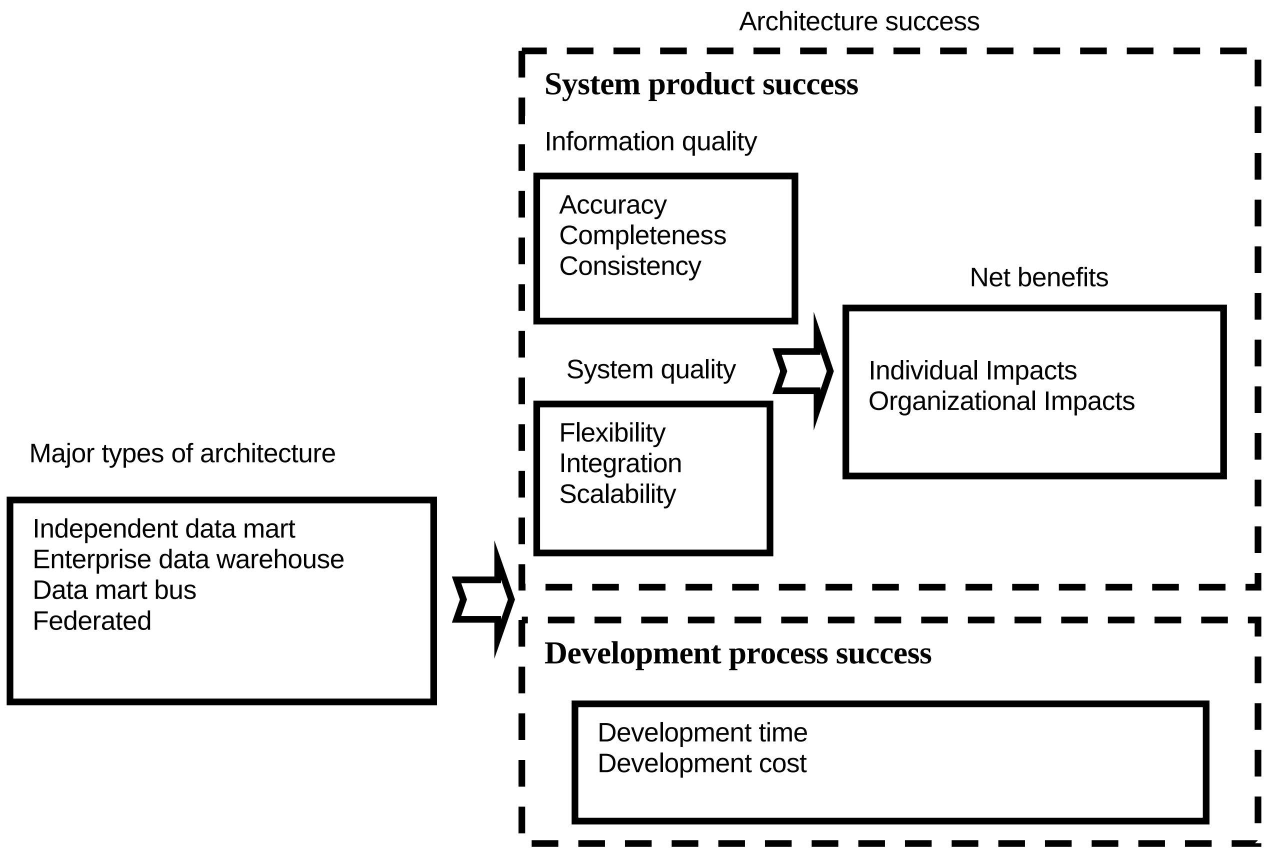


Figure 3.4: Data warehouse architecture success

CHAPTER 4 – HYPOTHESES

4.1 Introduction

Chapter three identified the most salient factors that affect data warehouse architecture selection using rational and social political theories of organizations and interviews with a panel of ten data warehouse experts. It also discussed the success factors applicable to the evaluation of data warehouse architecture success derived from theory and expert interviews.

Based on the selection factors and success factors described in chapter three, this chapter discusses the two research questions of this dissertation in terms of two main exploratory hypotheses and specific hypotheses derived from theory.

4.2 Propositions and Hypotheses

Propositions are specific conclusions about the relationship among concepts that are derived from theoretical review (Babbie 2000). Propositions make these assertions derived from theory in order to answer the research questions (Punch 1998). In contrast, hypotheses, describe a testable expectation about reality based on a theoretically derived proposition. In order to answer the two research questions that led to this investigation, hypotheses were formulated from theory and qualitative expert interviews. Next, each of the research questions is presented along with its hypotheses and research models. For reference purposes, the key constructs identified in the research models and their definitions are summarized in Table 4.1.

Table 4.1: Definitions of the key constructs in research models

Selection Factors	Definition
1. Horizontal information interdependence	The extent to which tasks and their outcomes were contingent upon information from one or more other organizational units.
2. Vertical information flow	The extent to which senior management's activities and performance were dependent on detailed information from lower organizational levels.
3. Urgency	The extent to which there was an urgent need to build the data warehouse.
4. Task Routineness	The extent to which users' jobs required non-routine data analyses.
5. Compatibility with Existing Systems	The extent to which the data warehouse architecture was compatible with existing systems.
6. View of the data warehouse	The extent to which implementing a data warehouse was viewed as being an infrastructure project important to supporting strategic initiatives.
7. Resource availability	The extent to which IT personnel, business unit personnel, and monetary resources were available for building the data warehouse.
8. Perceived ability of the in-house IT staff	The extent of the perceived ability of the in-house IT staff in terms of the relevant technical skills, successful experiences and confidence in developing data warehouses.
9. Source of Sponsorship	Source of sponsorship for the data warehousing initiative.
10. Expert Influence	The extent to which the organization's choice of data warehouse architecture was influenced by experts.
Success Factors	**Definition**
1. Flexibility (System Quality)	A measure of the flexibility of the data warehouse architecture to handle changes in information requirements.
2. Scalability (System Quality)	A measure of the scalability of the data warehouse architecture to meet increases in growth.
3. Integration (System Quality)	The extent to which the data warehouse integrates data from organizational source systems
4. Accuracy (Information Quality)	The accuracy of the data in the data warehouse.
5. Completeness (Information Quality)	The extent to which the data warehouse architecture enables all needed business rules, entities, and attributes to be included in the data warehouse.
6. Consistency (Information Quality)	The extent to which the data warehouse architecture supports data consistency across the entire organization.
7. Individual Impact (Net Benefits)	The extent to which the data warehouse architecture has results in benefits to individuals.
8. Organizational Impact (Net Benefits)	The extent to which the data warehouse architecture has results in benefits to the entire organization.

9. Cost of data warehouse (Development process success)	The cost of developing the data warehouse.
10. Development time (Development process success)	The duration to complete data warehouse architecture.

4.2.1 Research question on architecture selection

The first research question asked, “What is the relative importance of selection factors in the selection of a particular data warehouse architecture?” While the third chapter described several specific hypotheses for some of the relationships between selection factors and certain data warehouse architectures, the existing literature and the interviews did not help describe how each of the selection factors influence the selection of a particular architecture.

As a result, the part of the study designed to find answers to the first research question was more exploratory in nature. In order to identify what factors influence the selection of a particular data warehouse architecture, a general exploratory hypothesis was created. It suggests that each of the organizational factors impact the choice of a data warehouse architecture:

Hypothesis E1: *The selection factors influence the selection of a particular data warehouse architecture.*

The corresponding research model for the above hypothesis is presented in Figure 4.1.

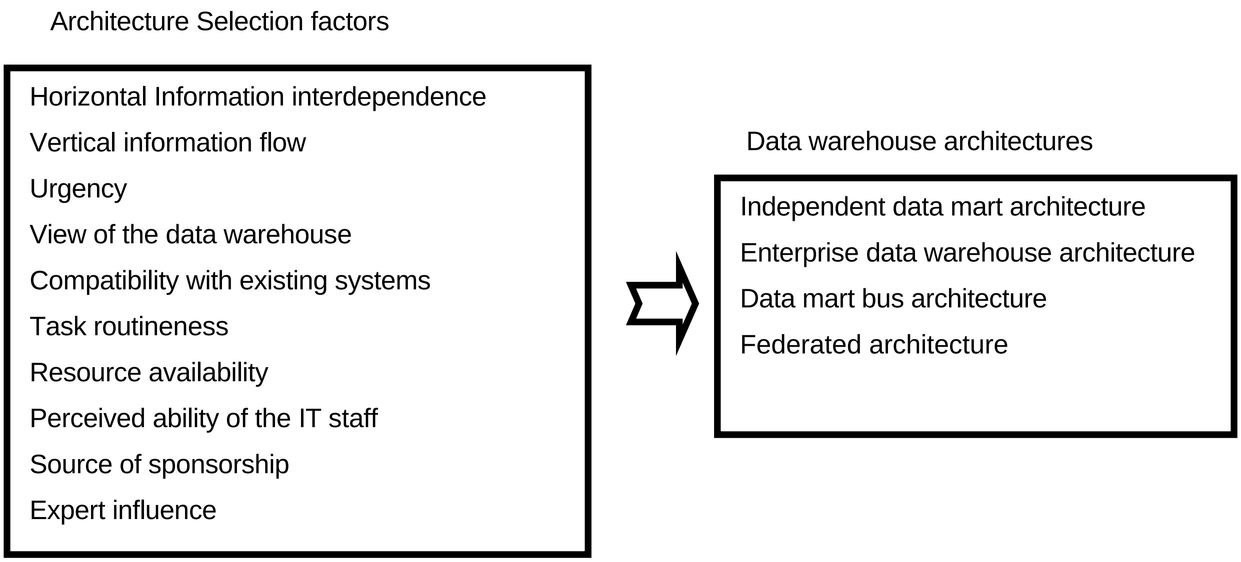


Figure 4.1: Hypothesis E1 on data warehouse architecture selection

Next, the impact of specific selection factors on a particular data warehouse architecture was considered.

4.2.2 Hypotheses on data warehouse architecture selection

All in all, ten specific hypotheses were created to describe the influence of organizational factors on architecture selection. Chapter three described the development of each of them. Several of these hypotheses indicate how different factors affect the selection between two particular architectures. Whenever possible, individual hypotheses are combined to provide some clarity to the presentation of hypotheses. Moreover, at times a specific hypothesis illustrates the influence of a selection factors on several architectures. In order to ease hypothesis testing, such hypotheses were further divided to mini hypotheses. Each mini hypothesis describes the influence of a selection factor on a single architecture when compared to another.

The following hypotheses describe the influence of information interdependence, vertical information flow, and view of the data warehouse on the selection of the EDW architecture when compared to the IDM architecture. The research model that combines and describes these hypotheses (i.e., hypotheses 1, 2, and 6) is presented in Figure 4.2.

Hypothesis 1: *Organizations with high information interdependence are more likely to select an EDW architecture than an IDM architecture.*

Hypothesis 2: *Organizations with high vertical information flow are more likely to select an EDW architecture than an IDM architecture.*

Hypothesis 6: *Organizations that view the implementation of a data warehouse as a short-term point solution rather than a strategic infrastructure project are more likely to select an IDM architecture than an EDW architecture.*

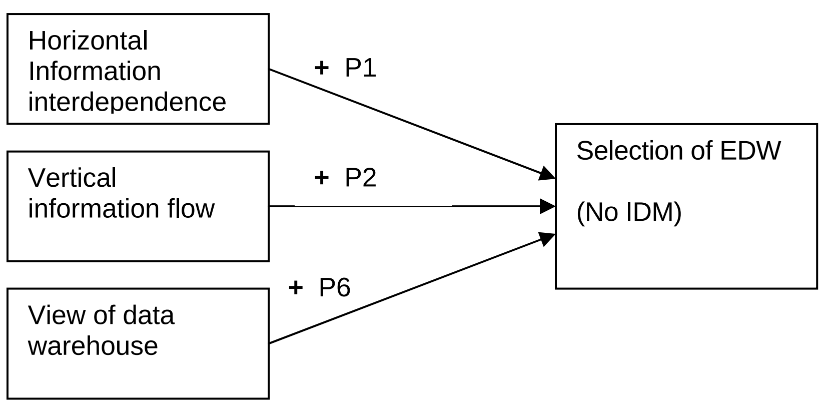


Figure 4.2: Research model for hypotheses: P1, P2, and P6

The next three hypotheses illustrate how three selection factors affect the choice between a DBA or an IDM architecture when compared to an EDW architecture. Specifically, they describe the influence of urgency, the perceived ability of the IT staff, and source of sponsorship. The research model that describes and combines these hypotheses (i.e., hypotheses 3, 8, and 9) is presented in Figure 4.3.

Hypothesis 3: *Organizations with high urgency are more likely to select a DBA or an IDM architecture than an EDW architecture.*

Hypothesis 8: *Organizations that have an in-house IT staff with low perceived ability are more likely to select an IDM or a DBA architecture than an EDW architecture.*

Hypothesis 9: *Organizations with upper management sponsorship is more likely to select an EDW architecture than a DBA or an IDM architecture.*

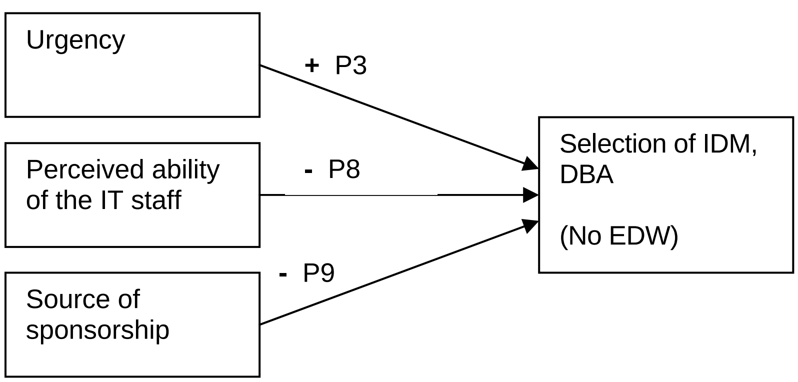


Figure 4.3: Research model for hypotheses: P3, P8, and P9

The mini hypothesis 3a, 8a, and 9a describe the influence of urgency, the perceived ability of the IT staff, and source of sponsorship on the selection of an IDM architecture when compared to an EDW architecture. On the other hand, the mini hypotheses 3b, 8b, and 9b look at the influence on DBA architecture compared to EDW architecture for the same selection factors.

Hypothesis 3a: *Organizations with high urgency are more likely to select an IDM architecture than an EDW architecture.*

Hypothesis 8a: *Organizations that have an in-house IT staff with low perceived ability are more likely to select an IDM architecture than an EDW architecture.*

Hypothesis 9a: *Organizations with upper management sponsorship is more likely to select an EDW architecture than an IDM architecture.*

Hypothesis 3b: *Organizations with high urgency are more likely to select a DBA architecture than an EDW architecture.*

Hypothesis 8b: *Organizations that have an in-house IT staff with low perceived ability are more likely to select a DBA architecture than an EDW architecture.*

Hypothesis 9b: *Organizations with upper management sponsorship is more likely to select an EDW architecture than a DBA architecture.*

The rest of the hypotheses each describe a unique relationship between a selection factor and the architectures. They are each presented next. Hypothesis four introduces the influence of task routineness on the choice between an EDW and a FED architecture. The research model that describes this hypothesis is displayed in Figure 4.4.

Hypothesis 4: *Organizations with tasks that are relatively more routine are more likely to select a FED architecture than an EDW architecture.*

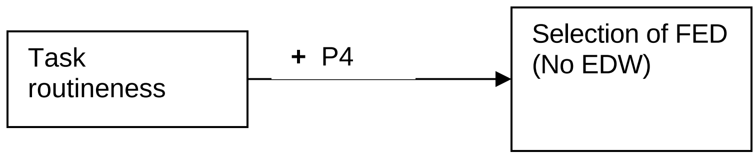


Figure 4.4: Research model for hypotheses: P4

The influence of resource availability on the selection of the DBA, IDM, or FED architectures when compared to EDW architecture is described in hypothesis seven. Figure 4.5 displays the research model that corresponds to this hypothesis. The influence of resource availability on IDM, DBA, and FED versus EDW architecture is expressed in the mini hypotheses 7a, 7b, and 7c respectively.

Hypothesis 7: *Organizations with low resources are more likely to select an IDM, DBA, or FED architecture than an EDW architecture.*

Hypothesis 7a: *Organizations with low resources are more likely to select an IDM architecture than an EDW architecture.*

Hypothesis 7b: *Organizations with low resources are more likely to select a DBA architecture than an EDW architecture.*

Hypothesis 7c: *Organizations with low resources are more likely to select a FED architecture than an EDW architecture.*

Figure 4.5: Research model for hypothesis: P7

Finally, the last two hypotheses on architecture selection, hypotheses 5 and 10, presents the impact of existing system compatibility and the influence of various sources of expertise on architecture selection. With each, the architecture selected depends on the existing systems or the influence of different sources of expertise. As a result, it is difficult to point out the specific data warehouse architecture selected from hypothesis 5 and 10.

Hypothesis 5*: Organizations in which existing systems are highly compatible with a particular data warehouse vendor product are more likely to select the data warehouse architecture that is aligned with that particular data warehouse vendor's products.*

Hypothesis 10*: Organizations with exposure to expert(s) are more likely to select the architecture aligned to the knowledge and services offered by the expert(s).*

4.2.3 Research question on architecture success

The second research question asked, "How successful are the alternative data warehouse architectures?" The third chapter discussed the success dimensions most suitable for this study and presented several specific hypotheses that describe some of the relationships between certain data warehouse architectures and success factors. However, the literature review of success in chapter three did not help fully understand the influence of each of the four data warehouse architectures on architecture success factors.

As a result, part of the study that was designed to find answers to research question two was more exploratory in nature to discover the impact of data warehouse architectures on success factors. As such, a general exploratory hypothesis was created to describe the current understanding of the overall relationship between the four data warehouse architectures and the success factors:

> ***Hypothesis E2:*** *The data warehouse architecture influences the success of the data warehouse development process and the system product success.*

The corresponding research model for the above hypothesis is presented in Figure 4.6.

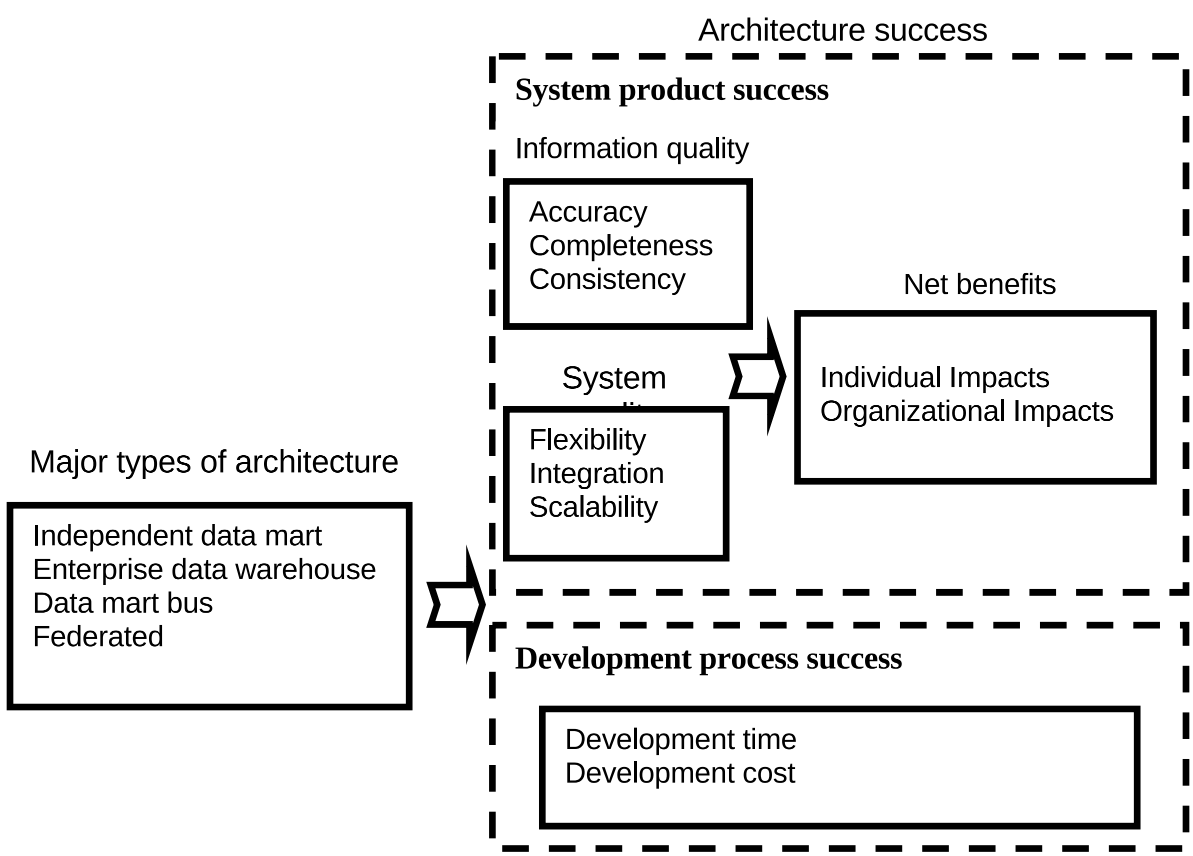


Figure 4.6: Data warehouse architecture success

Next, the impact of specific data warehouse architectures on success factors was considered. Chapter three presented the development of four specific hypotheses which describe the influence of specific data warehouse architectures on architecture success factors.

4.2.4 Hypotheses on data warehouse architecture success

The following hypothesis describes how flexibility (system quality factor) is affected by the use of the EDW architecture when compared to the IDM, DBA, or FED architectures. The research model that describes this hypothesis (i.e., hypothesis SQ1) is presented in Figure 4.7.

Hypothesis SQ1*: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than an IDM, DBA, or FED architecture.*

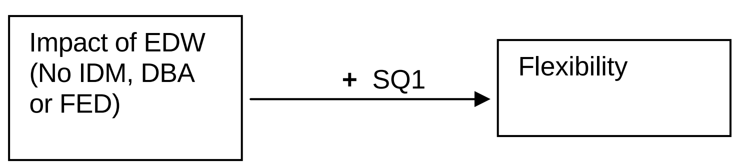


Figure 4.7: Research model for hypothesis: SQ1

Mini hypotheses SQ1a, SQ1b, and SQ1c describe the influence of EDW compared to IDM, DBA, and FED on flexibility.

Hypothesis SQ1a*: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than an IDM architecture.*

Hypothesis SQ1b*: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than a DBA architecture.*

Hypothesis SQ1c*: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than a FED architecture.*

The next hypothesis describes how integration (system quality factor) of the resulting data warehouse solution, is affected by implementing an EDW when compared to an IDM architecture. Figure 4.8 depicts the research model related to this hypothesis (i.e., hypothesis SQ2).

Hypothesis SQ2*: An EDW architecture is more likely to lead to the development of a data warehouse solution that enables high data integration than an IDM architecture.*

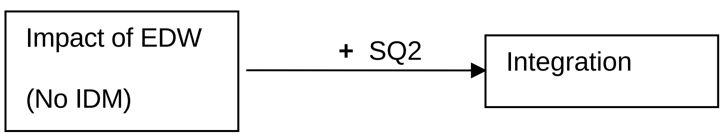


Figure 4.8: Research model for hypothesis: SQ2

How scalability of a data warehouse solution is influenced by implementing an EDW when compared to a DBA is explained in hypothesis SQ3. The research model that corresponds to this hypothesis (i.e., hypothesis SQ3) is presented in Figure 4.9.

Hypothesis SQ3*: An EDW architecture is more likely to be associated with the development of a scalable data warehouse solution than a DBA architecture.*

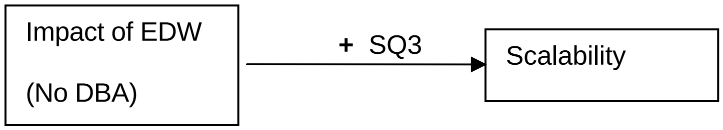


Figure 4.9: Research model for hypothesis: SQ3

Finally, hypothesis IQ1, describes how consistency (information quality factor), is affected by IDM architecture when compared to the rest of the architectures. The associated research model is graphically displayed in Figure 4.10.

Hypothesis IQ1*: An IDM architecture solution is less likely to be associated with high data consistency than an EDW, DBA, or FED architecture solution.*

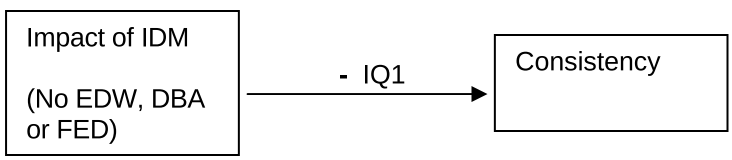


Figure 4.10: Research model for hypothesis: IQ1

Mini hypothesis IQ1a, IQ1b, and IQ1c describe how consistency is affect by IDM versus EDW, DBA, and FED.

Hypothesis IQ1a*: An IDM architecture solution is less likely to be associated with high data consistency than an EDW architecture solution.*

Hypothesis IQ1b*: An IDM architecture solution is less likely to be associated with high data consistency than a DBA architecture solution.*

Hypothesis IQ1c*: An IDM architecture solution is less likely to be associated with high data consistency than a FED architecture solution.*

4.3 Conclusion

This chapter first presented the hypotheses relevant to the first research question, which asked, “What is the relative importance of selection factors in the selection of a particular data warehouse architecture?” Next the chapter discussed the hypotheses relevant to the second research question, which asked, “How successful are the alternative data warehouse architectures?” For both architecture selection and success, it described the overall exploratory hypotheses and the specific hypotheses derived from theory, literature, and the expert interviews. It also presented the corresponding research models for each hypothesis. The fifth chapter describes the methodology and research design for the overall dissertation.

CHAPTER 5 – METHODOLOGY

5.1 Introduction

The current IS field is characterized by diversity in research methods (Benbasat et al. 1996). Several researchers have voiced their opinion on which methods are most appropriate to conduct IS research (Orlikowski et al. 1991). Ultimately, however, in designing a research methodology, a researcher should consider the phenomenon under study, the context, and balance the research needs with the practical limitations of the current methods available for conducting research (Mingers 2001).

This chapter describes the two-phased research methodology that was used to satisfy the research objectives. It addresses the research design, the sampling plan, the overall research process, and data analysis techniques.

5.2 Research Design

The research objectives, which drove the design for this study, were motivated by the need to describe the complex nature of data warehouse architecture selection and data warehouse architecture success. These research objectives required the use of methods that create a rich understanding of a complex phenomenon (Punch 1998). Furthermore, according to Lee (1991), the use of more than one research method results in a stronger research design because this approach takes advantage of the strengths of different methods.

The research design for this dissertation involved two phases: (1) qualitative interviews and (2) a quantitative field survey that provided a fuller understanding of data warehouse architecture selection and data warehouse architecture success. The following sections describe the research design for each of the two phases. A summary of the research design for both phases is provided in Table 5.1.

Table 5.1: research design

Phase One	
Purpose	Exploratory
Setting	Field
Timeframe	Cross-sectional
Approach	Qualitative interviews
Data collection method	Phone
Phase Two	
Purpose	Descriptive
Setting	Field
Timeframe	Cross-sectional
Approach	Survey
Data collection method	Web survey

5.2.1 Phase one: qualitative interviews

Selecting the appropriate architecture for IT infrastructure is a complex and challenging task (Ross 2003). Prior research has not identified or investigated specific variables related to data warehouse architecture selection, or how successful the alternative architectures are. As a result, the primary purpose of the first phase was to explore and identify the factors that affect data warehouse architecture selection and the factors to assess data warehouse architecture success.

Qualitative research methods allow a researcher to understand and explain the meaning of social phenomena with as little disruption of the natural setting as possible (Lincoln et al. 1985).

They generally involve investigating a small, nonrandom, purposeful sample of subjects for discovery, description, understanding, or hypotheses generating purposes. Merriam (1998) describes common types of qualitative research as basic, ethnography, phenomenology, grounded theory, and case study. She refers to basic studies as those that seek to discover and understand a phenomenon by identifying recurrent patterns in the form of themes or categories.

Data collection in basic studies can be conducted through interviews, observation, or document analysis. Personal interviews enable a researcher to clarify and probe to enhance the richness of research results (Merriam 1998). In addition, qualitative interviewing is often used in IS research in the initial exploratory phase of the investigation to gain a clear understanding of the phenomenon under study (Avison 1997). In order to discover the factors that affect data warehouse architecture selection and success, qualitative interviews were conducted.

The qualitative research tradition suggests that in order to discover, understand, and gain insight about a phenomenon of interest, the researcher must select a sample observation from which the most can be learned (Chein 1981). Similarly, to identify factors that affect data warehouse architecture selection, this study needed insight from individuals who have been involved in the process of architecture selection and have experience in data warehouse development. As an example, a recent study interviewed data warehouse experts such as renowned book authors, consultants, and experienced data warehouse managers to identify the factors that affect data warehouse success (Wixom et al. 2001). The expert interviews helped the authors generate hypotheses that were verified by quantitative analysis. This study used a similar purposive sample of data warehouse experts to examine data warehouse architecture selection and success.

Determining the adequate number of people to interview is generally challenging in qualitative research (Merriam 1998). Lincoln and Guba (1985) recommend sampling until a point of saturation or redundancy is reached. Specifically they claim, “In purposeful sampling, the size of the sample is determined by informational considerations. If the purpose is to maximize information, the sampling is terminated when no new information is forthcoming from new sampling units; thus redundancy is the primary criterion” (p.202, (Lincoln et al. 1985)). Following Lincoln and Guba’s advice, the number of data warehouse experts interviewed was limited to the point where no new information about data warehouse architecture selection and success was obtained.

The data warehousing experts were selected on the basis of being recognized authorities and spokespersons for the major architectures, leading consultants, highly recognized data warehouse managers, and Fellows of The Data Warehousing Institute. Appendix A presents the panel of experts interviewed in phase one. The results of phase one were described as part of chapter three.

5.2.1 Phase two: the field survey

Through the identification of the most salient factors applicable to data warehouse architecture selection and success in phase one, research models and hypotheses were generated to explore the relationships among selection factors, data warehouse architectures, and success factors. The second phase of this dissertation was conducted to examine the research models. The following subsections address the research design, the sampling plan, the research process, and data analysis techniques used.

5.2.1.1 Purpose

Researchers indicate that social research is generally conducted for exploratory, descriptive, or explanatory purposes (Babbie 2000; Punch 1998). Exploratory research investigates relatively new research areas in which variables are not well identified and defined. Researchers may also engage in research to describe situations and events in order to gain an accurate picture of a phenomenon under study. Descriptive studies look at the *what, when* and *how* questions in conducting social research (Babbie 2000). Finally, explanatory research seeks to explain relationships between phenomena under study.

The main purpose of the second phase of this study falls into the second category. Based on the findings from phase one, a research model was developed which indicated that certain organizational factors affect data warehouse architectures, and architectures vary with different success measures. Consequently, the primary purpose of the second phase was to validate the research model and describe the nature of the relationships between selection factors and data warehouse architectures, as well as the nature of the relationship between data warehouse architectures and architecture success. While the primary purpose of the study was descriptive, part of the second phase was to explain some of the relationships among data warehouse architectures, architecture selection, and success by testing several hypotheses.

5.2.1.2 Setting

Research can be conducted in either laboratory or field settings. Laboratory settings allow the researcher to have control over the phenomenon under study by manipulating the independent variables to observe different effects (Avison 1997). On the other hand, while field settings do not provide the degree of control provided by laboratories, they allow the researcher to collect information that reflects the true nature of phenomenon in a natural, real-world setting. In order to describe data warehouse architecture selection and success, this study required facts that reflected real-world data warehouse experiences and practices. As such, phase two was conducted in a field setting.

5.2.1.3 Timeframe

Researchers address timeframe considerations in their research in terms of cross-sectional or longitudinal research designs (Babbie 2000). A longitudinal study typically observes a small sample population over an extended period of time. It is characterized by relatively high investments in time, effort, and other resources. In contrast, cross-sectional studies observe and collect data for a population at a single point in time. The main problem with this design is that the researcher collects data at one point in time but generalizes the results to describe the nature of a social phenomenon overtime (Babbie 2000). Nevertheless, a cross-sectional timeframe was considered for the second phase due to its feasibility in terms of resources. In addition, a cross-sectional timeframe makes it possible to investigate a much larger number of organizations. The limitations of cross-sectional design were recognized as part of a discussion on the limitations of this study.

5.2.1.4 Approach

Classic IS literature describes three key methodologies used in empirical IS research: field studies, case studies, and laboratory experiments (Ives et al. 1980). Over time, the methodologies used in IS research have expanded. Today, quantitative IS research methods are categorized into

(1) field studies, (2) case studies, (3) laboratory experiments, and (4) field experiments (Boudreau et al. 2001).

Field studies often employ surveys that collect large amounts of data to examine the influence of one or more variables on a dependent variable (Avison 1997). Case studies focus on extensive data collection in a natural setting to capture the complexity of the problem. Laboratory experiments study the impact of dependent variables in a controlled environment. Finally, field experiments examine the impact of experimental manipulation on one or more variable in a natural setting.

Of these, a field survey approach was deemed most appropriate to address the objectives of this study. Compared to lab experiments, survey research examines phenomenon in a natural setting. Unlike field experiments, it does not attempt to control variables in natural settings to examine effects on the dependent variables. Finally, the researcher conducts survey research with clearly defined independent and dependent variables and a specific research model of the expected relationships at hand. Thus, unlike case studies, survey research is motivated to test expected relationships against observations of the phenomenon in a population.

In the past, survey research has been used for each of the three main purposes for conducting social research (i.e., exploratory, descriptive, or explanatory purposes). Descriptive survey research has been used in IS research to ascertain facts and to investigate situations, events, or attitudes that occur in a population (Pinsonneault et al. 1993). According to Pinsonneault et al (1993), survey research is most appropriate when the central question of interest are what, how and why questions. The main objectives of this study were to describe (1) *what* is the relative importance of selection factors in choosing a particular data warehouse architecture and (2) *what* is the relative success of different data warehouse architectures or *how* successful are the alternative data warehouse architectures? As such, the survey research approach was selected for this study as it presented the ability to observe a relatively large number of organizations in a real-world setting and to satisfy the research objectives.

5.2.1.5 Data collection method

There are several ways to administer surveys: face-to-face, telephone, mail, and via the Internet (i.e., e-mail or Web). While face-to-face surveys guarantee a high response rate, it is time consuming and may limit the study to a specific geographic area. Telephone surveys enable sampling a geographically dispersed population but can be very time consuming and costly. Usually mail surveys are less expensive to administer than either telephone or face-to-face surveys and enable the researcher to use long survey questionnaires if required (Babbie 2000). Recent studies comparing either email or Web surveys to paper-pencil mail surveys reveal that Internet-based surveys are relatively less costly than mail surveys (Dillman et al. 1999). Furthermore, electronic surveys reduce the time of conducting a survey and enables customization, randomization, and real-time changes to the questionnaire (Cook et al. 2000). The main concern in Internet surveying is coverage bias. Because the target population may not fully have Internet access, the sampled population may not represent the entire general target population (Couper 2000). However, when the general target population is a specific population in which respondents have high access to the Internet, coverage bias is less of a concern. As the next section (i.e., population and sampling) discusses, the potential respondents for this study are individuals who are familiar with the design, development and implementation of a data warehouse such as a data warehouse manager. As IT professionals, these individuals possess the required skills and access to the Internet. As such, Internet surveys were used for this study as it is relatively economical, feasible, and enables rapid data collection from a widely distributed sample population.

However, in addition to errors related to coverage biases, there are three other sources of error that survey design must overcome. They are sampling error, measurement error, and non-response error (Groves 1989). Compared to mail surveys, the use of online surveys is relatively new. A few researchers have presented guidelines on how to overcome potential sources of error when using Web surveys. Dillman et al. (1999) presents principles for the development of Web surveys. These are summarized in Table 5.2.

Table 5.2: Recommendations for Web questionnaire construction (Dillman et al. 1999)

- The first screen should be motivational, emphasize ease of responding, and provide instructions on actions to proceeding to the next page.
- Begin with a question that is fully visible on the first screen, which can be easily comprehended and answered by all respondents.
- Present questions in a conventional format similar to paper questionnaires.
- Limit line length to decrease likelihood of prose being extended across the browser screen.
- When the number of answer choices exceeds the number that can be displayed on one screen, consider double-banking.
- Use symbols or words to convey where the respondent is in the completion progress.

Dillman (2000) further suggests that many of the guidelines described for mail surveys apply to online surveys as well. For instance, a respondent-friendly questionnaire design, found to be important to improving response to self-administered mail questionnaires, is also important for the development of Web questionnaires. In the context of online surveys, a respondent-friendly Web questionnaire is one that is primarily compatible with the wide variety of computers and browsers possessed by respondents. Guidelines described by Dillman (2000) and Babbie (1998) were followed when constructing and administering the survey instrument.

5.3 Population and sampling

Sampling is the process of selecting the social phenomenon of interest from a population in a manner that enables the generalization of sample findings to the general target population (Babbie 2000). Sampling to achieve representativeness in the general target population can be achieved through many different strategies. However, the sampling logic must reflect the research purposes and questions of the study. For instance, when research questions focus on a relationship between variables, it is more appropriate to choose a purposive sampling strategy that maximizes the chances that the desired relationship will be observed (Punch 1998). The sampling plan should address the unit of analysis, the sampling frame, and the sample selection.

5.3.1 Unit of analysis

According to Babbie (2000), units of analysis are the individuals or entities whose characteristics researchers investigate in social research. Generally, the unit of analysis in social research can be an individual, a group, an organization, or a social artifact. The unit of analysis for this study was the data warehouse architecture of a completed data warehouse implementation. The person most closely involved with the design, development, and implementation of a data warehouse architecture, such as a data warehouse manager, *represented* the unit of analysis in the survey.

5.3.2 Sampling frame

The sample frame is a subset of the target population, which provides the basis for sampling the phenomenon of interest (Babbie 2000). In order to gain access to the general target population, many organizations and individuals were contacted to promote the study. As a result, the databases of several organizations and individuals (e.g., The Data Warehousing Institute's, Inmon Associates', and the Kimball Group) were identified and used as the sampling frame for this study.

5.3.3 Sample selection

Many of the organizations and individuals promoted the study using an electronic mailing, which described the study and provided a link to the Web-based survey. For example, TDWI sent an electronic newsletter to over 5,000 members. Inmon Associates promoted the study to 14,000 people on its electronic distribution list. Kimball Group sent out a message in their monthly 'Design Tip" to nearly 12,000 subscribers.

Since the online survey instrument was relatively long, study participants were offered a free copy of the study's findings and the opportunity to win one or four gift certificates from Sharper Image. Because of the role and recognition that the organizations promoting the study have in the data warehousing community, and the large number of potential respondents and companies that receive these mailings, it was believed that the respondents to the questionnaire would be representative of companies with data warehouses and the various architectures. Table 5.3 lists the various organizations and individuals that promoted the survey and the manner in which they promoted the survey.

Table 5.3: The organizations and the individuals that promoted the online survey

Sampling frame	**Method of promotion**
TDWI	e-newsletter (Flashpoint) Flyers at TDWI Boston Conference e-newsletter (BI This Week)

Bill Inmon	e-newsletter
Kimball Group	e-newsletter (Design Tips)
DM Review	e-newsletter (DM Direct) e-newsletter (Datawarehouse.com) Notice on DM Review Magazine July edition
Teradata	Notice on main Web page
Business Objects	Notice on Client Tech Support Web page
MicroStrategy	e-newsletter
DAMA International	Email announcement to data warehousing professional group (Microstrategy Direct)
Claudia Imhoff	Promotion to Intelligent Solutions Inc. clients
William McKnight	e-newsletter (McKnight Associates Inc.)

5.4 An overview of the research process

The overall research process for this dissertation involved two phases. First, in the initial phase, a panel of data warehouse experts was interviewed to further enrich the existing literature on selection factors and success measures. Second, using the expert interviews and existing literature, comprehensive research models were developed that described the salient selection factors that influence data warehouse architecture and measured architecture success. Consequently, in the second phase of the study, a survey was created using existing validated items whenever possible. When existing items did not meet the needs of the study, existing guidelines on survey construction (Babbie 1990) were utilized to generate multiple questionnaire items. Furthermore, when necessary, the wording of the existing items was modified to reflect the technology under study.

The online survey was constructed primarily based on feedback and recommendations from the panel of experts. Then, the questionnaire was reviewed and pre-tested by data warehouse experts for readability and content. According to Converse et al (1986), using a pilot sample of organizations that represents the target population to pre-test the survey is essential to the research process. Furthermore, previous research in data warehousing has also pre-tested with a group of organizations resembling the target population (Haley 1998). Consequently, twelve data warehouse managers (i.e., potential respondents of the survey) were asked to complete the online questionnaire and give reactions and the time taken to complete the survey. Based on feedback from the pretest, the online survey was estimated to take 15 minutes to complete. Next, the revised online survey was emailed to leading organizations and individuals in the industry requesting for their assistance to promote the survey to potential respondents. Then in May and June of 2004, those who offered to aid in data collection promoted the online survey to the target population using various communication channels such as electronic newsletters and listserv messages.

A series of statistical analyses were conducted on the data collected. First, descriptive statistics were generated for the independent variables and dependent variables. Next, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to identify potential problems and validate the overall measurement model. In order to assess validity and reliability, approaches previously suggested in the IS literature were used (Boudreau et al. 2001; Goodhue 1998).

Next, the data collected on the different selection factors and the architecture selected was analyzed using logistical regression to identify the relative importance of factors in the selection of a particular architecture. While part of this analysis was exploratory in nature, specific hypotheses linking particular architectures to selection factors were also tested. Finally, primarily using analysis of variance techniques, the architecture success factors were evaluated to identify the variations in data warehouse architecture success across different architectures.

5.5 Analysis techniques

Several data analysis techniques were utilized to analyze the data. The following subsections describe each of the data analysis techniques used.

5.5.1 Descriptive statistics

Descriptive statistics are used to describe the basic features of the data in a study. They provide a summary meaning of scores and the variance in observation. Mean and standard deviations were calculated on the survey data. Frequency distributions were calculated for each categorical variable. In addition to providing evidence of possible data entry mistakes and potential outliers, it also presented a preliminary understanding of the observations for each variable. Consequently, descriptive statistics presented a general overview of the demographic characteristics of the organizations, types of architecture included in the study, architecture success measures, and survey respondents.

5.5.2 Factor analysis

Factor analysis is an approach to examine the interrelationships among a large number of variables in order to condense the information in the original variables with minimum loss of the original information (Hair et al. 1998). It is designed to examine the covariance structure of a set of variables and determine the relationship between observed variables and underlying, unobserved constructs. Factor analysis is often used to verify the conceptualization of a construct of interest.

There are two types of factor analysis based on the intended purpose of the analysis: exploratory and confirmatory factor analysis (Stewart 1981).

5.5.2.1 Exploratory factor analysis

Exploratory factor analysis (EFA) is used to explore data to determine the number or the nature of factors that explain the covariation between variables (Stewart 1981). It is the most common form of factor analysis and is most appropriate when the researcher does not have sufficient evidence to form a hypothesis about the factor structure underlying the data. The data drives the discovery of the factor structure of the underlying data. Consequently, in the process of determining whether the identified factors are correlated, exploratory factor analysis evaluates construct validity (Kerby 1979).

In this study, EFA was used to gain an overall sense of the validity of the items on the survey instrument. The number of factors to retain was identified by the eigenvalues. An eigenvalue greater than one was used as the criterion to determine the factors to retain (Stewart 1981). Thus, EFA provided initial evidence confirming or disconfirming the factor structure suggested by theory in the measurement model. It also provided some insight into potential problems and alternatives to the hypothesized factor structure.

5.5.2.2 Confirmatory factor analysis

Confirmatory factor analysis (CFA) is used to determine if the number of factors and the factor loadings conform to what was suggested by theory (Stewart 1981). Thus, relationships can be hypothesized and tested using confirmatory factor analysis.

Indicator variables selected on the basis of theory are analyzed using factor analysis to determine if they load as predicted on the expected factor. Thus, it is a more viable method for evaluating construct validity. It enables a researcher to test the factor structure based on a model derived from theory that specifies the number and composition of the factors.

Structural equation modeling (SEM) was used to assess the measurement model and test improvements suggested by the EFA. While SEM is typically used to model causal relationships among latent variables (factors), it is equally possible to use SEM to explore CFA measurement models (Chin 1998). Indicator loadings estimate the strength of the relationship between a survey questionnaire item and a latent variable.

5.5.3 Logistic regression

Logistic regression is a specialized form of regression that is formulated to predict the impact of metric independent variables on a categorical dependent variable. While similar to discriminant analysis, logistic regression has the advantage of being less affected when the basic assumptions, particularly the multivariate normal distribution of variables, cannot be satisfied (Demaris 1992).

In logistic regression, the information on independent variables is used to predict the probability of an outcome (i.e., the category of interest). The regression coefficients are interpreted as either increasing or decreasing the probability of predicting the dependent variable. The estimated coefficient for an independent variable X_1 measures the change in the odds of being in a specific category of interest based on the one unit change in X_1 while controlling for all other independent variables (i.e., called Odds Ratio) (Agresti 1996).

In logistic regression, the test of global null hypotheses examines if all the regression coefficients are equal to zero. When rejected, the test indicates that at least one of the independent variables is helpful in predicting the probability of the dependent variable being a certain value. In addition, the Wald's chi-squares test statistic indicates the significance of each regression coefficient. However, authorities in logistic regression suggest that the likelihood ratio is a better statistical test to identify the significance of coefficients (e.g., Allison 1999, Agresti 1996).

Logistic regression comes in two basic forms: (1) binary logistic regression and (2) multinomial logistic regression. Binary logistic regression addresses the probability of predicting a categorical dependent variable restricted to two categories. Multinomial logistic regression allows for a categorical dependent variable with multiple categories.

Logistic regression is applicable to research questions with the objective of understanding group membership, whether the group compromises individuals, organizations, products, or any other object that can be evaluated through a series of independent variables. In the context of this study, logistic regression provides the basis to understand differences between architectures (i.e., a categorical dependent variable) in terms of variations in selection factors (i.e., metric independent variables).

In order to explore the influence of all ten selection factors on the four main architectures, a multinomial logistic regression analysis was conducted (i.e., to test hypothesis E1 which described the influence of the ten selection factors on the types of architecture). In addition, in order to gain more clarity of the impact of organizational factors on individual architectures, binary logistic regression analysis was utilized. For instance, using EDW architecture and the rest of the architectures as the binary categorical variable, binary logistic regression analysis was conducted to understand the impact of each independent variable on EDW architecture. A similar analysis was performed for each of the remaining four architectures. Finally, specific hypotheses that described relationships between specific organizational factors and certain architectures were also tested.

5.5.4 Multivariate analysis of variance

Multivariate analysis of variance (MANOVA) is used to explore the relationship between several categorical independent variables and two or more metric dependent variables. It evaluates whether an overall difference exists between groups. However, unequal sample sizes among groups affect the results of MANOVA.

According to Hair et al (1998), in order to use MANOVA, the sample size must exceed a minimum of 20 observations per cell. Before estimating mean differences between groups, there are several assumptions in MANOVA that must be tested: (1) correlation of dependent variables (using both univariate and multivariate tests) and (2) homogeneity of variance-covariance matrices (using Bartlett's test of sphericity). Finally, in order to assess the mean differences between groups, Hair et al (1998) recommends the following two test statistics: (1) Wilk's lamba and (2) the Pillai's trace.

In this study to explore and evaluate the differences in success of the four main categories of data warehouse architectures, MANOVA was used. In addition, several planned comparisons were performed to test specific hypotheses on success.

5.6 Conclusion

This chapter described the overall two-phased research design developed to explore the two research questions. Specifically, it discussed the methodology used to conduct the expert interviews in phase one and the methodology and issues relevant to the field survey in phase two. It presented the sampling plan and the overall research process for the dissertation. Finally, the chapter described the data analysis techniques used in this research. The next chapter first presents the operationalization followed by the validation of the survey instrument.

CHAPTER 6 – OPERATIONALIZATION

6.1 Introduction

This chapter describes how the study's constructs were measured using questions based on the academic literature, the practitioner literature, the analyses of the expert interview data, and feedback from the data warehousing experts. It also describes the instrument validation process used.

6.2 Construct measurement

One of the challenges of positivist quantitative research in information systems is accurately capturing and measuring the social entities under investigation (Straub 2004). Operationalization is the process of describing how a variable is observed and measured. A construct, which is an abstraction of a variable, represents what researchers actually measure or manipulate for research purposes (Babbie 1990). Well-developed constructs provide a researcher with a standardized language to communicate and understand other researchers' investigation of complex phenomena (Zmud et al. 1991).

Over the years, many researchers have presented guidelines to enhance the measurement accuracy of constructs (Grover et al. 1993; Sethi et al. 1991). When adequately validated items that represent a variable of interest are available, the use of existing measures is recommended. However, Swanson (1991) presents an interesting caveat about existing items. Existing research measures and questionnaires are deeply embedded in the research project that they pertain to. As a result, the context of such measures and questionnaires may not apply to a researcher's current project. It is necessary to exercise caution when adopting existing measures. Often, existing items serve only as useful starting points in operationalizing variables of interest (Zmud et al. 1991). Therefore, when pre-existing items that apply to the current context cannot be obtained from the literature, they must be developed.

Several sources of input guided the development of the constructs in this study. The existing literature and theory provided some conceptual understanding of the constructs of interest. However, the data warehousing literature and the interviews with experts often provided even more insights for shaping the constructs used in the study. Similarly, the development of the survey questionnaire items was primarily guided by the review of the data warehousing literature, the IS literature, and the analyses of expert interviews. Whenever possible, existing measures were adopted from the literature. When validated items that fit study objectives were unavailable, the items were generated.

In describing a process for developing measures for survey questionnaires, Hinkin (1998) suggests the use of the deductive approach to item generation when the researcher has a working knowledge of the constructs. The deductive approach is a process by which researchers generate the items through a thorough understanding of the phenomenon and by reviewing the literature (Hinkin 1995). In addition, Sethi et al (1991) advocates the use of an iterative approach to survey development, whereby items are further refined through each iteration of questionnaire evaluation. Using the working knowledge gained in the first phase of this study, multiple items were generated for the constructs of interest. Individual questions were constructed according to the recommendations of Babbie (2000). An attempt was made to create items that were brief, unbiased, and clear. The pool of all items were reviewed and refined in an iterative process with the help of other researchers.

After several iterations of initial review, the items were further refined through multiple iterations of review and feedback from a panel of leading data warehousing experts. More specifically, 20 experts reviewed the questionnaire for clarity and content.[4] While some provided comprehensive feedback through email, others opted to give feedback via follow-up calls. In addition, two experts went through the survey item by item while discussing their understanding of and answer to each question in the presence of the researcher. The expert reactions and suggestions were incorporated into the instrument. Finally, twelve data warehouse managers (i.e., potential respondents of the survey) were asked to complete the online questionnaire, give reactions, and indicate the time taken to complete the survey.

The section that follows presents the sample description, followed by the discussion of how the organizational selection factors, success variables, and the data warehouse architectures under investigation were operationalized. The operational definitions for the constructs in this study were previously presented in Table 4.2. Unless stated otherwise, the questions were asked using a seven-point likert scale anchored at strongly disagree (1) and strongly agree (7).

6.3.1 The sample

A purposive sample was employed for this study. The sampling objective was to capture a diverse group of organizations that have implemented various data warehouse architectures that are representative of the distribution of data warehouse architectures implemented in industry.

[4] The experts were Rob Armstrong, Karolyn Duncan, Wayne Eckerson, Vickie Ferrel, Jonathan Geiger, Jane Griffin, Douglas Hackney, Claudia Imhoff, Bill Inmon, Julie Kimball, Ralph Kimball, Pieter Mimno, Jim Revak, Margy Ross, Don Stoller, Ron Swift, Jim Thomann, Warren Thornthwaite, David Wells, and Todd Walters.

Demographic information about the respondent organizations was collected so that the sample as a whole could be described. The following questions with categorical responses were asked:

- Please indicate the location of your data warehouse: state/province and country
- Please check the activity that best describes the primary business of your company
- Please estimate the approximate 2004 gross revenues or operating budget of your company
- Please estimate the approximate 2004 number of employees in your company

These items were adopted from previous surveys on data warehousing and decision support systems (Eckerson 2002; Eckerson 2003; Watson et al. 1995; Wixom et al. 2001).

In addition, several questions captured information about the survey respondent and descriptive information about the data warehouse in the respondent's company. They were adopted from data warehousing and decision support systems surveys (Eckerson 2002; Eckerson 2003; Watson et al. 1995; Wixom et al. 2001). The questions about the respondent with categorical answer options were:

- Which of the following best describes your position in the organization?
- Were you actively involved in the selection of the data warehouse architecture in your organization?

The items on the data warehouse were:

- How long has the data warehouse been in use?
- How much raw data is in your data warehouse?

6.3.2 Data warehouse architectures

Despite the descriptions of the various data warehouse architecture in both the academic and practitioner literature, there is considerable confusion as to what the commonly used labels for each architecture mean. Often, two individuals may recognize a name or label that represents two very different architectures.

As a result, early in the process of developing a questionnaire, it was decided that the label used by the authority of each architecture, along with a simple diagram of the components of the architecture, would be used to describe each data warehouse architecture. In addition, in the survey instrument a detailed description of the various architectures was hyperlinked to the label to provide more information. Working closely with the authorities of each architecture and the other experts, the diagrams and the detailed descriptions were created. The question that resulted asked respondents to identify the architecture that most closely represented the one implemented in their organization. It provided data warehouse architecture labels and a diagram of each architecture.[5]

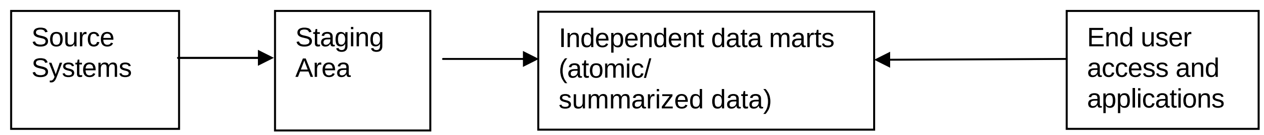
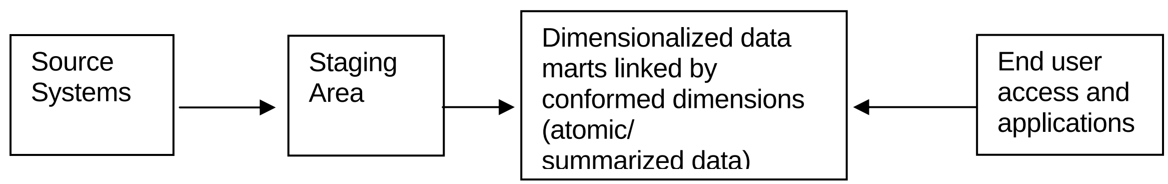
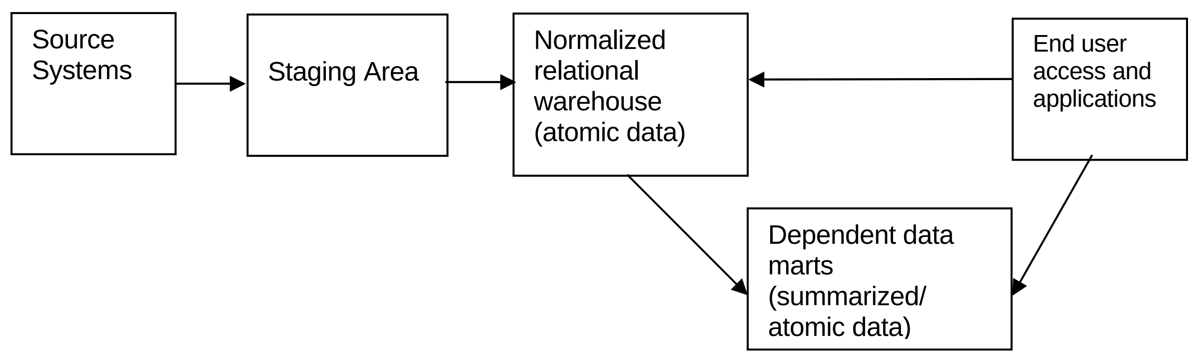
It was also deemed necessary to discover how closely a company's architecture matched the data warehouse architecture selected in the survey. As a result, a follow-up question ascertained the extent to which the architecture selected matched the architecture in the respondent's organization.

All in all, two questions were developed to identify the architecture implemented in respondents' organizations. The first asked them to identify the architecture currently in place in their organizations by providing familiar labels and a diagram of each architecture. The second question attempted to assess how closely the response selected in question one matched the architecture in their organization. Each of these questions is presented next.

---

[5] In presenting questions on data warehouse architectures, past studies and surveys have presented the EDW architecture category as two separate response categories. The two categories are hub and spoke architecture (HSA) and centralized data warehouse architecture (CDW). Both categories are identical except that the CDW architecture does not include dependent data marts. The two categories depict a structural difference but are essentially the same architecture. In past surveys, this presentation of the EDW architecture has enabled survey respondents to easily identify the EDW architecture. In a similar manner, the EDW architecture was presented as two separate response categories in this study and was later combined for data analysis purposes. Prior to combining the two responses categories, independent sample t-tests were run for all variables for the two groups. No significant difference between the two response categories was found.

- Please indicate which of the following best describes your current data warehouse architecture?

DATA WAREHOUSE ARCHITECTURE	COMPONENTS OF THE ARCHITECTURE
1. Independent data marts	
2. Data mart bus architecture with linked dimensional data marts	
3. Hub and spoke architecture, Corporate Information Factory, Enterprise data warehouse	

4. Centralized data warehouse, Enterprise data warehouse **(no dependent data marts)**	Source Systems → Staging Area → Normalized relational warehouse (atomic/ summarized data) ← End user access and applications
4. Federated	Existing Data Warehouses, Data Marts, and Legacy Systems → Logical/physical integration of common data elements ← End user access and applications
5. I do not know/ None of the above	

- How closely does your data warehouse architecture match the one you selected in the previous question?

Slight match			Moderate match			Exact match
1	2	3	4	5	6	7

6.3.3 The organizational selection factors

The third chapter described ten organizational factors that influence the selection of a data warehouse architecture. The operationalization of each is described below.

6.3.3.1 Information interdependence

Developing measures of interdependence is a formidable task. Due to the challenges of measuring interdependence in multiple organizational domains, most past research has focused on interdependence within a specific organizational domain of interest (Gattiker et al. 2003; Wybo et al. 1995). For example, Gattiker et al. (2003) limited their investigation to subunits within organizations. A data warehouse architecture can be implemented within different domains of an organization. As a result, this study evaluated the characteristics of the organizational setting of an architecture implementation within different domains of an organization. Limiting the scope of the study to a specific organizational domain (e.g., data warehouse architecture implementations in subunits within organizations) is one alternative to measurement; however; it limits the generalizability of the study findings to a particular domain. In addition, it narrows the sampling frame of organizations to firms with architecture implementations in a specific domain. As a result, the instrument included an item designed to capture the organizational domain in which the architecture was implemented. It was followed by questions on interdependence customized for each organizational domain.

There are many classifications of organizational domains or levels. Varadarajan et al. (2001) describe levels of a firm as corporate, business, and functional. Tushman et al. (1978) describe interdependence at the subunit level of the organization. Goodhue et al. (1988) describe domain as departmental, divisional, and corporate. In the context of this study, organizational domain is the business area or scope in the company for which the data warehouse architecture was implemented. Recent industry data warehousing reports published by The Data Warehousing Institute measured the various domains of data warehouse implementation as corporate, business unit, functional area, or subunit level (e.g., Eckerson 2003). Based on a review of the literature, insight from a committee member, and feedback from data warehousing experts, a question was created to ascertain the domain of the organization in which the data warehouse architecture was implemented. The question incorporated the four different types of domain described in past research, specifically the categorization described by Eckerson (2003). However, the final question had five response options. One more response option was added to capture the possibility that several but not all business units may encompass the domain of an architecture implementation.

- What is the domain of the business for which the data warehouse architecture was implemented (i.e., the business area or scope in the company for which the architecture was implemented)?
    a. Entire company (e.g., General Motors Corporation)
    b. Several but not all business units within the company (e.g., the Saturn and Chevrolet automobile divisions at General Motors)
    c. A single business unit with several functional areas (e.g., the Saturn automobile division at General Motors)
    d. A single functional area unit within a business unit (e.g., the marketing area within Saturn automobile division)
    e. A single subunit within a business unit (e.g., the used car fleet within Saturn automobile division)

A review of the literature revealed studies that describe several dimensions of interdependence (Gattiker 2000). McCann and Ferry (1979) present six dimensions of transactional interdependence including frequency of exchange and tolerance of slack resources. Wybo and Goodhue (1995) introduce perceptions of interdependence as another important measure of interdependence. The data warehouse manager, or the person most familiar with the selection of an architecture within an organization, may not be extremely knowledgeable and familiar with the inner workings of organizational units. For example, the he or she may not know the exact frequency of exchange between organizational units. As a result, the respondent is more likely to be able to answer questions on a perception of the information interdependence between units. Three items from the Wybo and Goodhue (1995) study were adapted with modification to measure the perception of information interdependence.

As mentioned before, the presentation of the items on interdependence were customized based on the domain. If the respondent indicated that the domain of the business for which the data warehouse architecture was being implemented was the entire company or several but not all business units, the following questions were asked.

- Close coordination among business units within the company was essential for them to successfully do their work.
- The decisions and actions of every business unit had important implications for the operation of the other units within the company.
- Information provided by other business units within the company was critical to each unit’s performance.

If the respondent stated that a single business unit, a single functional area, or a single sub unit was the domain for the architecture implementation, the same questions were asked with a slight modification. Any reference to domain in the questions was customized to match the domain they selected.

6.3.3.2 Vertical information flow

A data warehouse architecture implementation satisfies upper management's needs for information from lower levels of the organization by providing specific capabilities. A data warehouse can provide the ability to "drill down" to data elements at lower levels of the organization as well as provide the ability to "roll up" data to present an aggregated view (Kimball 1996b). Furthermore, in recent years, the data warehouse has become a vehicle that helps upper management adhere to regulations imposed by the SOX Act of 2002 (Lesar 2004). During expert interviews, Jane Griffin stated that dashboard front ends of data warehouses are providing top executives with drill down functionality to get at the detailed data they need.

A review of the literature revealed no validated items that measure the extent of upper management's information needs that can be satisfied by a data warehouse architecture implementation. Based on the data warehousing literature (Inmon 2001; Gray et al. 1998; Devlin 1997; Kimball 1996a) and feedback from data warehousing experts, four items were created to measure this construct. Of the four items, three were developed to measure the need to drill down and aggregate lower level information and one was developed to measure the need to adhere to SOX compliance.

Based on the domain for which the architecture was implemented, the vertical information flow questions were also customized in the final instrument. If a respondent indicated that the domain was the entire company or several business units within the company, any references to domain in the questions read as "entire company." Consequently, if the respondent selected any of the other domains as the domain for their data warehouse, any reference to domain in the questions read as "the business unit." The questions used to measure vertical information flow customized to the domain of a single business unit are as follows.

- Senior management of the business unit needed the ability to drill down into data from lower organizational levels in order to carry out their job responsibilities
- Upper management of the business unit needed the capability to drill through to detailed atomic level data.
- Senior management of the business unit needed an aggregated view of data from lower levels of the organization.
- In order to comply with regulations (e.g., Sarbanes-Oxley Act), upper management of the business unit needed a complete, accurate view of company information.

6.3.3.3 Urgency

Urgency has been described and measured in several literature streams in the social sciences. In psychology, time urgency is presented and operationalized as a multidimensional construct that has been studied at individual, group, and organizational levels (Conte et al. 1995; Landy et al. 1991). For instance, in organizational psychology, time urgency is defined as "the importance of completing the most amount of work in the shortest amount of time" (Maierhofer et al. 2000). In management research, Nutt (1993) describes urgency as time pressure to decision making. These conceptualizations of urgency in these literature streams did not match the needs of this research.

The IS research has also recognized that the urgency of the business need to implement information systems has an impact on the decision to implement a particular technology or system (Allen et al. 1991). The literature describes and measures numerous causes of the business need to develop a system. Environmental dynamism is one such cause that has received much attention in the literature (e.g., Bajwa et al. 1998). However, a review of IS literature did not present items that measure *the extent of urgency* of the business need.

The expert interviews shed some light on how to conceptualize and measure urgency to fit the needs of this study. For instance, Jim Thomann described urgency as “pressure to quickly meet business needs.” Based on the data warehousing literature, the expert interviews, and expert feedback, three items were created to measure urgency.

- There was pressure to build the data warehouse quickly.
- The business needs demanded a fast implementation of the data warehouse.
- A fast project turnaround time was critical in developing the data warehouse.

6.3.3.4 Task routineness

Based on the work of Goodhue (1995) and input from the expert interviews, task routineness was operationalized as the extent to which users’ jobs required non-routine data analyses. The review of the data warehousing literature revealed two studies that had previously measured the nature of tasks. Shin (2003) describes four items that measure task characteristics, but none of them fit the definition of task routineness in this study. Brohman (2000) primarily used items from Goodhue (1995) with slight modification to measure task routineness. Based on feedback from experts, three items were used to measure task routineness by modifying the items from Goodhue (1995) and Brohman (2000).

- Users often had questions that could not be addressed by structured queries and standard reports.
- Users often had to answer questions that were novel and unique.
- Users often faced questions that they had never answered before.

6.3.3.5 Compatibility with existing systems

The literature on information processing theory suggests that existing organizational capabilities for information sharing has an impact on the IS structure chosen to meet organizational information sharing requirements (Tushman et al. 1978; Lee et al. 1992). During the expert interviews, both Doug Hackney and Ron Swift described examples of how existing vendor platforms and systems impact the choice of the data warehouse architecture. Using insights from the academic literature and expert interviews, one item was created to measure the extent to which the data warehouse architecture was compatible with existing systems.

- Being compatible with the systems (e.g., applications, databases, ERP systems) that were already in place influenced the architecture that was implemented.

6.3.3.6 View of the data warehouse

Weill et al. (1998) describe an organization's view of IT infrastructure as none (i.e., satisfying a single business unit's need), utility (i.e., providing IT services at minimum cost), dependent (i.e., dependent on business strategy), and enabling (i.e., providing future options to implement strategy) (Broadbent et al. 1997). Of the four views, case studies of data warehouse implementations (e.g., Watson et al. 2001; Watson et al. 2002) and expert interviews indicated that none and dependent are the most applicable and influential views or perceptions of a data warehouse implementation. As a result, based on the conceptual understanding of the work by Weill and Broadbent (1998), the data warehousing literature, and expert interviews, three questions were created to measure 'none' and 'dependent' views.

- The implementation of a data warehouse was viewed as a point solution to meet a functional area unit(s) or subunit(s) need.
- The implementation of a data warehouse was viewed as an infrastructure project to support a range of applications.
- The implementation of a data warehouse was viewed as a solution to support strategic objectives.

6.3.3.7 Resource Availability

In IS research, previous studies have measured the extent to which the development of a system is constrained by both time and financial resources (e.g., Tait et al. 1988). In data warehousing, Haley (1998) describes and measures the availability of time, money, and people as the three key resources needed for a data warehouse implementation. The practitioner literature indicates the importance of having both adequate business personnel and IT personnel to develop a data warehouse (Inmon 2001; Kimball et al. 2002). As such, the key resource 'people' was further categorized to include IT and business unit personnel. In addition, the impact of time constraints on architecture selection was previously introduced in this study as a separate factor named "urgency." The expert interviews and feedback further confirmed that the choice of architecture may be constrained by money, IT personnel, and business personnel.

Consequently, the respondent's perception of resource availability in terms of money, IT personnel, and business personnel was measured in this study. In order to determine how these resources constrain architecture selection, three questions were created.

- The availability of business personnel constrained the choice of architecture.
- The availability of IT personnel constrained the choice of architecture.
- The availability of monetary resources constrained the choice of architecture.

6.3.3.8 The perceived ability of the in-house IT staff

The perceived ability of the in-house IT staff was defined and measured through the input from several sources. A review of studies in the organizational learning and self efficacy literature indicated that past experience and existing skills influence the performance of computer-related activities. Based on Agarwal et al.'s (2000) conceptualization of specific software self efficacy and Bandura's (1977) conceptualization of enactive mastery, a question was developed to measure the respondent's perception of the in-house IT staff's prior experience in implementing a data warehouse. Next, using Fichman et al.'s (1997) work on the influence of existing skills and knowledge on lowering knowledge barriers systems and Haley's (1998) assessment of the need for right technical skills to development a data warehouse, an item was created to capture the respondent's perception of the extent to which the in-house IT staff had relevant technical skills for developing the data warehouse.

During the interviews and feedback sessions, several experts stressed the importance of the development team's confidence in their ability to implement a data warehouse as influencing architecture selection. Consequently, a third question was developed to estimate how the respondent perceived the in-house IT staff's confidence in their ability to develop a data warehouse.

- The in-house IT staff had the necessary technical skills for developing the data warehouse.
- The in-house IT staff had prior experience successfully implementing a data warehouse.
- The IT staff was confident that it could successfully implement the data warehouse.

6.3.3.9 Source of sponsorship

Previous studies in information systems have measured the various sources of sponsorship for the implementation of a decision support system (e.g., Watson et al. 1991). Based on input from experts and the literature, a single question was created to determine the source of sponsorship for the data warehouse. The question contained the possible sources of sponsorship based on the work of Watson et al. (1998) and Eckerson (2003). The resulting question contained answer categories that ranged from a single business unit to sponsorship permeated throughout the organization.

- The main source of sponsorship for the data warehouse:
    a. was from a single functional area unit or a subunit.
    b. was from a single business unit.
    c. was from IT.
    d. was from multiple business units.
    e. had senior business management (i.e., CXO) support.
    f. sponsorship permeated the entire company.

6.3.3.10 Expert influence

Past research in management describes several sources of external influence on decision making in organizations, including consulting firms, mass media, and business schools (Abrahamson 1991). In the data warehousing realm, consultants have been recognized as a key external expert that influences data warehouse implementations (Hackney 2000a). In addition, during expert interviews, Wayne Eckerson indicated that attending data warehousing conferences or reading a book authored by a leader in the field can influence an organization to select a particular architecture. Based on the insights from the academic literature and expert interviews, three items were created to measure the extent to which the organization's choice of data warehouse architecture was influenced by consultants, the data warehousing literature, and attendance at data warehousing conferences and seminars.

- The choice of data warehouse architecture was strongly influenced by data warehousing consultants.
- The choice of data warehouse architecture was strongly influenced by the data warehouse literature.
- The choice of data warehouse architecture was strongly influenced by attendance at data warehousing seminars and conferences.

6.3.4 The architecture success factors

The IS literature is abound with studies that have described and measured IS success (Delone et al. 2003). In recent years, there has been a several studies that have assessed data warehousing success by adapting existing measures of IS success and by developing new items of success that reflect the data warehouse environment (Wixom et al. 2001; Shin 2003). *Architecture* success is an area of IS success not previously measured in the literature. As a result, the expert interviews and feedback sessions were used to adapt existing items and develop new questions to obtain the final pool of questions to measure *data warehouse architecture* success.

In the literature review chapter, architecture success was described as being composed of product success and development process success. The three dimensions of product success relevant to architecture success were identified as information quality, system quality, and net benefits. Development process success was described in terms of two dimensions: development time and development cost. The operationalization of each of them is described below.

6.3.4.1 Information Quality

The IS literature describes several dimensions of information quality (e.g., Bailey et al. 1983). The expert interviews helped identify the dimensions of information quality most appropriate for measuring the respondent's perception of data warehouse architecture success. They were accuracy, completeness, and consistency. Three questions were used to measure each dimension.

Of the three questions on accuracy, two of them were created based the work of a leading authority on information quality, Larry English. His definition and description of information quality was used to create the items (English 1999). Another item was adapted from Goodhue (1998) with slight modification to reflect the data warehouse environment.

- Data warehouse structured queries and reports contain few data errors.
- Your data warehouse provides the level of data correctness needed for its intended purpose.
- The data in the data warehouse correctly represent the real world objects and events.

In data warehousing, Wixom et al (2001) measured completeness using two questions adapted with modification from Ives et al. (1983) but later dropped both due to low reliability. Consequently, the original measures from Ives et al. (1983) were adapted with modification based on input from the expert interviews and feedback. A third question was based on the definition of completeness from Jarke et al. (1999) and input from experts.

- Your data warehouse architecture includes data about all the business processes and subject areas that are required by users and applications.
- All necessary decision support data is available within the data warehouse.
- Your data warehouse architecture provides all the data needed by users and applications.

Vaduva et al.'s (2001) definition of data consistency was the foundation for the development of all three questions on consistency. In order to create items that reflected data warehouse practitioners' language for data consistency, phrases from the expert interviews and data warehouse practitioner literature were used in the questions. Specifically, two questions were created based on the writings of Inmon (1999; 2001). Another measure was based on a statement made by Rob Armstrong during his expert interview.

- Your data warehouse architecture provides "a single version of the truth."
- Your data warehouse architecture reduces data inconsistencies.
- Your data warehouse architecture provides "a single system of record" for decision-support data.

6.3.4.2 System Quality

The IS literature and expert interviews helped identify the dimensions of system quality that are the most applicable for measuring architecture success: flexibility, scalability, and integration. Questions were developed to identify the respondent's perception of system quality of the data warehouse architecture implementation.

The definitions and items on flexibility used in previous study's (e.g., Ives et al.1983; Duncan 1995; Goodhue 1998) were too general to assess the flexibility of a data warehouse architecture. For instance, Ives et al. (1983) has an item that assesses a system's ability to flexibly adjust to new demands. This question does not tap into new demands specific to a data warehouse environment, such as the ability to add new business processes and subject areas. The expert interviews and practitioner literature presented language and phrases that described flexibility that matched the context of this study. As such, with past literature as guidance, three items were developed to measure flexibility incorporating verbiage specific to data warehousing. A similar approach was adopted to create items for scalability and integration.

- Your data warehouse architecture makes it is easy to add new business processes and subject areas.
- Your data warehouse architecture provides the capability to satisfy new requirements quickly.
- Your data warehouse architecture provides the capability to easily support future application needs.

Using industry success stories and white papers of data warehousing implementations (Bhend 1999; Ciol 2002; Heinecke 2000) and expert interviews, three aspects of scalability for a data warehouse were identified. They were increases in the number of users, the volume of data volume, and the complexity of queries. A question was created to assess each aspect of these aspects of scalability.

- Your data warehouse architecture is scalable to handle increases in the number of users without negatively impacting system performance.
- Your data warehouse architecture is scalable to handle increases in the complexity and number of simultaneous queries without degrading system performance.
- Your data warehouse architecture is easily scalable to handle increases in the volume of data.

Integration is a key distinguishing feature of a data warehouse (Inmon 1992). The academic literature (e.g., Galbraith 1974; Goodhue 1992), practitioner literature (e.g., Hoss 2001; Leon 2003), and the expert interviews offered a conceptual representation of integration as it applied to this study. Drawing on these sources, three questions were generated to measure integration. Of the three, one was based on the definition of integration described by Gattiker et al. (2003). A second question was based on comments made by Rob Armstrong and Margy Ross during the expert interviews and feedback. The final item was founded on a definition of integration presented in Gray et al. (1998).

- Your data warehouse architecture supports and facilitates the integration of data from multiple systems.
- Your data warehouse architecture supports and facilitates the integration of internal and external data sources.
- Your data warehouse architecture supports and facilitates the integration of all needed data around primary keys.

6.3.4.3 Net benefits

Individual impacts and organizational impacts were identified as the two benefit categories used to estimate the value gained by implementing a data warehouse architecture. As mentioned previously in the literature review, the experts strongly recommended the addition of these categories to assess architecture success. In recent years, researchers have generated scales to assess the impacts of IS (Jiang et al. 2002; Torkzadeh et al. 1999) as well as the benefits of data warehousing (Wixom et al. 2001). Some of the items were adapted from past research. In order to create measures that indicated variations in benefits due to differences in the implementation of data warehouse architectures, interview data and language from practitioner literature were also used.

Four items were used to assess the respondent's perceptual assessment of individual impacts. Three of the items were based on expert interviews and the dimensions of individual impact described by Torkzadeh et al. (1999). Another question was adopted with modification from Wixom et al (2000).

- Users can access the data more easily and quickly because of the data warehouse.
- The data in the data warehouse is easy and intuitive for users to understand and use.
- Your data warehouse enables users to think about, ask questions, and explore issues in ways that were previously not possible.
- Your data warehouse has improved the decision-making capabilities of end users.

Three questions on organizational impacts were adopted with modification from Haley's (1998) research on data warehousing. The rest were founded on the comments from several expert interviews. Both Julie Kimball and Rob Armstrong stressed the importance of improved communication and coordination from data warehouses, while Ron Swift mentioned gains from a high, measurable return on investment (ROI). Finally, Wayne Eckerson, Rob Armstrong, and Dave Wells pointed out a data warehouse implementation can lead to the widespread adoption of business intelligence in an organization. Overall, six questions were used to measure the respondent's perceptions of organizational impacts of data warehouse architecture implementations.

- Your data warehouse architecture has met the business requirements for which it was implemented.
- Your data warehouse architecture has greatly facilitated the use of Business Intelligence.
- Your data warehouse architecture has enabled improvements in business processes.
- Your data warehouse architecture has supported the achievement of strategic business objectives.
- Your data warehouse architecture has led to high and measurable ROI.
- Your data warehouse architecture has improved communication and cooperation across organizational units.

6.3.4.4 Development time

Data warehousing is often described as "a journey not a destination," indicating its evolving nature overtime. It is difficult to assess the total time at which it is completed, since it is continually evolving. As a result, it is necessary to assess the time taken for key milestones.

As mentioned previously in chapters one and two, data warehouse architectures are implemented incrementally in an iterative manner. Organizations often use the time taken to implement the first subject area or first business process as one of the key measures that indicate the progress made towards developing a data warehouse (Hackney 1998). Furthermore, previous studies assessing development time have used questions with categories that measure time taken to implement, as well as questions with categories that measure time taken relative to an established schedule (Wateridge 1998; Keil 1995). In data warehousing, Eckerson (2003) presents questions on development time with answer choices that measure time ranges applicable to data warehouse implementations. One question was developed to measure development time based on the work of Eckerson (2003).

- How much time was required to develop and roll out the first business process(es) or subject area(s) in the architecture?
    a. 3 months or less
    b. 4 – 6 months
    c. 7 – 12 months
    d. 13 – 24 months
    e. Over 24 months
    f. Not sure

6.3.4.5 Development cost

Similar to assessing the time taken to implement a data warehouse, estimating the cost is challenging. It is difficult to compute a total cost of development because implementing a data warehouse is an ongoing process. Much like appraising the time taken, organizations assess the cost of completing key milestones at different points in time. The data warehousing experts also stressed that measuring cost required looking at development costs at key points in time as data warehouses are developed incrementally. As a result, the interviews and data warehousing literature were analyzed to discover specific ways to measure development cost. Consequently, three specific types of cost relevant to architecture success were identified. They were: the cost of the first subject area or business process, overall cost to date, and annual cost. Using dollar cost categories from Eckerson's (2003) work on development cost, three questions were developed that addressed the three types of cost.

- What was the cost (US $) (e.g., hardware/software, personnel) of developing the first business process(es) or subject area(s) ?
  a. Less than 100,000
  b. 100,000 – 500,000
  c. 500,000 – 1 million
  d. 1 million – 5 million
  e. 5 million – 10 million
  f. Above 10 million
  g. Not sure

- What is the annual cost (US $) (e.g., hardware/software, personnel) of maintaining (not enhancements) the architecture?
  a. Less than 100,000
  b. 100,000 – 500,000
  c. 500,000 – 1 million
  d. 1 million – 5 million
  e. 5 million – 10 million
  f. Above 10 million
  g. Not sure

- What is the cost to date (US $) (e.g., hardware/software, personnel) of developing (both maintenance and enhancements) the overall architecture?
  a. Less than 100,000
  b. 100,000 – 500,000
  c. 500,000 – 1 million
  d. 1 million – 5 million
  e. 5 million – 10 million
  f. Above 10 million
  g. Not sure

6.4 Instrument Validation

Prior to performing most univariate and multivariate statistical analyses, it is essential that researchers demonstrate that the instrument used to measure phenomena of interest meets minimal standards of validation (Straub 1989). The two central issues that are addressed during instrument validation are reliability and validity.

6.4.1 Reliability

Reliability refers to the extent to which a measurement item on a scale is free from random error (Kerlinger 1986; Nunally 1978). According to Babbie (2000), it refers to the consistency or stability with which an instrument produces results. The most commonly accepted and recognized measure of reliability is Cronbach’s alpha (Bollen 1989; Sethi et al. 1991; Straub 1989). Low reliability can create problems in the analyses of results. For instance, it increases the possibility of making a Type II error. According to Gefen et al. (2000), alpha values of .60 are a commonly used threshold value for acceptable reliability in exploratory studies. An alpha value of .60 was used in assessing the reliability of the instrument in this study.

6.4.2 Validity

According to Babbie (2000), validity refers to the extent to which an item sufficiently reflects the true meaning of the concept under consideration. While reliability is concerned with the consistency of results, validity is concerned with getting results that accurately reflect the construct being measured. It is usually assessed in terms of content validity, construct validity, and external validity.

6.4.2.1 Content validity

Content validity assesses how well the scale measures what it is supposed to measure. Babbie (2000) suggests that content validity indicates if an item covers “the range of meanings included within a concept” (pg. 144). According to Straub (1989), it can be established through literature reviews and expert judges or panels. More specifically, he recommends using several rounds of pre-testing the questionnaire with different sets of experts. Following Straub’s advice, the questionnaire was developed using a set of expert interviews and a literature review. Next, several pre-tests were conducted with experts to establish content validity (Previously described in 6.2 Construct Measurement).

6.4.2.2 Construct validity

Construct validity is defined as the degree to which an item measures what it claims to be measuring (Nunnally 1978). There are three variants of construct validity: convergent validity, discriminant validity, and nomological validity.

Convergent validity refers to the extent to which measuring the same construct using two or more different methods are in agreement. More recently, researchers have described convergent validity as the degree to which "items thought to reflect a construct converge or show significant, high correlations with one another, particularly when compared to the convergence of items on other constructs, irrespective of method" (Straub et al. 2004, p.391). Discriminant validity refers to the extent to which items measuring different constructs can be distinguished. When items measuring the same construct correlate significantly higher than with items reflecting other constructs, discriminant validity is achieved (Straub 1989). Using the recommendations of past researchers, convergent and discriminant validity were assessed using confirmatory factor analysis.

Finally, nomological validity describes the extent to which predictions about the constructs of interest within a formal theoretical network are confirmed. In domains where theories and measures are not well-developed and a formal theoretical network is not well established, nomological validity is hard to assess (Goodhue 1998). In such domains, predictive validity is calculated. By confirming previously unproven but theoretical defended predictions about the constructs of interest, confidence in constructs of interest is enhanced. As a result, specific hypotheses based on existing theory were tested in this study.

6.4.2.3 External validity

External validity refers to the generalizability of results (Straub et al 2004). External validity can be attained through proper sampling methods that enhance the representativeness of the sample collected. While purposive sampling was used in this dissertation, the following methods were used to ensure external validity:[6]

- Authorities for each major architecture were asked to send out an electronic message to promote the study.
- Several major vendors of data warehouse products sent out a message to client listservs.
- One of the main magazines and the leading institution within the data warehousing space promoted the study to their subscribers.

A wide representation of organizations that have implemented different data warehouse architectures was obtained in this manner.

6.5 Conclusion

This chapter first described the manner in which the constructs of interest were operationalized based on expert interviews, feedback from experts, theory, and the literature. Next, the instrument validation process used in this study in terms of validity and reliability of measures were described. The chapter that follows presents the results of the statistical analysis of the data collected.

[6] Table 5.3 lists all the organizations and individuals that helped promote the study to gain a sample representative of the industry population.

CHAPTER 7 – DATA ANALYSIS

7.1 Introduction

This chapter contains the results from the analysis of the data collected from the online survey. First, the data file is described and demographic data are presented. Next, the measurement model for selection and success is examined. The general exploratory hypothesis and specific hypotheses for architecture selection are tested mainly using binary and multinomial logistic regression. Similarly, the general exploratory hypothesis and specific hypotheses for architecture success are tested mainly using analysis of variance techniques.

7.2 Setting up the data file

During May and June of 2004, various professional organizations and leading authorities in the field were asked to promote the online survey to their clients/subscribers/members via electronic communication channels. At the end of June, 652 responses from organizations were gathered. Of the 652 responses, many surveys were omitted from the final sample for a variety of reasons. First, 97 were omitted from organizations that did not have a data warehouse in place at the time of the survey. Also, many surveys had data for only the first few questions and they were not used. Approximately 117 survey responses fell into this category and were discarded.

Next, the data file was examined to identify valid responses based on the data warehouse architectures identified in this study. Ten organizations were not sure of the data warehouse architecture they had implemented and were removed from the data set. The remaining organizations identified their architecture as one of the four main data warehouse architectures.

However, these responses were further assessed to omit the organizations that described their architecture as only a slight match to one of the four main architectures. Twenty-seven organizations indicated their architecture was less than 4 on a 1-7 scale that asked how closely their architecture matched one of the four main data warehouse architectures (Question 6 in Appendix B).

The resulting distribution of the four main data warehouse architectures was as follows:

1. Independent data mart architecture (IDM) = 44
2. Data mart bus architecture (DBA) = 104
3. Enterprise data warehouse architecture (EDW) = 235
4. Federated (FED) = 17

This distribution is consistent with other data warehouse surveys (e.g., Eckerson 2003). While a relatively small sample size for the FED architecture was expected prior to data collection, the seventeen responses collected was lower than expected. As a result, the FED sample size was insufficient to conduct statistical analysis techniques such as Multinomial logistic regression and MANOVA.[7] Consequently, FED was dropped as a category for the data analysis. The final data set contained 384 usable data points (See Table 7.1).

[7] For instance, in MANOVA, to achieve a power of 0.80 and detect at a minimum a very large effect size - for a four group, eight dependent variable design - 26 responses per group is required (Hair et al. 1995).

Table 7.1: Summary of the data file

Description of data file	Number of responses
Initial data set	652
No data warehouse in place	97
Incomplete responses	117
Not sure of the data warehouse architecture	10
Implemented architecture only slightly matched one of the four major architectures	27
FED architecture responses	17
Final data set	384

7.3 Demographics

A description of the demographic information of the final data file in terms of organizations, the respondents, and the data warehouse architectures is presented below.

7.3.1 Organizations

The organizations that participated in the study represented nations across the world with 58.1 percent of them from the United States (See Figure 7.1). Within the United States, organizations represented all the different regions of the country (See Table 7.2a). The mean gross revenues were 660 million with a minimum value of 5 million, and a maximum value of 12.5 billion. The mean number of employees was 3,750 with a minimum of 50 and a maximum of 12,500. The organizations represented a wide variety of industries (See Table 7.2b).

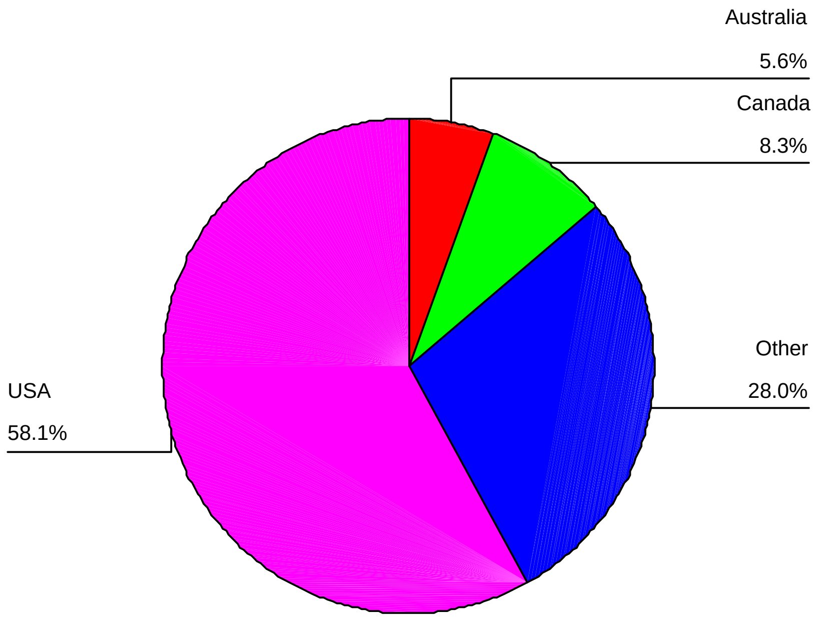


Figure 7.1: Organizations by nation

Table 7.2a: Participants by region

Region	Frequency	Percentage
Mid-Atlantic	41	18.8
Midwest	52	23.9
New England	15	6.9
South	40	18.3
Southwest	26	11.9
West	44	20.2
Total	218	100.0

Table 7.2b: Organizations by industry

Industry	Frequency	Percentage
Consulting/professional services	58	14.1
Financial services/banking	51	12.5
Insurance	36	9.6
Retail/Wholesale/ distribution	34	9.1
Manufacturing (non-computer)	33	8.8
Other	32	8.5
Telecommunications	31	8.3
Government	27	7.2
Software/Internet	18	4.8
Education/publishing	15	4
Healthcare	15	4
Transportation/logistics	14	3.7
Utilities	12	3.2
Computer manufacturing	8	2.1
Total	384	100.0

7.3.2 Respondents

The survey respondents were asked two questions about their position in the organization and if they were actively involved in the selection of the data warehouse architecture. Most of the respondents were data warehouse or IS managers (43%). Others were data warehousing staff members (21%), systems integrators/consultants (20%), and the rest were employees holding some other position in the organization (e.g., CIO). Sixty-eight percent of the respondents were actively involved in architecture selection.

7.3.3 Data warehouse architecture

Several questions asked for general information about the data warehouse, including:

- The size of raw data (in Gigabytes);
- The number of active users; and
- The time since implementation (in months).

The means, standard deviations, minimum values, and maximum values for these questions are presented in Table 7.3.

Table 7.3: General information about the data warehouses

	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
The size of raw data	371	1.00	9.00	4.56	2.13
Number of active users	371	1	30,000	691.02	2,303.96
Time since implementation	375	1	240.00	49.90	38.17

The data show that on average the data warehouse was implemented 4 years prior to the survey and contains approximately 4.5 Gigabytes of raw data. The number of active users that access the warehouse at least one a month ranges from 1 to 30,000, with most organizations having close to 700 active users on average.

7.4 The measurement model

Several analysis techniques were used to assess the reliability and validity of the measurement models. The variables in the study were first divided into two groups based on the research question they pertained to: (1) the architecture selection group and (2) the architecture success group. Next, as response variables were expected to co-vary with their predictors, variables in the success group were further divided into two groups. As a result, variables fell into three different groups for measurement model assessment. The architecture selection group consisted of all variables intended to measure the organizational selection factors that affect data warehouse architecture.[8] The architecture quality group consisted of variables intended to measure information quality and system quality.[9] Finally, the architecture impacts group consisted of variables intended to measure the individual and organizational impacts of data warehouse architecture.

7.4.1 Exploratory factor analysis

Using the SPSS 11.5 factor analysis procedure, exploratory factor analysis (EFA) was conducted to assess the unidimensionality of items and constructs. The final data set sample size of 384 met the minimal sample size requirements for EFA of seven to 10 responses per item (Nunnally 1978). Principal components analysis with a promax rotation was employed to conduct EFA in two steps. In the first step, the EFA included all the items with the number of factors specified to be the number of latent constructs. Items were expected to load on their proposed construct with a loading greater than 0.60 and to load on other factors by less than 0.30 (Comrey et al. 1992). In the second step, an EFA was conducted for the items of each latent factor separately using the surviving items from the first step. In this step, the number of factors was unspecified.

7.4.1.1 Architecture selection group

The items on architecture selection were first examined to check if items were appropriate for factor analysis. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy of 0.75 indicated that items met the requirements for factor analysis. In step one of EFA, the initial factor matrix indicated that all items loaded above 0.60 on the intended factor except for three items. Two of the

---

[8] The items on existing system compatibility, source of sponsorship, and expert influence were not included in the assessment of the measurement model as existing system compatibility and source of sponsorship was a single item factor and expert influence was a causal factor with three items.
[9] The items on time and cost were not included in the assessment of the measurement model as time was a single item factor and cost was a causal factor with three items: initial cost, annual cost and overall cost.

items addressed vertical information flow while the other addressed the view of the data warehouse. Each variable in the analyses and its corresponding coded name for measurement model analysis purposes is presented in Table 7.4.

A closer review of the VIF questions revealed that the first two addressed upper management's need to drill down to lower level data, the third referred to the need to aggregate lower level data, and finally the fourth described upper management's need to comply with regulations. As the fourth VIF question specifically refers to information needs resulting from new regulations established in the last couple of years, and cross loaded on other factors, this item was removed from the analysis.

Kimball (1996) describes the ability to drill down and roll up (i.e., aggregate) as fundamental data access needs of upper management. Although the data warehousing literature refers to drill down and roll up abilities together as a single functionality of a data warehouse, it is possible that upper management perceives the ability to drill down and roll up as separate. The third VIF item loaded only 0.501 on the underlying VIF factor and cross loaded on the TSK factor. Consequently, it was omitted from further analysis.

Next, the first question on VIEW was reverse coded to reflect a similar scale as the other two items. This item loaded only 0.530 on the underlying VIEW. It is possible that the reverse coded question was not clear to the respondents. As such, the item was removed from further analysis. The final factor matrix for step one of the EFA for the selection group is presented in Table 7.5. Step two of the EFA for architecture selection indicated that items from step 1 for each construct loaded on its intended latent factor above 0.60.

Table 7.4 Coded names for variables used in data analyses

Variable name	**Code name**
Selection factors	
Horizontal information interdependence (Interdependence)	INT
Vertical information flow (vertical info)	VIF
View of the warehouse (strategic view)	VIEW
Urgency	URG
Resource availability (resource constraints)	RES
Task routineness (task)	TSK
Perceived ability of the IT staff (perceived ability)	SKL
Source of sponsorship (sponsorship level)	SPN
Success factors	
Flexibility	FLX
Scalability	SCAL
Integration	INTG
Accuracy	ACUR
Completeness	COMP
Consistency	CONS
Individual impacts	INDV
Organizational impacts	ORG

Table 7.5: The results of EFA for architecture selection – step1

Item	Factors						
	1	2	3	4	5	6	7
INT1	-.007	**.858**	.027	.053	-.014	.005	-.058
INT2	-.016	**.903**	-.066	-.006	.054	-.046	.046
INT3	.025	**.880**	.034	-.024	-.025	.019	.040
VIF1	.012	-.032	-.006	.046	.056	**.879**	-.005
VIF2	-.018	.007	-.029	-.036	-.027	**.949**	.028
VIEW2	-.088	.021	-.043	-.100	.044	-.016	**.970**
VIEW3	.103	.008	.120	.113	-.098	.078	**.644**
URG1	**.933**	-.044	-.022	-.014	-.014	-.025	-.018
URG2	**.927**	.041	.014	-.001	-.009	.028	-.051
URG3	**.905**	.008	.000	-.023	.031	-.010	.023
RES1	.011	.053	.091	-.047	**.776**	.122	-.073
RES2	-.084	.024	.010	-.048	**.899**	-.009	-.072
RES3	.116	-.073	-.076	.136	**.736**	-.096	.185
TSK1	-.006	-.021	-.020	**.803**	.088	-.012	-.057
TSK2	-.024	-.007	.086	**.910**	-.052	-.005	-.056
TSK3	-.012	.054	-.083	**.812**	-.027	.023	.054
SKL1	-.023	-.013	**.909**	.047	.003	.014	-.001
SKL2	.065	.066	**.873**	-.047	-.018	.011	-.059
SKL3	-.053	-.062	**.862**	-.008	.048	-.064	.078

7.4.1.2 Architecture quality group

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy for items in the architecture quality group was 0.91, indicating that the items were appropriate for factor analysis. In step one of EFA, all items loaded above the suggested guidelines on their intended factor. Step two of EFA further confirmed that each set of items load well on its intended latent factor. The final factor matrix for step one of the EFA for the architecture quality group is presented in Table 7.6.

Table 7.6: The results of EFA for architecture quality group – step1

Item	Factors 1	2	3	4	6	7
FLX1	.026	-.010	-.091	**.977**	.070	-.037
FLX2	.038	.030	.056	**.898**	-.079	-.034
FLX3	-.068	-.044	.037	**.857**	.081	.086
SCAL1	-.046	.015	**.955**	.037	-.053	.013
SCAL2	-.002	-.036	**.926**	.032	-.013	-.027
SCAL3	.036	-.004	**.839**	-.123	.132	.044
INTG1	-.095	-.095	.073	.101	**.714**	.146
INTG2	.052	-.031	-.026	-.028	**.886**	.020
INTG3	.075	.167	.013	.040	**.786**	-.175
ACUR1	.007	**.915**	-.015	-.052	.041	.014
ACUR2	-.006	**.802**	-.008	-.037	-.002	.070
ACUR3	-.053	**.750**	.001	.088	.010	.101
COMP1	**.932**	-.042	-.014	-.019	.018	-.006
COMP2	**.920**	.043	.007	.026	-.039	.041
COMP3	**.913**	-.037	-.018	-.007	.085	.074
CONS1	.072	.143	.038	.033	-.052	**.863**
CONS2	-.001	.103	-.043	-.022	.050	**.837**
CONS3	.049	-.016	.030	-.008	-.039	**.901**

7.4.1.3 Architecture impacts group

The items in the architecture impacts group displayed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy over 0.90, thus establishing their suitability for factor analysis. All except for two items on organizational impacts (ORG) loaded above the suggested guidelines in step one of the EFA.

The first question on ORG evenly loaded on individual impact as well. The question specifically measured the extent to which the data warehouse met the business requirements for which it was implemented. Business requirements can include both individual impacts such as individual decision-making effectiveness, as well as organizational impacts, such as improved communication and coordination across the organization. Consequently, this item was removed.

The second item, addressed "the extent to which the data warehouse facilitated business intelligence (BI) initiatives." This item also cross-loaded on individual impacts and was omitted from the final analysis.

The final factor matrix for step one of the EFA for the architecture impacts group is presented in Table 7.7. Step two of the EFA indicated that items from step 1 for each construct loaded on its intended latent factor above 0.63.

Table 7.7: The results of EFA for architecture impacts group – step1

Item	Factor	
	1	2
INDV1	-.012	**.884**
INDV2	-.108	**.744**
INDV3	.088	**.793**
INDV4	.211	**.789**
ORG3	**.740**	.231
ORG4	**.690**	.177
ORG5	**.929**	-.175
ORG6	**.730**	.067

7.4.2 Reliability

Cronbach's alphas were assessed for items associated with each latent construct. According to Gefen et al. (2000), Cronbach's alpha for constructs in exploratory studies should exceed 0.60, and in confirmatory studies they should exceed 0.70. Furthermore, Nunnally (1978) suggests that a reliability of at least 0.6 suffices for early stages of basic research. The internal consistency of constructs in this study surpassed the suggested guidelines for exploratory studies. Table 7.8 presents the Cronbach's alphas for the architecture selection factors. Table 7.9 shows the Cronbach's alphas for both the quality and impact factors.

Table 7.8: Reliabilities of architecture selection factors

Construct	**Alpha**
INT	0.8625
VIF	0.8094
VIEW	0.6326
URG	0.9049
RES	0.7536
TSK	0.7781
SKL	0.8492

Table 7.9: Reliabilities of the architecture quality success factors and impacts of architecture factors

Construct	**Alpha**
FLX	0.9345
SCAL	0.9073
INTG	0.7649
ACUR	0.8729
COMP	0.9401
CONS	0.9267
INDV	0.8876
ORG	0.8618

7.4.3 Confirmatory factor analysis

The final assessment of the measurement model for each group was performed using confirmatory factor analysis (CFA) in structural equation modeling (SEM). SEM was conducted using LISREL 8.54 and it provides several criteria to examine the quality of a measurement model.

For instance, indicator loadings are estimates of the strength of the relationship between a questionnaire item and a latent construct. The statistical significance of indicator loadings is one estimate of the validity of the model (Shumacher et al. 1996). In addition, LISREL also provides numerous goodness-of-fit indices to judge the quality of a measurement model. The absolute fit measures the degree to which the covariance matrix produced by the data is similar to the specified model (Hair et al. 1995). The likelihood ratio chi-square statistic is the most common of these fit measures. However, it is sensitive to sample size. As a result, the ratio of chi-square to degrees of freedom is often examined instead. A ratio less than 3:1 is considered acceptable (Gefen et al. 2000).

Additionally, authorities encourage the use of other fit indices to assess the measurement model (Hair et al. 1995; Vandenberg et al. 2000). Vandenberg et al. (2000) recommend the following two fit indices: Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). They also recommend the Tucker Lewis Index (TLI or NNFI), an incremental fit index, to assess the specified model. In reviewing SEM practices in the IS literature, Gefen et al. (2000) recommends and recognizes the Adjusted Goodness of Fit Index (AGFI) as the most commonly used measure of model fit in the IS literature. The recommended target values for each of these fit statistics by each of these authorities are presented in Table 7.10.

According to Straub et al. (2004), SEM facilitates the assessment of convergent and discriminant validity through CFA by examining the “correctness of the measurement model” using fit indices. If the fit statistics are above the accepted thresholds, they indicate that the specified measurement model is well supported by the data. Furthermore, assessing the factor loadings to ensure that items load cleanly on constructs that they are presupposed to load on provides further evidence of the validity of the model.

In this dissertation, the convergent and discriminant validity of the models was established by examining the overall fit statistics and factor loadings for each model. The measures of fit mentioned above were used to evaluate the measurement models specified in this study.

7.4.3.1 Architecture selection group

The measurement model for the architecture selection group was tested using LISREL and is presented in Figure 7.2. As shown in Table 7.10, all five fit statistics for the model were above the recommended thresholds. Furthermore, as presented in Table 7.11, all indicator loadings were significant at $p < 0.05$ ($t = 1.96$). The total sample size used for the analysis was 384. Overall, the CFA revealed that the model was a good fit for the data.

Table 7.10: The goodness of fit statistics for architecture selection measurement model

Measure	Target value	Architecture selection
Chi-square	< 3 times degree of freedom	235.99*
RMSEA	<0.08	0.043
SRMR	<0.10	0.042
NNFI	>0.90	0.96
AGFI	>0.80	0.92

*Degree of freedom 131

Table 7.11: Completely standardized factor loadings for architecture selection measurement model

Observed Variables – Selection		
ITEM	**Factor loading**	**t- value**
INT1	0.81*	
INT2	0.85	16.95
INT3	0.81	16.47
VIF1	0.93*	
VIF2	0.73	7.94
VIEW2	0.54*	
VIEW3	0.86	7.00
URG1	0.85*	
URG2	0.92	21.88
URG3	0.86	20.56
RES1	0.73*	
RES2	0.81	10.61
RES3	0.61	10.01
TSK1	0.60*	
TSK2	0.84	10.81
TSK3	0.78	10.85
SKL1	0.88*	
SKL2	0.76	15.53
SKL3	0.78	15.80

* Indicates a parameter fixed at 1.0 in the original solution.

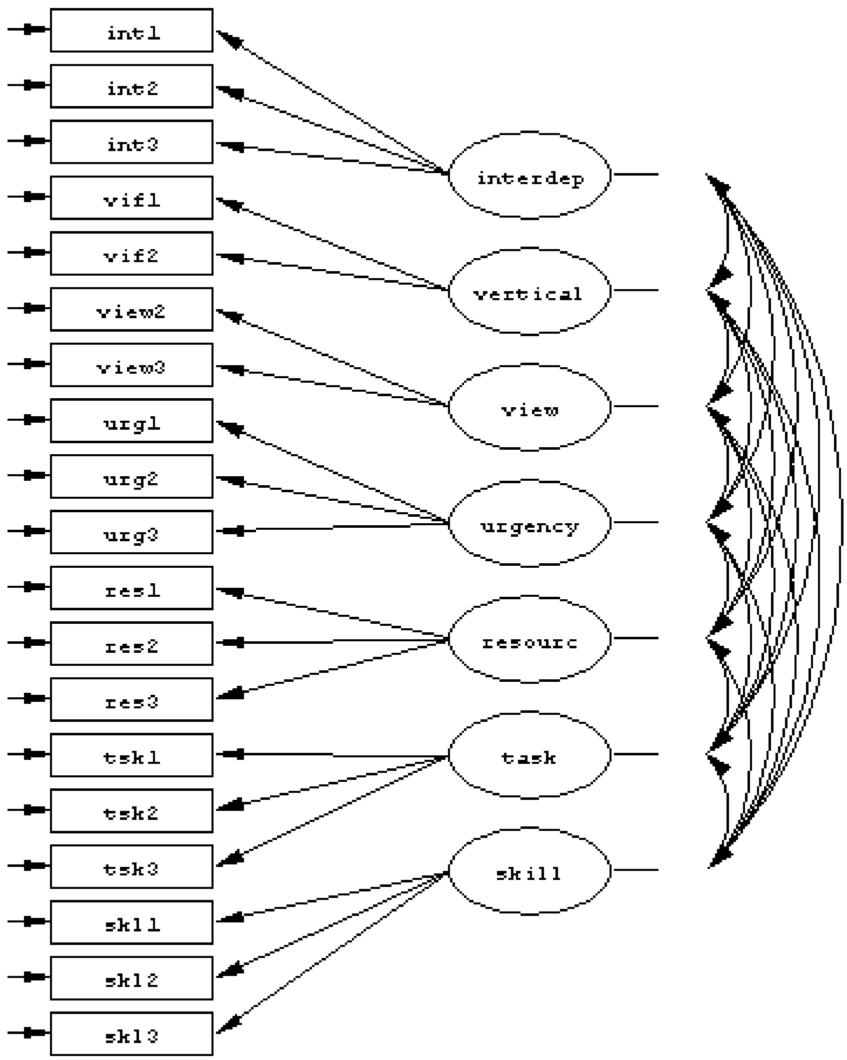


Figure 7.2: Measurement model for architecture selection group

7.4.3.2 Architecture quality group

The measurement model for the architecture quality group is presented in Figure 7.3. As shown in Table 7.12, all five fit statistics for the model were above the recommended guidelines and all indicator loadings were significant at $p < 0.05$ (See Table 7.13). The total sample size used for this analysis was less than the data set used for the selection group because 24 responses had incomplete answers for the questions on success. The revised sample size used for the analysis of the architecture quality group model was 360. The results from the CFA indicated that the model was a good fit for the data.

Table 7.12: The goodness of fit statistics for architecture quality measurement model

Measure	Target value	Architecture quality
Chi-square	< 3 times degree of freedom	247.89*
RMSEA	<0.08	0.049
SRMR	<0.10	0.041
NNFI	>0.90	0.99
AGFI	>0.80	0.91

*Degree of freedom 128

Table 7.13: Completely standardized factor loadings for architecture quality measurement model

Observed variables			Latent variables		
ITEM	**Factor loading**	**t- value**	**Factor**	**Std. structural coefficient**	**t-value**
FLX1	0.90*		Flexibility	0.83	15.71
FLX2	0.92	27.07			
FLX3	0.91	26.31			
SCAL1	0.87*		Scalability	0.72	12.78
SCAL2	0.92	23.37			
SCAL3	0.94	20.42			
INTG1	0.81*		Integration	0.80	12.92
INTG2	0.75	13.12			
INTG3	0.62	10.99			
ACUR1	0.80*		Accuracy	0.80	13.49
ACUR2	0.88	18.38			
ACUR3	0.83	17.32			
COMP1	0.88*		Completeness	0.58	10.59
COMP2	0.96	28.29			
COMP3	0.91	26.03			
CONS1	0.91*		Consistency	0.95	18.79
CONS2	0.89	25.80			
CONS3	0.90	26.19			

* Indicates a parameter fixed at 1.0 in the original solution

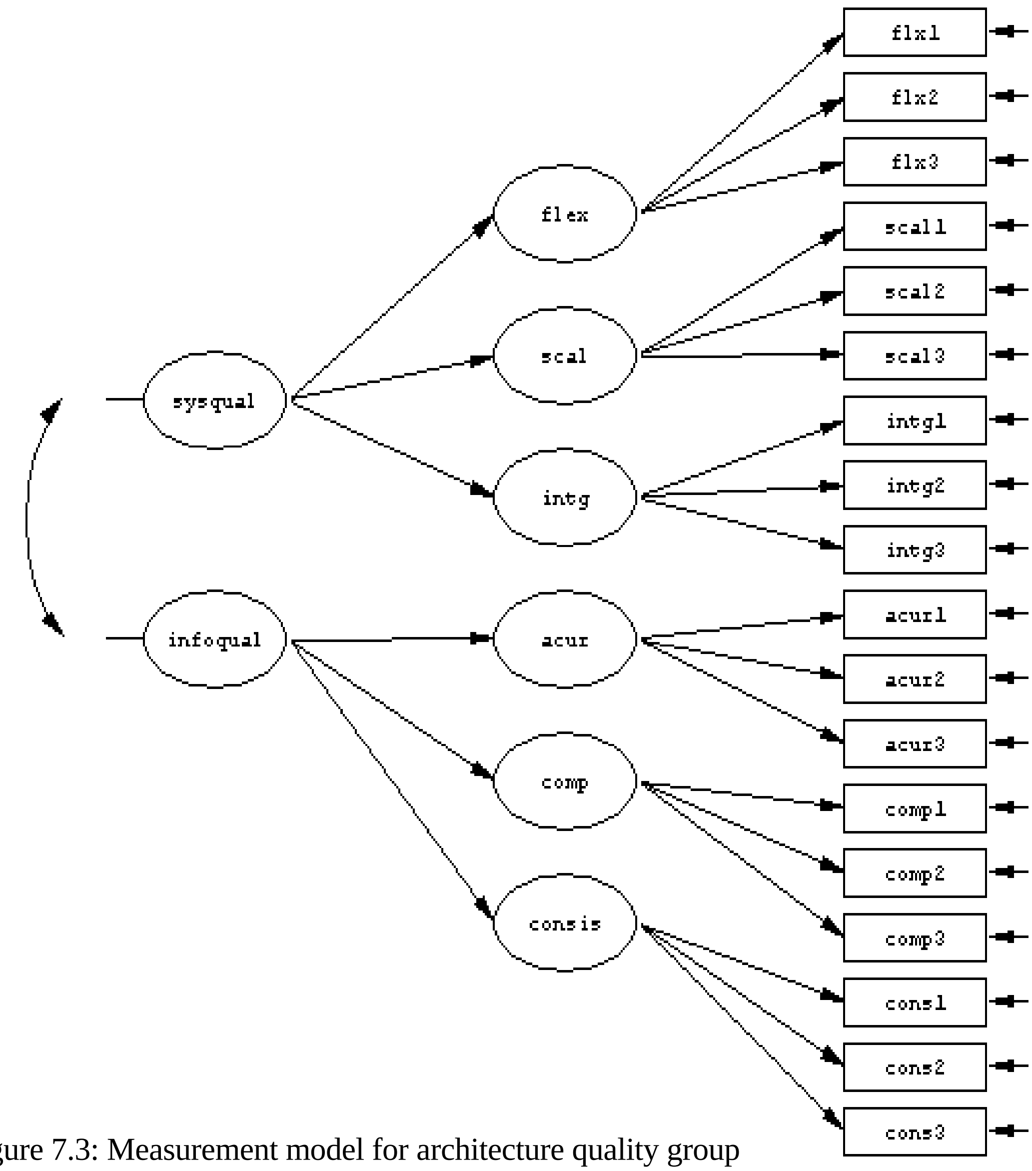


Figure 7.3: Measurement model for architecture quality group

7.4.3.3 Architecture impact group

The measurement model for the architecture impact group is presented in Figure 7.4. The results from the CFA using LISREL indicated that the model was a good fit for the data. Each of the five fit statistics for the model, displayed in Table 7.14, was above the recommended values. In addition, the indicator loadings, presented in Table 7.15, were above the 0.05 significance level (t = 1.96). The total sample size used for this analysis was 360.

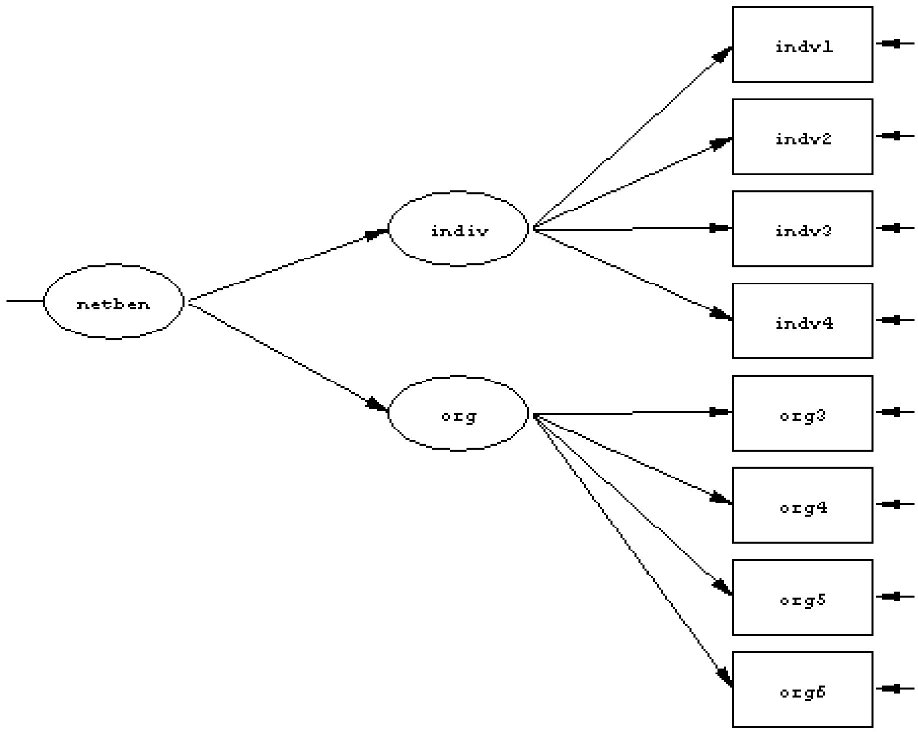


Figure 7.4: Measurement model for architecture impacts group

Table 7.15: Completely standardized factor loadings for architecture impact measurement model

Observed variables			**Latent variables**		
ITEM	**Factor loading**	**t-value**	**Factor**	**Std. structural coefficient**	**t-value**
INDV1	0.81*		Individual Impacts	0.90	15.51
INDV2	0.69	14.01			
INDV3	0.85	18.09			
INDV4	0.92	20.94			
ORG3	0.85*		Organizational Impacts	0.92	16.70
ORG4	0.86	19.36			
ORG5	0.74	15.81			
ORG6	0.70	14.53			

* Indicates a parameter fixed at 1.0 in the original solution

Table 7.15: The goodness of fit statistics for architecture impact measurement model

Measure	**Target value**	**Architecture impact**
Chi-square	< 3 times degree of freedom	32.28
RMSEA	<0.08	0.044
SRMR	<0.10	0.023
NNFI	>0.90	0.99
AGFI	>0.80	0.96

*19 Degrees of freedom

7.4.4 Summary of the measurement model assessment

The measurement model was examined for three groups (i.e., architecture selection, architecture quality, and architecture impacts) through the use of exploratory factor analysis, reliability estimation, and confirmatory factor analysis. After eliminating only four items from all three groups through the EFA, the CFA for each group revealed good model fit. Next, the results of testing hypotheses on architecture selection are presented, followed by a description of the testing of the hypotheses related to architecture success.

7.5 Architecture selection

The data analyses on architecture selection were divided into two parts. First, the exploratory hypothesis on architecture selection (i.e., Hypothesis E1: The selection factors influence the selection of a particular data warehouse architecture) was tested through several sets of data analyses. The second part of the data analyses involved testing specific hypotheses.

7.5.1 Architecture selection: exploratory hypothesis

The exploratory hypothesis on architecture selection involved testing the full research model for architecture selection (See Figure 4.1), which describes the influence of all selection factors on the selection of data warehouse architectures.

In order to understand the influence of selection factors on architecture in less complex circumstances and in preparation for testing the full research model, a set of analyses were undertaken to identify which factors influence the selection between two architectures choices. The three specific steps of the data analysis used to assess the exploratory hypothesis on architecture selection were:

Step one: Analyses to identify which selection factors influence the selection of one architecture when compared to the rest of the architectures (e.g., EDW versus the rest).

Step two: Analyses to identify which selection factors influence the selection of one architecture when compared to another (e.g., EDW versus IDM).

Step three: Analyses to identify which selection factors influence the selection of all three data warehouse architectures.

Also, an additional new variable (e.g., domain) was examined to investigate if it influenced architecture selection.

As the dependent variable was dichotomous in nature, binary logistic regression (BLR) was used to perform steps one and two. Multinomial logistic regression (MLR) was applied to estimate the factors that influence the selection of one of the three data warehouse architectures: IDM, DBA, and EDW in step three. Data warehouse architecture was entered as the dependent variable. Eight of the ten original selection factors (See Figure 4.1) described in hypothesis E1 were entered as the independent variables[10]. Existing system compatibility and expert influence were excluded from the MLR analysis, as these factors do not point towards the selection of a particular architecture. A statistical analysis of these two variables is described in a separate section later in the chapter. The descriptive statistics for the eight selection factors for each architecture category entered in the MLR are presented in Table 7.16a, b, and c.

In presenting guidelines for the application of logistic regression, Hosmer et al. (2000), suggest the use of the stepwise procedure when the outcome being studied is relatively new, the important covariates may not be known, and associations with the outcome are not well understood. Given the exploratory nature of this study, stepwise multinomial logistic regression was used to test the hypothesis.

A crucial aspect of using stepwise logistic regression is the choice of the alpha level to judge the importance of a variable to include in the final model. An extensive examination of the issue of significance level in logistic regression by Lee et al. (1997) indicates that the choice of p = 0.05 is too stringent in stepwise logistic regression, resulting in exclusion of key variables. Instead, they recommend a p-value ranging from 0.15 to 0.20 for inclusion of a variable in the final model. As a result, the forward Wald method using an alpha level of 0.20 for entry was applied to conduct the stepwise analyses. The data analyses in step one are described next.

[10] The multiple items that survived the measurement model assessment were averaged to create each selection variable.

Table 7.16a: Descriptive statistics for selection variables in IDM

	N	Minimum	Maximum	Mean	Std. Deviation
Interdependence	44.00	1.00	7.00	4.30	1.51
Vertical info	44.00	1.00	7.00	4.13	1.83
Strategic view	44.00	1.00	7.00	4.19	1.72
Urgency	44.00	1.00	7.00	4.73	1.67
Resource constraints	44.00	1.00	7.00	4.67	1.41
Task	44.00	1.00	7.00	4.71	1.41
Perceived ability	44.00	1.00	7.00	3.36	1.55
Sponsorship level	44.00	1.00	6.00	2.43	1.23
Total sample size	44.00				

Table 7.16b: Descriptive statistics for selection variables in DBA

	N	Minimum	Maximum	Mean	Std. Deviation
Interdependence	105.00	1.00	7.00	4.97	1.44
Vertical info	105.00	1.00	7.00	4.71	1.70
Strategic view	105.00	1.00	7.00	5.04	1.43
Urgency	105.00	1.00	7.00	5.14	1.43
Resource constraints	105.00	1.00	7.00	4.16	1.57
Task	105.00	1.00	7.00	5.16	1.28
Perceived ability	105.00	1.00	7.00	3.99	1.75
Sponsorship level	105.00	1.00	6.00	2.59	1.36
Total sample size	105.00				

Table 7.16c: Descriptive statistics for selection variables in EDW

	N	Minimum	Maximum	Mean	Std. Deviation
Interdependence	235.00	1.00	7.00	4.87	1.36
Vertical info	235.00	1.00	7.00	4.47	1.57
Strategic view	235.00	1.00	7.00	5.35	1.28
Urgency	235.00	1.00	7.00	4.91	1.41
Resource constraints	235.00	1.00	7.00	4.27	1.46
Task	235.00	1.00	7.00	5.20	1.25
Perceived ability	235.00	1.00	7.00	4.23	1.57
Sponsorship level	235.00	1.00	6.00	2.85	1.30
Total sample size	235.00				

7.5.1.1 Architecture selection: step one

To understand which selection variables influence the selection of a particular architecture when compared to the rest of the architectures, the following stepwise binary logistic regression models were examined: EDW compared to the rest, DBA compared to the rest, and IDM compared to the rest. The outcome of each of these analyses is presented next. First, the stepwise BLR model for EDW compared to the rest.

The likelihood ratio test for the logistic regression model (chi-square =22.33 [4df]) was statistically significant at $p < 0.0001$. The Nagelkerke pseudo R-squared for the model was 0.160. The Hosmer and Lemeshow goodness-of-fit test was non-significant at $p<0.05$, implying that the model's estimates fit the data at an acceptable level. The stepwise procedure produced vertical info, strategic view, perceived ability, and sponsorship level as the factors that influence the selection of an EDW architecture when compared to the rest (See Table 7.17). Parameter estimates and odds ratios for each independent variable are shown in Table 7.17. The model results show the selection variable parameter estimates and odds ratios for the EDW architecture category with the rest of the architectures (i.e., IDM and DBA) as the reference category.

Table 7.17: BLR output – step one: parameter estimates for EDW vs. rest

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
EDW	Vertical info	-.123*	.885	3.042	1	.081
	Strategic view	.265**	1.303	10.456	1	.001
	Perceived ability	.131*	1.140	3.824	1	.051
	Sponsorship level	.126	1.134	2.150	1	.143
	Constant	-1.207	.299	5.793	1	.016

a The reference architecture category is: The rest of the architectures
** = $p < 0.05$; * = $p < 0.1$[11]

[11] Parameter estimates were assessed at $p < 0.10$ level of significance in steps one and two of architecture selection in order to explore the influence of selection factors on architecture choice. In step three, only $p < 0.05$ was used to identify influential selection factors.

The parameter estimates from the BLR model for EDW compared to the rest show that at $p < 0.1$, only vertical info, strategic view, and perceived ability influence architecture choice. More specifically, the odds of selecting EDW as a data warehouse architecture when compared to selecting either IDM or DBA architecture increases by 87.7 percent[12] for a one unit increase in strategic view while the other selection variables remain constant. Similarly, as the perceived ability of the IT staff increases, the odds of an organization selecting an EDW architecture increases when compared to the odds of selecting one of the other two architectures. Furthermore, when the upper management's needs for information (i.e., vertical info) decrease, a company is more likely to select an EDW architecture. Next, the logistic model for DBA versus the rest is described.

The stepwise BLR model for DBA compared to the rest was significant at a 0.05 level of significance (chi-square = 10.74 [4df]). The Nagelkerke R-squared was 0.04. In addition, the non-significance of the Hosmer-Lemeshow goodness-of-fit test indicated adequate model fit. The stepwise procedure identified four variables as influencing DBA architecture selection. They were vertical info, urgency, resource constraints, and sponsorship level.

Table 7.18: BLR output – step one: parameter estimates for DBA vs. rest

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
DBA	Vertical info	.148**	1.160	3.882	1	.049
	Urgency	.154*	1.167	3.248	1	.072
	Resource constraints	-.146*	.864	3.266	1	.071
	Sponsorship level	-.170*	.844	3.541	1	.060
	Constant	-1.353	.259	5.367	1	.021

a The reference architecture category is: The rest of the architectures
** = $p < 0.05$; * = $p < 0.1$

[12] The odds ratio for view when comparing EDW and other architectures is 0.123. One minus 0.123 is 0.877

The parameter estimates from the BLR model for DBA compared to the rest show that at $p < 0.1$, all variables selected by the stepwise process were significant (See Table 7.18). Specifically, as upper management's needs for information (i.e., vertical info) and urgency increases, organizations are more likely to implement a DBA architecture. In addition, when the constraints on resource availability are high and the source of sponsorship elevates to higher levels within the organization, organizations are less likely to select a DBA architecture when compared to the rest. Finally, the binary logistic analysis for IDM versus the rest is reported.

The stepwise BLR for IDM compared to the rest provided a significant model (chi-square=32.62[3]) with 0.160 Nagelkerke R-squared. The stepwise procedure indicated three variables influenced the selection of IDM architecture when compared to the rest of the architectures. They were strategic view, perceived ability, and resource constraints. The resulting parameter estimates were all significant at $p < 0.05$ (See Table 7.19).

Table 7.19 BLR output- step one: parameter estimates for IDM vs. rest

Architecture category	**Variables**	**Beta**	**Odds Ratio [Exp(Beta)]**	**Wald**	**Degrees of Freedom**	**Significance**
IDM	Strategic view	-.452**	.636	17.450	1	.000
	Resource constraints	.355**	1.426	6.961	1	.008
	Perceived ability	-.304**	.738	7.523	1	.006
	Constant	-.339**	.713	.207	1	.649

a The reference architecture category is: The rest of the architectures
** = p < 0.05; * = p < 0.1

According to the parameter estimates, as the view of the data warehouse as a strategic initiative and the perceived ability of the IT staff decrease, organizations are more likely to implement an IDM architecture rather than the rest of the architectures. Finally, as resource constraints increase, firms are more likely to implement an IDM. The data analyses in step two are described next.

7.5.1.1 Architecture selection: step two

According to Hosmer et al. (2000), researchers often conduct binary logistic regression analysis between different combinations of two categories of the dependent variables as a first step towards identifying significant relationships between the predictor variables and the multiple category dependent variable. Begg et al. (1984) state that conducting BLR's to ascertain significant independent variables approximates multinomial logistic regression without losing much efficiency. In this study, three stepwise binary logistic regression models were tested to distinguish the selection variables that influence the different combinations of architecture categories. They were IDM versus EDW, DBA versus EDW, and IDM versus DBA.

The stepwise BLR analysis for IDM compared to EDW and produced an adequate model. The likelihood ratio test for the BLR model (chi-square =34.86 [3df]) was statistically significant at $p < 0.0001$. The Nagelkerke pseudo R-squared for the model was 0.202. The Hosmer and Lemeshow goodness-of-fit test was non-significant at $p<0.05$, suggesting acceptable fit. The stepwise procedure produced strategic view, resource constraints, and perceived ability as the factors that influence the selection of an IDM when compared to an EDW architecture (See Table 7.20).

Table 7.20 BLR output- step two: parameter estimates for IDM vs. EDW

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
IDM	Strategic view	-.504**	0.346	18.174	1	.000
	Resource constraints	.349**	1.294	5.842	1	.016
	Perceived ability	-.317**	.627	7.180	1	.007
	Constant	-.392	.324	.234	1	.629

a The reference architecture category is: EDW architecture
** = p < 0.05; * = p < 0.1

The results from IDM versus EDW were similar to the parameter estimates obtained in step one for IDM versus the rest of the architectures. The output suggested that as the strategic view and perceived ability decrease and resource constraints increase, organizations are more likely to implement an IDM than an EDW. The fit statistics of the BLR for DBA versus EDW indicated an adequate model fit. The overall model was significant at $p < 0.01$ (chi-square =11.92 [4df]). The pseudo R-squared was only 0.05 but the Hosmer and The Lemeshow goodness-of-fit test was not significant, suggesting good fit. The stepwise procedure identified the following variables as affecting the selection of a DBA when compared to an EDW: vertical info, strategic view, urgency, and sponsorship level (See table 7.21)

Table 7.21: BLR output- step two: parameter estimates for DBA vs. EDW

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
DBA	Vertical info	.146*	1.136	3.440	1	.064
	Strategic view	-.196**	.783	4.376	1	.036
	Urgency	.147*	1.137	2.779	1	.095
	Sponsorship level	-.143	.846	2.303	1	.129
	Constant	.807	.251	1.539	1	.215

a The reference architecture category is: EDW architecture
** = p < 0.05; * = p < 0.1

The parameter estimates from the BLR model for DBA versus EDW show that at $p < 0.1$, only vertical info, strategic view, and urgency influence the selection of DBA. As vertical info and urgency increases, firms are more likely to select a DBA. Moreover, when strategic view is low, the odds of selecting a DBA when compared to the odds of selecting an EDW increase. The final BLR of step two, DBA versus IDM is presented next.

The log likelihood ratio and the Hosmer and Lemeshow test indicated adequate fit of the stepwise BLR for DBA compared to IDM. The pseudo R squared was 0.194. The stepwise procedure identified strategic view, urgency, resource constraints, and perceived ability, as the factors that significantly influence the selection of a DBA versus an IDM (See table 7.22).

Table 7.22: BLR output - step two: parameter estimates for DBA vs. IDM

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
DBA	Strategic view	.328**	1.388	6.425	1	.011
	Urgency	.215	1.239	2.607	1	.106
	Resource constraints	-.445**	.641	8.378	1	.004
	Perceived ability	.290**	1.337	5.149	1	.023
	Constant	-.808	.446	.759	1	.384

a The reference architecture category is: IDM architecture
** = $p < 0.05$; * = $p < 0.1$

According to the parameter estimates, as strategic view, perceived ability, and resource availability increases, organizations are more likely to select a DBA than an IDM architecture. The third and final step of the exploratory analysis on architecture selection is described next.

7.5.1.1 Architecture selection: step three

In step three, the stepwise MLR was conducted using the eight selection factors and three categories of data warehouse architecture. The EDW architecture was treated as the reference category.[13] The main model was analyzed using Likelihood ratio test and model fitting information including chi-square and R-squared.

The goodness-of-fit statistic (chi-square = 757.80; significance = 0.609) indicated that the multinomial logistic regression model was not significantly different from a perfect model, which correctly classifies all responses into one of the three data warehouse architectures. Furthermore, the likelihood ratio test for the overall model was significant at the 0.05 level, further indicating good model fit. For the pseudo R-squared statistics, the Cox and Snell and Nagelkerke were rather low at 0.111 and 0.133, respectively. According to Hosmer et al. (2000) unlike linear regression, low pseudo R-squared statistics are the norm in this analysis technique.

The likelihood ratio and the model chi-square assess the overall logistic model but do not indicate which independent variables are important to the final model. The stepwise MLR process revealed five independent variables as significant in affecting architecture selection. They were strategic view, resource constraints, perceived ability, vertical info, and urgency. Parameter estimates and odds ratios for each independent variable are shown in Table 7.23. The model results show the selection variable parameter estimates and odds ratios for IDM and DBA architecture categories with EDW as the reference architecture.

[13] It is the category against which the other categories are compared.

Table 7.23: Initial MLR output – step 3: parameter estimates

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
IDM	Vertical info	-.003	.997	.001	1	.979
	Strategic view	-.505**	.604	17.380	1	.000
	Urgency	-.032	.968	.072	1	.788
	Resource constraints	.352**	1.421	6.183	1	.013
	Perceived ability	-.330**	.719	8.310	1	.004
	Intercept	.594		.446	1	.504
DBA	Vertical info	.145	1.157	3.425	1	.064
	Strategic view	-.217**	.805	5.560	1	.018
	Urgency	.150	1.162	2.909	1	.088
	Resource constraints	-.084	.919	1.032	1	.310
	Perceived ability	-.071	.932	.897	1	.344
	Intercept	-.451		.425	1	.515

a The reference architecture category is: EDW
** = $p < 0.05$

The table shows that strategic view, resource constraints, and perceived ability influence the selection of a data warehouse architecture when comparing the IDM and EDW architectures. More specifically, the odds of selecting IDM as a data warehouse architecture when compared to selecting EDW, decreases by 39.6 percent[14] for a one unit increase in strategic view while the other selection variables remain constant. In a similar manner, the odds ratio for perceived ability when comparing IDM and EDW suggest that as perceived ability increases by one unit[15] , the odds of selecting an IDM as compared to EDW decrease by 28.1 percent. In addition, with a higher constraint on resource availability (i.e., a one unit increase in resource), there is a 42.1 percent increase in the odds of selecting an IDM compared to an EDW architecture.

On the other hand, strategic view was significant at $p < 0.05$ in influencing the odds of selecting the DBA architecture when compared to EDW. Specifically, the odds of selecting a DBA

[14] The odds ratio for view when comparing IDM and EDW is 0.604. One minus 0.604 is 0.396
[15] A 'one unit' increase refers to the one point increase in a user's perception of the view of the warehouse on a one to seven scale of strongly disagree to strongly agree.

architecture when compared to selecting an EDW architecture decreased by 19.5 percent as strategic view increased by one unit. While vertical info and urgency were significant only at 0.1 level of significance, the analysis indicated that perceived ability and resource constraints did not influence the chance of selecting a DBA when compared to the chance of selecting an EDW architecture. The results from the MLR for the full model were similar to the results obtained in step two.

The multinomial model was also assessed for its discriminating power. The classification accuracy achieved by the model was 62.8 percent. Blindly predicting the largest architecture category (i.e., EDW) for all responses would yield a classification accuracy of 61 percent (235/384). As such, the MLR model is only marginally better at predicting architecture selection than blindly predicting the category with the largest number of responses. However, according to Kleinbaum (1994) when the goal of the model is to identify and understand the impact of independent variables rather than prediction, predictive accuracy of a model is less of a concern.

7.5.1.3 Analysis of omitted selection variables

Existing system compatibility and expert influence were not included in the previous exploratory analysis because they were not measured in a manner that suggests any particular architecture. However, it is conceivable that the need for greater compatibility with existing systems or high expert influence may affect the selection of a particular architecture more than others. As such, independent sample t-tests were conducted to investigate the influence of these selection variables. Expert influence was composed of influence from consultants, the data warehouse literature, and seminar/conference attendance.

Three t-tests were conducted. The first tested if the influence of the selection factors was different for the DBA and EDW architectures. The means and standard deviations for the selection variables in each architecture category are presented in Table 7.24a.

The results showed that the influence of the data warehouse literature was significantly different between DBA and EDW at $p < 0.1$[16] (See Table 7.24b). None of the other selection variables showed significant mean differences.

Table 7.24a: t-test output: DBA and EDW – means and std. deviations of variables

Variable	Architecture	Mean	Std. Deviation
System compatibility	DBA	4.36	2.048
	EDW	4.62	1.914
Consultant influence	DBA	4.58	2.134
	EDW	4.72	2.083
Literature influence	DBA	5.18	1.555
	EDW	4.85	1.461
Seminar/conference influence	DBA	4.24	1.811
	EDW	4.12	1.632

Table 7.24b: t-test output: DBA vs. EDW – result of t-test

Variable	t value	Degrees of freedom	Significance	Mean difference (DBA-EDW)
System compatibility	-1.103	337	.271	-.25
Consultant influence	-.578	338	.563	-.14
Literature influence	1.885	338	.060*	.33
Seminar/conference influence	.578	338	.563	.11

* = $p < 0.1$

The second t-test considered the IDM and EDW architecture categories. The means and standard deviations for selection variables in each architecture category are depicted in Table

[16] Although $p < 0.1$ is not considered an acceptable level of significance in IS research, both expert interviews and practitioner literature provided strong support for the influence of data warehouse literature on EDW and DBA architecture. Consequently, the result of the t-test for the influence of data warehouse literature on EDW and DBA architecture was interpreted as an outcome of this study.

7.25a. The results of the t-test indicated that there was no significant mean difference between IDM and EDW architecture (See Table 7.25b).

Table 7.25a: t-test output: IDM vs. EDW – means and std. deviations of variables

Variable	Architecture	Mean	Std. Deviation
System compatibility	IDM	4.67	1.973
	EDW	4.62	1.914
Consultant influence	IDM	4.45	1.946
	EDW	4.72	2.083
Literature influence	IDM	4.64	1.658
	EDW	4.85	1.461
Seminar/conference influence	IDM	3.98	1.562
	EDW	4.12	1.632

Table 7.25b: t-test output: IDM vs. EDW – result of t-test

Variable	t value	Degrees of freedom	Significance	Mean difference (IDM-EDW)
System compatibility	.185	275	.853	.06
Consultant influence	-.794	277	.428	-.27
Literature influence	-.875	277	.382	-.21
Seminar/conference influence	-.549	277	.584	-.15

* = $p < 0.1$

The final t-test was for the IDM and DBA architectures. The means and standard deviations for the selection variables for the IDM and DBA architectures are presented in Table 7.26a. The outcome of the final t-test revealed that the IDM and DBA architectures differ in terms of the influence of the data warehouse literature at the 0.1 significance level (Refer Table 7.26b).

Table 7.26a: t-test output: IDM vs. DBA – means and std. deviations of variables

Variable	Architecture	Mean	Std. Deviation
System compatibility	IDM	4.67	1.973
	DBA	4.36	2.048
Consultant influence	IDM	4.45	1.946
	DBA	4.58	2.134
Literature influence	IDM	4.64	1.658
	DBA	5.18	1.555
Seminar/conference influence	IDM	3.98	1.562
	DBA	4.24	1.811

Table 7.26b: t-test output: DBA vs. EDW – result of t-test

Variable	t value	Degrees of freedom	Significance	Mean difference (IDM-DBA)
System compatibility	.852	146	.396	.31
Consultant influence	-.338	147	.736	-.13
Literature influence	-1.912	147	.058*	-.54
Seminar/conference influence	-.834	147	.406	-.26

* = $p < 0.1$

The three t-tests conducted to identify the influence of existing system compatibility and expert influence on data warehouse architecture suggested that the influence of the data warehouse literature may be higher for DBA when compared to EDW and IDM.

7.5.1.4 Additional analysis: other variables

The data analyses on architecture selection indicated that horizontal information interdependence had no influence on architecture selection. The domain of the organization, in which interdependence is measured, may influence interdependence and in turn influence architecture selection. In order to investigate if domain has more of an influence on architecture selection than interdependence, a multinomial logistic regression was performed.

All of the selection variables previous analyzed in architecture selection step one, as well as the new variable 'domain,'[17] were entered into a stepwise MLR. As before, EDW architecture was used as the reference category and parameter estimates were obtained for the remaining two architectures. The overall likelihood ratio test (chi-square = 57.88[12df]) was significant at $p < 0.0001$, indicating a good model fit.

The pseudo R-squared statistic was 0.167. In addition to the five variables that were previously identified as influencing architecture in the initial MLR (undertaken in step one), domain was also identified as a significant variable that influences architecture selection. Again, interdependence was not identified as a significant variable[18]. The parameter estimates from the MLR model is presented in Table 7.27a.

Similar to the initial MLR in step one, strategic view, resource constraints, and perceived ability were identified as significant variables (at alpha 0.10) that influence the selection of IDM architecture when compared with the EDW architecture. Furthermore, similar to the initial model, strategic view, vertical info, and urgency were also significant in influencing the likelihood of selecting a DBA architecture when compared to EDW. The new variable, domain, was significant at $p < 0.05$, indicating that a reduction in the size of the domain of the organization would increase the likelihood of an organization selection an IDM or a DBA when compared to an EDW architecture.

A summary of the results from the exploratory analysis is presented in Table 7.27b. The selection factors that were identified as significant are identified by check marks.

---

[17] As mentioned in chapter six on operationalization, domain was captured with five responses. However, they represented three specific domains. In order to enhance interpretation, the responses for domain were recoded. The response 'entire company' was coded as one, 'several business units' as two and the rest as three.

[18] In order to see if domain influences the manner in which other selection variables (specially interdependence) influence architecture selection, it was entered as the first block followed by all the other factors in a another block into a MLR analysis. This analysis with block separation, used to control domain in the rest of the analysis, did not give results that was different from the results presented in this section.

Table 7.27a: MLR output: parameter estimates for MLR with domain

Architecture category	Variables	Beta	Odds Ratio [Exp(Beta)]	Wald	Degrees of Freedom	Significance
IDM	Vertical info	.010	1.010	.007	1	.932
	Strategic view	-.455**	.635	13.534	1	.000
	Urgency	-.022	.978	.033	1	.856
	Resource constraints	.281*	1.325	3.789	1	.052
	Perceived ability	-.348**	.706	9.076	1	.003
	Domain	.572**	1.772	5.826	1	.016
	Intercept	-.486		.232	1	.630
DBA	Vertical info	.155*	1.168	3.801	1	.051
	Strategic view	-.177*	.838	3.534	1	.060
	Urgency	.159*	1.173	3.169	1	.075
	Resource constraints	-.139	.871	2.607	1	.106
	Perceived ability	-.088	.915	1.360	1	.243
	Domain	.492**	1.636	9.267	1	.002
	Intercept	-1.376		3.232	1	.072

a The reference architecture category is: EDW
** = $p < 0.05$; * = $p < 0.1$

7.5.2 Architecture selection: specific hypotheses

Each of the specific selection hypotheses described in Chapter four, except for those that referred to the FED architecture, was tested. The hypotheses that described a comparison of the FED architecture with another data warehouse architecture could not be tested because of the small number of FED in the data set. To test each specific hypothesis, separate binary logistic regression analyses were conducted. The results of testing each hypothesis along with the beta coefficient and significance of the parameter estimate of the corresponding BLR are presented in table 7.28.

Table 7.27b Architecture selection exploratory hypotheses testing results

Data analysis type	Strategic view	Resource constraints	Perceived ability	Vertical info	Urgency	Sponsorship level	Interdependence	Task	System compatibility	Expert influence - consultants	Expert influence - literature	Expert influence – seminars/conferences	Domain
Step 1													
EDW vs. rest	√**		√*	√*									
DBA vs. rest		√*		√**	√*	√*							
IDM vs. rest	√**	√**	√**										
Step 2													
IDM vs. EDW	√**	√**	√**										
DBA vs. EDW	√**			√*	√*								
DBA vs. IDM	√**	√**	√**										
Step 3													
Full selection model													
IDM vs. EDW	√**	√**	√**										
DBA vs. EDW	√**			√*	√*								
Omitted variable analysis													
DBA vs. EDW											√*		
IDM vs. EDW													
DBA vs. IDM											√*		
Additional variable analysis													
Inclusion of Domain in full model													
IDM vs. EDW	√**	√**	√**										√*
DBA vs. EDW	√**			√*	√*								√*

** = $p < 0.05$; * = $p < 0.1$

Table 7.28: Specific hypothesis testing for architecture selection

Hypothesis	Beta	Significance	Result
Hypothesis 1: Organizations with high information interdependence are more likely to select an EDW architecture than an IDM architecture.	-.280**	0.014	Supported
Hypothesis 2: Organizations with high vertical information flow are more likely to select an EDW architecture than an IDM architecture.	0.129	0.195	Not supported
Hypothesis 3a: Organizations with high urgency are more likely to select an IDM architecture than an EDW architecture.	-.027	.818	Not supported
Hypothesis 3b: Organizations with high urgency are more likely to select a DBA architecture than an EDW architecture	.152*	0.088	Supported
Hypothesis 4: Organizations with tasks that are relatively more routine are more likely to select a FED architecture than an EDW architecture.			Not tested
Hypothesis 6: Organizations that view the implementation of a data warehouse as a short-term point solution rather than a strategic infrastructure project are more likely to select an IDM architecture than an EDW architecture.	-.449**	.001	Supported
Hypothesis 7a: Organizations with low resources are more likely to select an IDM architecture than an EDW architecture.	0.370**	0.011	Supported
Hypothesis 7b: Organizations with low resources are more likely to select a DBA architecture than an EDW architecture.	-0.102	0.228	Not supported
Hypothesis 7c: Organizations with low resources are more likely to select a FED architecture than an EDW architecture.			Not tested
Hypothesis 8a: Organizations that have an in-house IT staff with low perceived ability are more likely to select an IDM architecture than an EDW architecture.	-.331**	.004	Supported
Hypothesis 8b: Organizations that have an in-house IT staff with low perceived ability are more likely to select a DBA architecture than an EDW architecture.	-0.083	0.276	Not supported
Hypothesis 9a: Organizations with upper management sponsorship is more likely to select an EDW architecture than an IDM architecture.	-0.255*	0.05	Supported
Hypothesis 9b: Organizations with upper management sponsorship is more likely to select an EDW architecture than a DBA architecture.	-0.184*	0.059	Supported

** = $p < 0.05$; * = $p < 0.1$

7.6 Architecture success

The data analyses on architecture success were divided into two parts. In the first part, the exploratory hypothesis on success was tested by examining the impact of data warehouse architectures on the success variables[19] in the following steps:

Step one: Analyses to identify which success variables are affected by the three data warehouse architectures.

Step two: Analyses to identify which success variables are affected by one architecture when compared to the rest (e.g., EDW versus the rest).

Step three: Analyses to identify which success variables are affected by a combination of two data warehouse architectures (e.g., EDW and DBA; EDW and IDM).

Second, specific hypotheses on architecture success were tested. The statistical analyses relevant to the exploratory hypothesis on success are presented next.

7.6.1 Architecture success: exploratory hypothesis

Multivariate analysis of variance was conducted to test the exploratory hypothesis on architecture success (i.e., Hypothesis E2: The data warehouse architecture influences the success of the data warehouse development process and resulting data warehouse solution). As there were a rather large number of dependent success variables, in order reduce the complexity of the analysis, the success variables were divided into two groups: (1) the variables on system product success[20]

---

[19] As with the selection factors, multiple items that survived the measurement model assessment for architecture success were averaged to create each success variable.

[20] Although the research model for architecture success shows no direct paths from data warehouse architectures to the dimensions of net benefits (i.e., individual impacts and organizational impacts) they were also included in the analyses together with the information and system quality variables to discover if architectures differed in terms of net benefits as well. The SEM analysis described tests the impacts of architecture on system product success as described in the research model for architecture success.

and (2) variables on development process success. The three-step data analyses for the system product success variables are described first.

7.6.1.1 System product success - step one

Prior to applying MANOVA, the multivariate normality of the data was assessed using PRELIS. Multivariate kurtosis was 1.23, which is well below the threshold of 2.0. In addition, Bartlett's Test of Sphericity indicated there were significant collective intercorrelations between dependent variables, providing justification for the use of MANOVA. Finally, analysis of the equivalence of the variance covariance matrices across groups indicated that the Box test was not significant at an alpha of 0.001.[21]

The overall multivariate test of the MANOVA indicated that there is a difference in the means of the success variables for the three categories of data warehouse architectures. The overall multivariate test results are presented in Table 7.29. Specifically, Wilks' lambda, the most common traditional test used when there are more than two groups of an independent variable, was significant at an 0.05 alpha level.

Table 7.29: MANOVA output – step one of system product success: overall multivariate test

Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	2.815	24.000	.000
Wilks' Lambda	2.859(b)	24.000	.000
Hotelling's Trace	2.902	24.000	.000
Roy's Largest Root	4.729(c)	1DBA0	.000

*alpha = 0.05

[21] An alpha level of 0.001 is recommended when analyzing groups with unequal sample sizes.

The univariate ANOVA tests assessing the impact of architecture on each success factor separately indicated that architecture significantly influenced all eight success variables (See Table 7.30).

Table 7.30: MANOVA output - step one of system product success: one-way ANOVA output

	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	42.54	2.00	21.27	10.38	0.00
Scalability	14.65	2.00	7.33	3.97	0.02
Integration	44.75	2.00	22.38	17.15	0.00
Accuracy	12.27	2.00	6.13	3.70	0.03
Completeness	19.34	2.00	9.67	3.70	0.03
Consistency	48.79	2.00	24.39	12.84	0.00
Individual impacts	19.54	2.00	9.77	7.37	0.00
Organizational impacts	9.84	2.00	4.92	3.33	0.04

*p=0.05

Next, pairwise multiple comparison tests were conducted to examine the exact nature of the influence of architecture on success factors. The Bonferroni method is recommended for multiple comparisons when the number of groups is small. The Bonferroni pairwise comparison tests the mean difference of a success variable between any two architecture categories. The results of this test are presented in Table 7.31. Overall, the results indicated that success variables are significantly different for the IDM architecture category when compared to one or both other two architecture categories. Specifically, the mean of flexibility, scalability, integration, consistency, and individual impacts was less for the IDM architecture in pairwise comparisons with DBA and EDW architecture. Furthermore, organizational impacts and completeness were significant and their means were lower for the IDM when compared to the EDW architecture. The IDM significantly differed from DBA architecture in terms of the mean of accuracy. In contrast, none of the means of the success variables were different for pairwise comparisons between DBA architecture and EDW architecture at $p < 0.05$ or $p < 0.1$.

Statistical power analysis was conducted to determine whether the insignificant results were due to the relatively small sample size. The Eta squared for the overall model was 0.17 and suggested a medium effect size. In order to achieve a power of 0.80, the sample size requirements for a three group, eight dependent variable design is 72 responses per group (Lauter 1978). The smallest architecture category (i.e., IDM architecture) had only 43 responses and did not meet these minimum requirements. As such, it is possible that non-significant results in this current MANOVA analysis were due to inadequate sample size.[22]

7.6.1.2 System product success - step two

In order to further identify how a particular data warehouse architecture category affects success measures, a set of MANOVA analyses was conducted. These analyses assessed the influence of each data warehouse architecture category on success measures when compared to the rest of the architectures combined. The first MANOVA considered the grouping of IDM architecture versus the rest of the architectures.

The results of the MANOVA for IDM architecture versus the rest of the architectures are presented in Table 7.32. The overall multivariate test was significant, indicating that success measures differ between the IDM and the rest of the architectures. The univariate ANOVA tests for each success variables were significantly different for the IDM architecture versus the rest (See Table 7.32).

[22] The statistical power analysis for a MANOVA with only EDW and DBA revealed that there was adequate sample size to achieve 0.80 power. This suggests that only the non-significant results associated with the differences between IDM and the other two architectures maybe due to sample size inadequacy.

Table 7.31: MANOVA output - step one of system product success: multiple comparisons using Bonferroni method

Success variable	1st Data warehouse architecture	2nd Data warehouse architecture	Mean Difference of success variable between 1st and 2nd architecture	Significance
Flexibility	IDM	DBA	-1.1732*	.000
	IDM	EDW	-.9464*	.000
	DBA	EDW	.2267	.583
Scalability	IDM	DBA	-.6337*	.034
	IDM	EDW	-.6161*	.021
	DBA	EDW	.0175	1.000
Integration	IDM	DBA	-1.0446*	.000
	IDM	EDW	-1.1017*	.000
	DBA	EDW	-.0571	1.000
Accuracy	IDM	DBA	-.6393*	.021
	IDM	EDW	-.4760	.082
	DBA	EDW	.1633	.896
Completeness	IDM	DBA	-.5919	.140
	IDM	EDW	-.7324*	.021
	DBA	EDW	-.1405	1.000
Consistency	IDM	DBA	-1.1327*	.000
	IDM	EDW	-1.1359*	.000
	DBA	EDW	-.0031	1.000
Individual impacts	IDM	DBA	-.8061*	.000
	IDM	EDW	-.6045*	.005
	DBA	EDW	.2016	.455
Organizational impacts	IDM	DBA	-.4992	.077
	IDM	EDW	-.5139*	.035
	DBA	EDW	-.0146	1.000

* = The mean difference is significant at the .05 level.

On the other hand, the MANOVA for DBA architecture versus the rest of the architectures indicated that success variables could not be differentiated between the two groupings of architecture. The Wilks' lambda for the MANOVA was not significant at 0.05 level of alpha (See Table 7.33 for results). However, the univariate F tests indicated that individual impacts and flexibility were significantly different for the DBA architecture when compared to the rest of the architectures combined. At first glance, this statistic suggests that the DBA architecture maybe more flexible in adapting to changes, as well as could result in more individual impacts when compared to the rest of the architectures.

Table 7.32: MANOVA output - step two of system product success: IDM vs. rest

IDM vs. the rest of the architectures

Multivariate tests of significance

Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	5.713	8.000	.000
Wilks' Lambda	5.713	8.000	.000
Hotelling's Trace	5.713	8.000	.000
Roy's Largest Root	5.713	8.000	.000

Univariate F tests

	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	39.069	1	39.069	19.024	.000
Scalability	14.630	1	14.630	7.946	.005
Integration	44.533	1	44.533	34.213	.000
Accuracy	10.471	1	10.471	6.306	.012
Completeness	18.009	1	18.009	6.895	.009
Consistency	48.787	1	48.787	25.744	.000
Individual impacts	16.802	1	16.802	12.637	.000
Organizational impacts	9.829	1	9.829	6.662	.010

*p=0.05

Table 7.33: MANOVA output - step two of system product success: DBA vs. rest

DBA vs. the rest of the architectures

Multivariate tests of significance

Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	1.464	8.000	.169
Wilks' Lambda	1.464	8.000	.169
Hotelling's Trace	1.464	8.000	.169
Roy's Largest Root	1.464	8.000	.169

Univariate F tests

	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	10.292	1	10.292	4.823	.029
Scalability	.986	1	.986	.525	.469
Integration	1.061	1	1.061	.746	.388
Accuracy	4.114	1	4.114	2.451	.118
Completeness	.032	1	.032	.012	.913
Consistency	2.347	1	2.347	1.159	.282
Individual impacts	6.385	1	6.385	4.700	.031
Organizational impacts	.338	1	.338	.225	.635

*p=0.05

Finally, the overall multivariate test for the MANOVA for EDW versus the rest was significant (See results in Table 7.34). The ANOVA tests indicated that consistency and integration were significantly different for EDW when compared to the rest of the architectures. The results of the ANOVA tests are presented in Table 7.34.

Table 7.34: MANOVA output - step two of system product success: EDW vs. rest

EDW vs. the rest of the architectures			
Multivariate tests of significance			
Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	2.443	8.000	.014
Wilks' Lambda	2.443	8.000	.014
Hotelling's Trace	2.443	8.000	.014
Roy's Largest Root	2.443	8.000	.014

Univariate F tests					
	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	1.530	1	1.530	.709	.400
Scalability	2.688	1	2.688	1.434	.232
Integration	12.244	1	12.244	8.799	.003
Accuracy	.094	1	.094	.055	.814
Completeness	8.904	1	8.904	3.376	.067
Consistency	10.561	1	10.561	5.276	.022
Individual impacts	.182	1	.182	.132	.717
Organizational impacts	2.418	1	2.418	1.616	.204

*p=0.05

7.6.1.3 System product success - step three

In step three, another set of MANOVA analyses was conducted to look at the impact of architecture on success measures when considering two architectures at a time. The first MANOVA was conducted to see if success factors could be differentiated for the IDM when compared to the DBA architecture. The overall Wilk’s Lambda statistic was significant and the univariate ANOVA tests showed that each success measure could be differentiated between the

IDM and DBA architectures except for completeness. The results of MANOVA for IDM compared to DBA are presented in Table 7.35.

The multivariate test for the MANOVA for the IDM compared to the EDW architecture was also significant. The univariate F tests for each success measure indicated that success measures differed for the IDM when compared to the EDW architecture (See Table 7.36 for results). In contrast, the overall multivariate test for the MANOVA for DBA compared to EDW was not significant and none of the univariate F tests were significant. The results of the MANOVA for DBA compared to EDW are presented in Table 7.37. Overall, the outcome of the three MANOVAs conducted in step three presented very similar results to the results described in the multiple comparisons of the MANOVA in the first step with a few exceptions.

Table 7.35: MANOVA output - step three of system product success: IDM vs. DBA

IDM vs. DBA architecture

Multivariate tests of significance

Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	3.350	8.000	.002
Wilks' Lambda	3.350	8.000	.002
Hotelling's Trace	3.350	8.000	.002
Roy's Largest Root	3.350	8.000	.002

Univariate F tests

	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	41.005	1	41.005	17.945	.000
Scalability	11.963	1	11.963	5.902	.016
Integration	32.509	1	32.509	20.324	.000
Accuracy	12.175	1	12.175	6.698	.011
Completeness	10.437	1	10.437	3.721	.056
Consistency	38.227	1	38.227	15.544	.000
Individual impacts	19.359	1	19.359	12.257	.001
Organizational impacts	7.426	1	7.426	4.423	.037

*p=0.05

Table 7.36: MANOVA output - step three of system product success: IDM vs. EDW

IDM vs. EDW architecture

Multivariate tests of significance

Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	5.716	8.000	.000
Wilks' Lambda	5.716	8.000	.000
Hotelling's Trace	5.716	8.000	.000
Roy's Largest Root	5.716	8.000	.000

Univariate F tests

	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	32.243	1	32.243	14.964	.000
Scalability	13.664	1	13.664	7.235	.008
Integration	43.692	1	43.692	33.509	.000
Accuracy	8.154	1	8.154	4.919	.027
Completeness	19.309	1	19.309	7.461	.007
Consistency	46.441	1	46.441	26.558	.000
Individual impacts	13.156	1	13.156	9.953	.002
Organizational impacts	9.505	1	9.505	6.754	.010

*p=0.05

Table 7.37: MANOVA output - step three of system product success: DBA vs. EDW

DBA vs. EDW architecture

Multivariate tests of significance

Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	1.376	8.000	.206
Wilks' Lambda	1.376	8.000	.206
Hotelling's Trace	1.376	8.000	.206
Roy's Largest Root	1.376	8.000	.206

Univariate F tests

	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Flexibility	3.466	1	3.466	1.863	.173
Scalability	.021	1	.021	.012	.913
Integration	.220	1	.220	.187	.666
Accuracy	1.798	1	1.798	1.128	.289
Completeness	1.331	1	1.331	.521	.471
Consistency	.001	1	.001	.000	.985
Individual impacts	2.739	1	2.739	2.248	.135
Organizational impacts	.014	1	.014	.010	.921

*p=0.05

7.6.1.4 Development process success – step one

MANOVA was also used to look at the influence of architecture on the development time and development cost. Architecture was entered as the independent variable while initial cost, annual cost, overall cost, and initial time were entered into the analysis as the dependent variables.

The overall multivariate test for the MANOVA conducted in step one was significant, indicating that cost and time variables can be differentiated by data warehouse architectures (See Table 7.38 for results). The univariate ANOVA indicated that each of the four variables could be differentiated by architectures (Table 7.38). The multiple comparisons using the Bonferroni method (See Table 7.39) indicated that initial cost, annual cost, and overall cost could be significantly differentiated by the EDW architecture when compared to the other two. The corresponding mean differences for the EDW architecture were higher, indicating that each of the cost measures were higher for the EDW architecture when compared to the other two architectures. While the initial time variable was not significant at a 0.05 alpha level, at 0.1, initial time could be differentiated for EDW when compared to the DBA architecture. As with the cost measures, initial time taken was higher for the EDW architecture. None of the variables could be significantly differentiated between DBA and IDM architecture.

An assessment of sample size adequacy revealed that for the medium effect size present in the current analysis, a minimum sample size of 56 is required per group to achieve a power of 0.80. The minimum sample size threshold was above the sample size for the smallest architecture category. As such, it is possible that insignificant results may have been due to inadequate sample size.

Table 7.38: MANOVA output - step one of development process success

Multivariate tests of significance			
Test of group difference	F value	Degrees of freedom	Significance*
Pillai's Trace	2.188	8.000	.027
Wilks' Lambda	2.217	8.000	.025
Hotelling's Trace	2.245	8.000	.023
Roy's Largest Root	4.458	4.000	.002

Univariate F tests					
	Sum of Squares	Degrees of freedom	Mean Square	F value	Significance*
Initial cost	16.613	2	8.307	4.933	.008
Annual cost	17.077	2	8.538	6.680	.001
Overall cost	29.783	2	14.892	6.993	.001
Initial time	7.419	2	3.709	3.670	.027

*p=0.05

Table 7.39: MANOVA output: multiple comparisons using Bonferroni method

Success variable	1st Data warehouse architecture	2nd Data warehouse architecture	Mean Difference of success variable between 1st and 2nd architecture	Significance
Initial cost	IDM	DBA	-.22	1.000
	IDM	EDW	-.68*	.045
	DBA	EDW	-.45*	.043
Annual cost	IDM	DBA	-.22	1.000
	IDM	EDW	-.68*	.015
	DBA	EDW	-.46*	.013
Overall cost	IDM	DBA	-.39	.734
	IDM	EDW	-.96*	.007
	DBA	EDW	-.57*	.019
Initial time	IDM	DBA	-.09	1.000
	IDM	EDW	-.41	.167
	DBA	EDW	-.33	.069

* = The mean difference is significant at the .05 level

The results of the MANOVA analyses in step two are summarized next. The Wilk's Lambda for the overall multivariate test for IDM compared to the rest was not significant. However, univariate F tests indicated that IDM could be significantly differentiated from the others in terms of annual cost and overall cost. Similarly, the multivariate test for DBA compared to the rest was also not significant and the univariate ANOVA test for each variables showed that DBA could be differentiated from the rest in terms of annual cost and overall cost. Next, the overall multivariate test for the MANOVA for EDW versus the rest was significant and the univariate F tests for each variable was significant at the 0.05 alpha level.

The third and final step of the exploratory analysis for development process success variables (i.e., comparing a combination of two architectures at a time) produced the same results as the results produced in the pairwise multiple comparison tests conducted in the first step.

7.6.2 Architecture success: specific hypotheses

Each of the specific hypotheses on success described in Chapter four, except for ones that referred to the FED architecture, was tested. Each of the hypotheses that described a comparison of FED and another architecture was not tested due to very low sample size for FED architectures. The results from the univariate ANOVA tests conducted in testing the exploratory hypothesis on success provided the needed information to test each of the remaining hypotheses. Each of the hypotheses is listed along with the corresponding results from the ANOVA in Table 7.40.

Table 7.40: Specific hypothesis testing for architecture success

Hypothesis	**F value**	**Significance**	**Result**
Hypothesis SQ1a: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than an IDM architecture.	14.96	0.000	Supported
Hypothesis SQ1b: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than a DBA architecture.	1.86	0.173	Not supported
Hypothesis SQ1c: An EDW architecture is more likely to be associated with the development of a flexible data warehouse solution than a FED architecture.			Not tested
Hypothesis SQ2: An EDW architecture is more likely to lead to the development of a data warehouse solution that enables high data integration than an IDM architecture.	33.51	0.000	Supported
Hypothesis SQ3: An EDW architecture is more likely to be associated with the development of a scalable data warehouse solution than a DBA architecture.	0.012	0.913	Not supported
Hypothesis IQ1a: An IDM architecture solution is less likely to be associated with high data consistency than an EDW architecture solution.	26.558	0.000	Supported
Hypothesis IQ1b: An IDM architecture solution is less likely to be associated with high data consistency than a DBA architecture solution.	15.544	0.000	Supported
Hypothesis IQ1c: An IDM architecture solution is less likely to be associated with high data consistency than a FED architecture solution.			Not tested

7.6.3 Structural model for research question on success

In addition to using MANOVA, an attempt was made to understand the impact of data warehouse architectures on success factors using structural equation modeling. The data warehouse architectures were included in the analysis as dummy variables with the EDW architecture as the reference dummy variable.

The structural model examined the impact of architecture on system product success, which included the following second order latent variables: system quality, information quality, and net benefits. The paths from each architecture variable to system quality and information quality were of primary interest in the specified model. In addition, the paths from information quality and system quality to net benefits were also tested. The graphical representation of the structural model is presented in Figure 7.5. The fit statistics for the model indicated sufficient goodness-of-fit. The fit statistics for the model are listed in Table 7.41.

Table 7.41: The goodness of fit statistics the structural models testing influence of architecture on architecture success factors

Measure	**Target value**	**The structural model**
Chi-square	< 3 times degree of freedom	825.96
RMSEA	<0.08	0.063
SRMR	<0.10	0.21
NNFI	>0.90	0.98
AGFI	>0.80	0.83

*340 Degrees of freedom

The results of the hypothesized causal relationships displayed graphically in Figure 7.5 are listed in Table 7.42. The results indicated that when compared to the EDW, the IDM architecture had a significant lower influence on information quality and system quality. In contrast, the DBA was not significantly different from the EDW architecture in its influence on information quality and system quality. The R-squared statistics that corresponded to the relationship between architecture and quality factors indicated that architecture explains only 5 percent of the variance in information quality and only 7 percent of the variance in system quality. On the other hand, the structural model indicated that information quality and system quality explain 71 percent of the variance in net benefits.

Table 7.42: SEM results: Standardized parameter estimates for the second order structural model

	IDM architecture	DBA architecture	R Squared
Information Quality	-0.21 (4.35)	0.02 (0.60)	0.05
System Quality	-0.26 (3.65)	0.04 (0.32)	0.07
	Information Quality	System Quality	R Squared
Net benefits	0.67 (8.53)	0.47 (11.21)	0.71

The t-value for each estimate is presented in parentheses

The reduced form equations from the LISREL output indicated that if a causal path was estimated between the IDM architecture and each of the first order success variables in the model (such as flexibility, scalability, etc…), the IDM architecture would have a significantly lower impact on the success variables when compared to the EDW architecture. In contrast, the reduced form equations indicated that the DBA was not significantly different from the EDW architecture in terms of its influence on the first order success variables in the analysis.

Overall, the SEM assessment of the impact of architecture on system product success factors confirmed the broad findings of the set of MANOVA analyses, providing further assurance for the validity of results for architecture success.

7.7 Conclusion

This chapter described the analysis of data to test the hypotheses on architecture selection and success. First, the measurement model for architecture selection and success was thoroughly assessed. Next, in systematic steps, a set of analyses were undertaken to assess the influence of architecture selection factors on data warehouse architecture. In a similar manner, the influence of architecture on success variables was determined. In the next chapter, the results of the data analyses are discussed in detail.

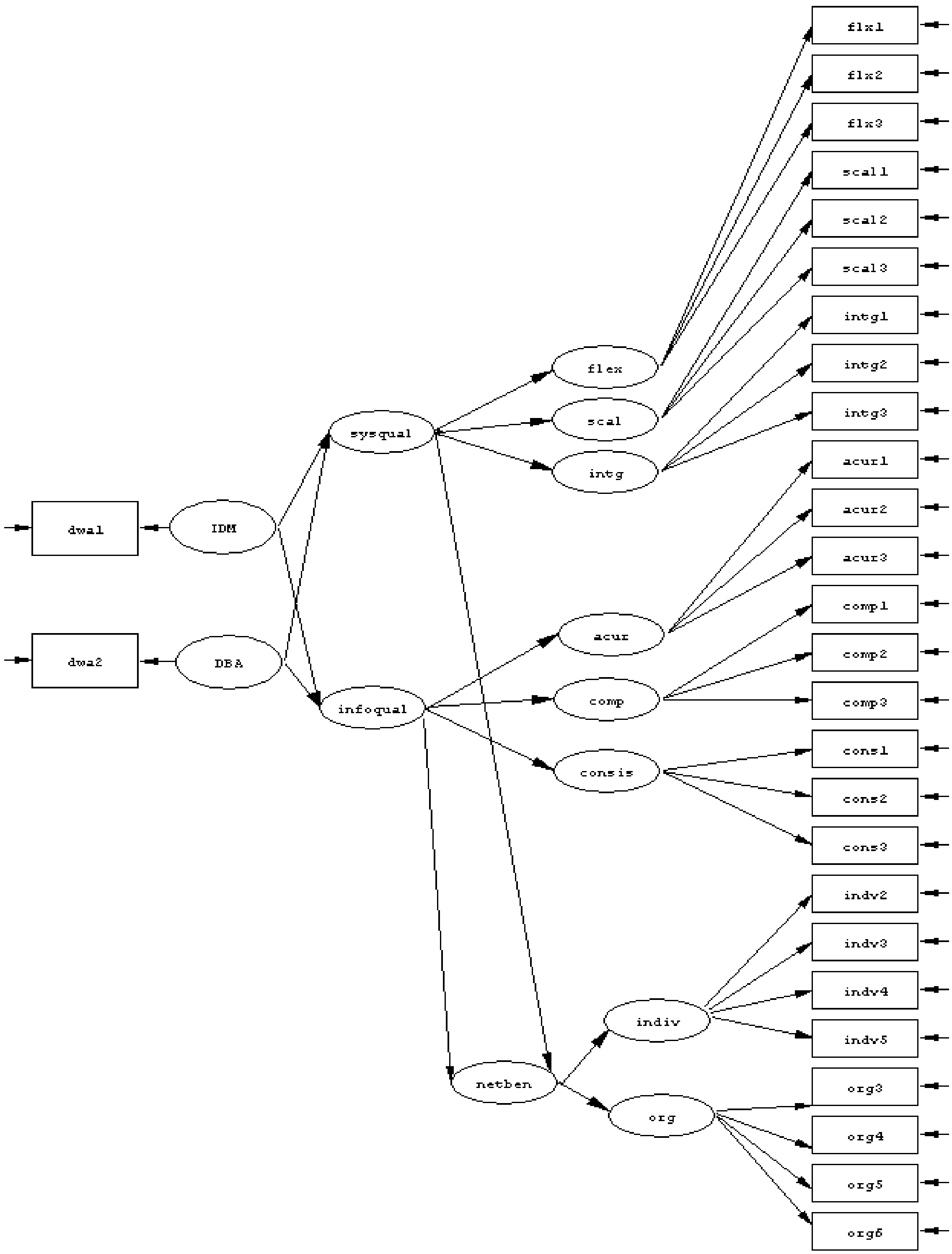


Figure 7.5: The structural model for system product success

CHAPTER 8 – DISCUSSION

8.1 Introduction

In this chapter, the findings from the research are discussed. The findings are two-fold. First, the findings related to architecture selection are presented followed by the findings on architecture success.

8.2 Architecture selection

The research question on architecture selection required investigating the relative importance of organizational selection factors. Chapter seven reported the results of the tests of specific hypotheses and the exploratory hypothesis on architecture selection. Prior to presenting the possible explanations for the unsupported hypotheses, the overall findings on architecture selection are discussed.

8.2.1 Overall findings on architecture selection

The overall findings from architecture selection are examined in two ways. First, the influence of each factor on architecture selection as a whole is considered. Second, the findings are discussed in terms of the factors that were identified as important in influencing the selection of a particular architecture versus another.

8.2.1.1 Selection factor influence on overall architecture selection

It is first useful to look at the influence of the individual selection factors on architecture selection in general. In this section, the influence of each selection factor is described. To facilitate the presentation of findings, at times the influence of several selection factors are described together.

8.2.1.1.1 *View of the warehouse, the perceived ability of the IT staff, and resource availability*

The findings showed that the view of the data warehouse, the perceived ability of the IT staff, and resource availability are the factors with the greatest predictive power. The importance and influence of these factors in IS have been recognized in the adoption and contingency literature. Past studies lend credence to the importance of these factors in this study.

View of the warehouse

Based on case studies of 50 organizations, Broadbent et al (1997) suggest that when IT is viewed as being strategic, organizations invest in extensive firm-wide IT infrastructure services in a centralized way. On the other hand, a view of IT as not being strategic leads a firm to invest in IT services in a less integrative, comprehensive manner. The results of this study corroborate Broadbent et al. (1997)'s findings. When a data warehouse is viewed as infrastructure, it is more likely that a firm selects an EDW or a DBA architecture rather than an IDM. Both EDW and DBA provide an enterprise view of the data and are generally developed to support multiple decision-support initiatives. In contrast, when viewed as strategic infrastructure, an organization is less likely to implement an IDM architecture, which is often described as a relatively inexpensive, quick point solution that meets an individual business unit's needs.

The findings further suggest that as the view of the warehouse as strategic increases, organizations are more likely to implement an EDW rather than a DBA architecture. At first glance, the difference in view of the warehouse between the two architectures is not easy to understand. However, a closer analysis of their characteristics suggests that there may be other organizational conditions that work together to influence the architecture choice. This possibility is further explored later in this section.

Perceived ability of the IT staff

Other studies also provide support for the influence of perceived IT ability on architecture selection. Teo et al's (1995) study of the adoption of financial EDI in Singapore indicated that the perceived complexity of an EDI was a key inhibitor to the intent to adopt. As previously described, the perceived ability of the IT staff to implement a data warehouse changes based on the perceived complexity of the architecture. Consequently, organizations are influenced to adopt a particular architecture based on the perceived ability of the IT staff.

The literature also provides some details of the nature of the relationship between perceived ability and infrastructure. Weill et al. (1998) present evidence that more skilled IT staffs are associated with infrastructures with more capabilities. In addition, Byrd et al. (2001) discovered that high perceived technical skills were important to achieve greater IT infrastructure flexibility.[23] In the context of this study, the findings indicated that when faced with low perceived ability, organizations are more likely to select an IDM architecture. An IDM architecture leads to the implementation of a point solution that does not provide the infrastructure flexibility of an EDW or a DBA (Hackney 1998). On the other hand, when the perceived ability of the IT staff is high, organizations are more likely to implement an EDW or a DBA.

[23] IT infrastructure flexibility was measured in terms of integration and modularity.

The results further suggest that the perceived ability of the IT staff does not influence the selection between EDW and DBA architectures. According to Breslin (2003), implementing an EDW architecture requires greater specialized skills and experience from the implementation team than needed for a DBA implementation. As such, organizations that house an IT staff with low perceived ability should be more prone to implement a DBA architecture. The inspection of the means of perceived ability for both architectures indicates that the mean for DBA was considerably lower than the mean for EDW. Therefore, the difference in means was in the anticipated direction. Statistical power analysis was carried out in order to identify whether the non-significant result was due to the sample size. The power analysis indicated that in order to detect a medium effect (Cohen 1988) with 0.80 power, a total sample size of at least 1,500 responses would be required (Hsieh et al. 1998). Therefore, it is possible that the non-significant relationship was due to the study's sample size.

Resource availability

Previous studies also indicate some support for the influence of resource availability on architecture selection. According to Iacovou et al (1995), the organizational readiness to adopt an EDI is influenced by financial resource availability. The empirical investigation of the factors that influence the adoption of EDI, by Chwelos et al. (2001), suggests that the perceived readiness of the organization influences the intent to adopt an EDI. By the same token, the results of this study showed that the perceived constraints on resource availability influence the selection of a particular data warehouse architecture.

According to Grover et al. (2004), when organizations consider IT investment decisions, they may be more prone to invest in IT infrastructure when slack resources are available. The current research supports this finding. When resource constraints are low, a firm is more likely to select an EDW or DBA rather than an IDM architecture. Much like the influence of perceived ability of the IT staff, the findings indicate that resource availability had no influence on the selection between the DBA and EDW architectures. This was not as expected. Again, the mean difference of resource availability for DBA and EDW was in the direction anticipated. An examination of sample size adequacy to test this relationship revealed that the number of observations were insufficient to determine a medium effect size for the relationship.

8.2.1.1.2 *Vertical information flow*

According to the data analysis, vertical information flow did not have a significant influence on the choice of an IDM versus an EDW architecture. The initial expectation was that as the demand for detailed data from lower levels of the organization rises, organizations would opt to select an EDW. When a system design is implemented to meet the needs of a local functional unit, consistently integrating it to meet upper management's needs for lower level data is difficult (Goodhue et al. 1992). The IDM architecture is often implemented to meet the unique needs of a specific area. As such, it may not satisfy management's need for lower level data. As a result, this finding was puzzling.

Furthermore, the results also suggest that as the upper management's need for detailed information from lower levels of the organization increases (i.e., vertical information flow increases), organizations are more likely to implement a DBA rather than an EDW architecture. This finding was also somewhat surprising since the ability to drill down to atomic level data is publicized as a key feature of an EDW. Closer scrutiny of the two architectures suggested a possible explanation.

From a technical standpoint, it is feasible to have the capability to drill down to data in lower levels of the organization with both architectures. However, a DBA architecture is designed specifically to provide easy and fast access to multidimensional data. In an EDW architecture, the central data store with relational data structures is implemented first, after which dependent marts are created with multidimensional data. Therefore, as the findings suggest, when faced with high vertical information flow, organizations may be more likely to implement a DBA than an EDW.

The above explanation suggests that organizations chose a DBA over an EDW due to the *ease* with which a DBA gives the ability to drill down to detailed information from lower levels. The data analysis also suggests that vertical information flow does not influence the selection of an IDM when compared to an EDW. Given these findings and the possible explanations for non-significance, it is possible that this selection factor does not correctly capture organizations' mental model of vertical information flow. Furthermore, during the measurement model assessment, one item from this scale was removed as it had a very low factor loading and there was a possibility that respondents may have conceptualized it differently than intended. This further suggests that there may have been a problem with the manner in which the respondents interpreted vertical information flow.

From a theory perspective, vertical information flow and horizontal information interdependence were included in the research model to examine the importance of horizontal and vertical information sharing on the selection of an architecture based on the concept of interdependence in the literature. Since the importance of interdependence is captured in horizontal information interdependence, perhaps omitting the other does not hinder the investigation of the research question. Furthermore, the ability to drill down to lower levels of granularity within a given domain has become more of a norm rather than a special capability of a particular architecture. According to Codd's 12 OLAP rules, the ability to query to the desired level of data granularity is a basic OLAP capability.[24] As such, perhaps this variable does not help differentiate between architectures in the manner that was expected. Consequently, it was omitted from the discussion of results with the hope of investigating and re-conceptualizing it in the future.

8.2.1.1.3 *Urgency*

Another finding from the data analyses was that increases in the urgency to implement a data warehouse influence the likelihood of selecting a DBA architecture when compared to EDW. While this finding was only marginally significant (i.e., at $p < 0.1$), this result confirmed the majority opinion of authors in the data warehousing literature (e.g., Hackney 1998; Joshi 1999) and the opinion of experts that were interviewed. As a result, this finding was interpreted as a significant result of the dissertation.

The time required to implement an EDW architecture is higher than a DBA architecture (Hackney 2000a). For example, creating an enterprise data model is a much more extensive and time-consuming process in implementing an EDW. The finding also provided support for the OIPT theory (Tushman et al. 1978) by showing that as urgency increases, organizations are more likely to implement an architecture that, at least initially, provides lower information processing capacity.

Urgency did not have a significant influence on the selection of an IDM architecture when compared to the rest. The IDM architecture is recognized as a relatively inexpensive,

---

[24] The 12 OLAP evaluation rules are largely accepted within the IT industry and are used by vendors to revise their products to better conform to Dr Codd's OLAP criterion. (Watson et al. 1992).

small, and quick solution to fulfill specific needs in an organizational unit (Marco 2003). As such, it would seem that when urgency is high, an organization would opt to build an IDM architecture rather than an EDW or a DBA. For instance, when compared to an IDM, both architectures require greater investment in upfront planning to create the enterprise data model. The history of IDM implementations may offer a possible reason for the non-significant result obtained.

At the beginning of the decision support systems evolution, independent data marts were introduced and implemented in organizations to meet task specific needs for multidimensional data that transaction processing systems could not satisfy. Overtime, organizations have recognized that while an IDM might be an expedient, point solution, it does not support a broad decision support environment. As such, though there may be some IDM architecture implementations that were driven by urgency for a quick solution, most implementations in organizations today may not have been influenced by urgency. They were simply the norm at the time for creating a repository of decision support data.

8.2.1.1.4 *Existing system compatibility and expert influence*

The importance of existing system compatibility and expert influence on the selection of a particular architecture could not be directly assessed. These factors were not measured in a manner that would permit such an analysis. However, t-tests were used to investigate whether these factors influence architecture selection in general.

The findings suggested that organizations that implemented a DBA architecture were more influenced by the literature on data warehousing than the other two architectures. The set of books written by Ralph Kimball (i.e., the authority on the DBA architecture) provide specific, detailed instructions on how to build a DBA data warehouse. There is no recent literature that advocates the use of an IDM. It is possible that the influence of literature on the selection of DBA over the other architecture stems from the popularity of Ralph Kimball's books.

8.2.1.1.5 *Horizontal information interdependence, task routineness, and source of sponsorship*

The MLR data analyses on architecture selection suggested that horizontal information interdependence, task routineness, and source of sponsorship do not have a significant impact on the selection of any data warehouse architecture, while individual BLR analyses suggested otherwise. The results obtained when multiple selection factors were entered together were contrary to expectations. The inclusion of interdependence, task, and level of sponsorship constructs as selection variables is well supported by theory. An examination of the means of each selection factor for each architecture category (Refer to Table 7.16a, b, and c) indicated that there were considerable differences between the means for the architectures. For instance, the mean of interdependence for IDM when compared to EDW and DBA was considerably lower. The means of task and sponsorship revealed similar results.

The multinomial logistic regression analysis, with interdependence as the only predictor variable and architecture as the dependent variable, indicated that as interdependence decreases, organizations are more likely to implement an EDW or a DBA than an IDM. A similar outcome was obtained from the analysis of the task construct. The MLR analyses further indicated that interdependence or task do not influence the likelihood of selecting a DBA versus an EDW, suggesting that there is no difference in the influence of interdependence and task between the two. However, this finding was not counterintuitive. For instance, both DBA and EDW are recognized as data warehouse architectures that lead to the creation of an integrated data solution to satisfy high information sharing needs among organizational units. While a DBA architecture provides a logically integrated data solution to satisfy high interdependence,[25] the EDW architecture provides a physically integrated solution (Eckerson 2002). Likewise, it is possible that EDW and DBA do not differ in terms of their support for task routineness. According to the authorities, both architectures provide a data warehouse that can hold an abundance of atomic and summarized level data (Inmon et al. 2001; Kimball et al. 2003). As such, as task non-routineness increases, both a DBA and an EDW architecture can satisfy user needs to run queries in novel ways by combining data elements with various degrees of granularity.

An additional MRL carried out using source of sponsorship as the predictor variable indicated that the level of sponsorship influences the likelihood of selecting both the IDM and DBA architectures when compared to an EDW. Specifically, the analysis suggested that as the level of sponsorship within the organization decreases, firms are more likely to implement an IDM when compared to the rest. Further, when considering the choice of selecting between DBA and EDW, as the level of sponsorship rises, firms are more likely to choose an EDW architecture. These relationships were expected and were included as specific hypotheses.

In the case of all three constructs, when strategic view was added to the MLR model, the influence of interdependence, task, and level of sponsorship on the likelihood of selecting an IDM when compared to EDW became non-significant. In addition, when strategic view was entered into the MLR, the influence of the level of sponsorship on the likelihood of choosing a DBA versus an EDW also became non significant.

---

[25] Although a DBA architecture implementation begins with implementing a data mart, which is of considerably smaller scope than a EDW architecture implementation, it is built according to an enterprise level data model with the view of logically integrating architecture marts. At first glance, an EDW architecture seems the more appropriate choice when faced with high interdependence as it creates an in- depth enterprise data model prior to building an enterprise data warehouse. However, it is possible that organizations do not perceive a distinction between selecting an architecture that creates a physically integrated enterprise data warehouse or a logically integrated set of architecture data marts.

This initial analysis suggests that strategic view may be mediating the relationship between these selection constructs and architecture selection. From a data warehousing standpoint, it is conceivable that the need to satisfy high interdependence, non-routine task requirements, and an increasing the level of sponsorship may lead an organization to view the implementation of a data warehouse as a strategic infrastructure initiative. Many past case studies on data warehousing, such as the implementation of a data warehouse at the Blue Cross Blue Shield of North Carolina (BCBS) (Watson et al, forthcoming), provides evidence for the possibility of this mediating relationship. At BCBS, there was a need to consolidate dispersed data sources to facilitate greater information sharing across the organizations. High non-routine task requirements and sponsorship from the top management of the organization led the company to perceive the implementation of their data warehouse as an infrastructure investment that would help them achieve long-term strategic goals.

The existing IS literature also provides evidence of this possible mediating relationship. Information processing theory suggests that organizations may differ in the manner they perceive an IT solution based on the source of uncertainty they face. For example, a company faced with uncertainty from a dynamic environment may perceive an IT infrastructure as a mechanism that provides long-term competitive advantage (Duncan 1995). Interdependence and task routineness are two key sources of uncertainty identified in OIPT that shape organizational business needs.

According to Broadbent et al (1997) and Weill et al (1998), recognizing overall business needs in the company as whole, or the business needs of individual units, leads a company to identify business maxims. Business maxims guide the identification of IT maxims, which in turn shape the organization’s perception of the IT infrastructure solution considered for investment. In order to demonstrate the manner in which the view of IT changes based on business needs, the authors present case study illustrations of companies faced with various business needs, their business maxims, IT maxims, and the view of the IT infrastructure that emerged. Some of the characteristics of the business needs they present can be described as the ability to satisfy high information interdependence or task routineness within a company. Furthermore, based on their findings from multiple case analyses, Broadbent et al (1997) conclude that upper management sponsorship often influences the perception of IT. As the source of sponsorship rises, organizations are more likely to view IT as strategic infrastructure.

In order to test if interdependence, task, and sponsorship level influence the view of the warehouse, SEM was used. The fit statistics for the model were all higher than the recommended thresholds. The model indicated that all three factors have a significant, positive influence on view of the warehouse, explaining 39 percent of the variance in view. This analysis supports the possibility that strategic view mediates the relationship between the three selection variables and architecture selection.

Tests for mediation effects, using the causal steps approach described by Baron et al (1986), were conducted to assess whether the view of the warehouse mediated the influence of interdependence, task, and sponsorship level on selection of an IDM versus EDW. The first three causal steps of the mediation test for interdependence provided significant path coefficients. The fourth step, the path from interdependence to architecture selection while controlling for strategic view, was zero, suggesting that view of the warehouse fully mediates the relationship. Similar findings were obtained through the mediation tests for the remaining two selection factors. Thus, the tests indicated that view of the warehouse mediated the influence of interdependence, task, and sponsorship level on the likelihood of selecting an IDM when compared to an EDW. Overall, they suggested that increases in interdependence, task non-routineness, and level of sponsorship raises the organization's perception of the warehouse as a strategic infrastructure initiative. Consequently, a strategic view of the warehouse increases the likelihood of selecting an EDW architecture when compared to an IDM.

In a similar manner, a mediation test showed that strategic view mediated the relationship between sponsorship level and the selection of a DBA versus an EDW. The results suggest that as sponsorship for the data warehouse initiative rises within the organization, the view of the warehouse as a strategic infrastructure initiative increases, leading to the selection of an EDW. This conclusion is supported by the data warehousing literature. Generally, when business units sponsor data warehouse initiatives, when considering the choice of implementing an EDW or a DBA, they select a DBA architecture solution (e.g., Watson et al. 2001). This is due to the fact that a DBA architecture is a more feasible implementation for business units as it gives control to those units in addition to meeting their budgetary constraints.[26] For instance, it gives business units the opportunity to influence the development of conformed dimensions and facts used to implement a DBA. On the other hand, case studies reveal that when upper management sponsors a data warehouse initiative, they select an EDW architecture rather than a DBA to acquire a comprehensive solution that gives them a global view of the entire organization. The EDW architecture gives upper management the opportunity to influence the data management practices of the entire organization. Therefore, one possible reason for the difference in the influence of the view of the warehouse for both the architectures may stem from the difference in the source of sponsorship.

Furthermore, mediation tests confirmed that the influence of interdependence and task on the selection of an IDM when compared to a DBA is mediated by the view of the warehouse.

---

[26] As previously mentioned in the literature review, in addition to social political reasons such as to gain more control, a sponsor may influence the selection of an architecture driven by rational reasons as well.

Much like an EDW, the DBA architecture provides a global data warehouse solution to satisfy business needs that may arise from high interdependence and task non-routineness. Not surprisingly, source of sponsorship failed the test, which investigated whether the influence of the level of sponsorship on the selection of a DBA versus an IDM is mediated by strategic view. Source of sponsorship did not influence the selection of an IDM when compared to a DBA. The data warehousing literature suggests that when the sponsorship for a data warehouse initiative is from business units or functional areas, organizations are more likely to implement either a DBA or an IDM architecture (Hackney 2000a).

8.2.1.1.6 *Domain*

In addition to the selection variables described in the research model, the data analyses revealed that the domain of the business in which the architecture was implemented has an influence on architecture selection. The results suggest that as the domain of the business increases from a subunit to the entire company, organizations are more likely to select an EDW architecture than a DBA or IDM architecture. Furthermore, when considering just DBA and IDM, as the domain increases, organizations are more likely to implement a DBA architecture. These findings are consistent with the opinions and prescriptive advice on data warehousing practices offered in the practitioner literature (e.g. Hackney 2000a; Griffin 1998). According to Griffin (1998), when the scope of the implementation is small, such as in a functional area, an IDM maybe more appropriate than the other architectures. Consequently, Goodhue et al. (1992) suggest that when the domain of system design is a local business unit, combining data elements in a consistent manner to facilitate global integration becomes challenging. In the IS academic literature, the importance of domain to data management has been previously suggested. In

Goodhue et al.'s (1988) investigation of organizational data management practices, the authors identify scope as an important factor that influences data management choices.

8.2.1.2 Selection factor influence on one architecture compared to another

The exploratory hypothesis analysis revealed that some selection factors appear to be more important than others in the selection of a particular architecture. The overall findings on architecture selection suggest that different combinations of certain factors affect the likelihood of selecting one architecture when compared to another. In addition, the tests for mediation effects provided some insight on how some of the selection factors might influence architecture selection.

Three selection factors emerged as being important in choosing between the EDW and IDM architectures. They are the perceived ability of the IT staff, resource availability, and view of the warehouse. The analysis suggested that on average, based on the nature of these selection factors, the likelihood of selecting one architecture over the other varies. Specifically, when considering the selection of an EDW as compared to an IDM architecture, when the perceived ability of the IT staff is high, the resource constraints are low, and the organization's perception of a warehouse as a strategic infrastructure initiative is high, there is a greater the likelihood that a firm will select an EDW. In addition, further analysis revealed that interdependence, task routineness, and the source of sponsorship influence the selection of an EDW through strategic view. When interdependence is high, task routineness is low, and the level of sponsorship is high, organizations perceive the implementation of a data warehouse as a strategic initiative. A model describing the influence of selection factors on the selection on an EDW architecture when compared to an IDM architecture is presented in Figure 8.1.

According to the exploratory analysis, two selection factors appear to be significant when deliberating selecting between EDW and DBA. They are urgency and view of the warehouse. Specifically, when considering the selection of an EDW, as the urgency to build a data warehouse decreases and strategic view of the warehouse increases, organizations are more likely to select an EDW. Further analysis indicated that the source of sponsorship influences the selection of an EDW when compared to a DBA through view of the warehouse. When the sponsorship level is very high, an organization is more likely to perceive the implementation of a data warehouse as a strategic infrastructure project of high importance, leading to the selection of an EDW architecture. A model describing the influence of selection factors on the selection on an EDW architecture when compared to a DBA architecture is presented in Figure 8.2.

Finally, three factors surfaced as being important to selecting between a DBA and an IDM. They are perceived ability of the IT staff, view of the warehouse, and resource availability. In particular, when considering the selection of a DBA, on average, the higher the perceived ability of the IT staff and the organization's perception of a warehouse as a strategic infrastructure initiative, and lower the resource constraints, the greater the likelihood that an organization selects a DBA.[27] Further analysis indicated that interdependence and task routineness also influence the selection of a DBA. Specifically, when interdependence is high and task routineness is low organizations are more likely to view the implementation of a data warehouse as an infrastructure project, leading to the selection of a DBA over an IDM. A model describing the influence of selection factors on the choice of a DBA when compared to an IDM architecture is presented in Figure 8.3.

[27] Dividing the odds ratios for IDM versus EDW by DBA versus EDW gave the parameter estimates needed to examine the odds of IDM versus DBA.

Next, possible explanations for the unsupported hypotheses on architecture selection are presented.

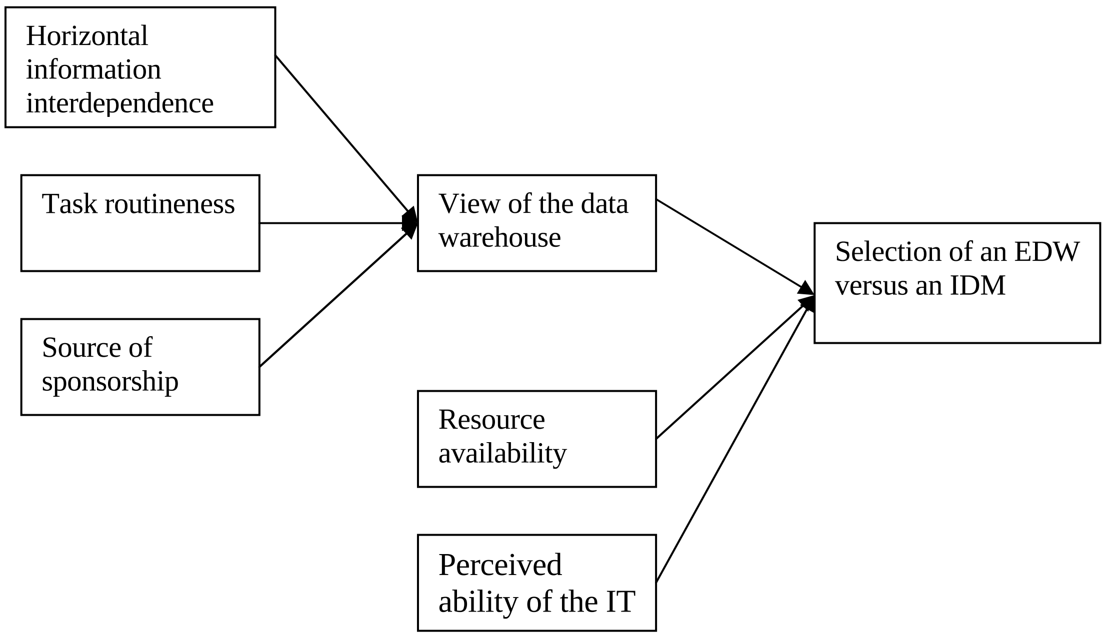


Figure 8.1: Model for factors that influence the selection of an EDW an IDM

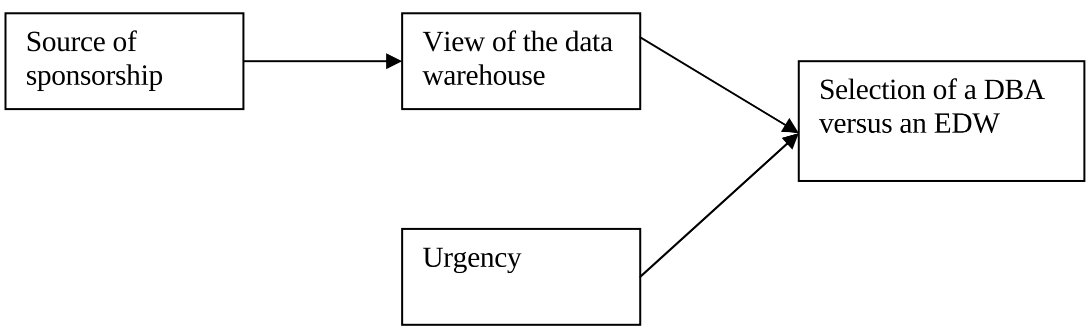


Figure 8.2: Model for factors that influence the selection of an EDW versus a DBA

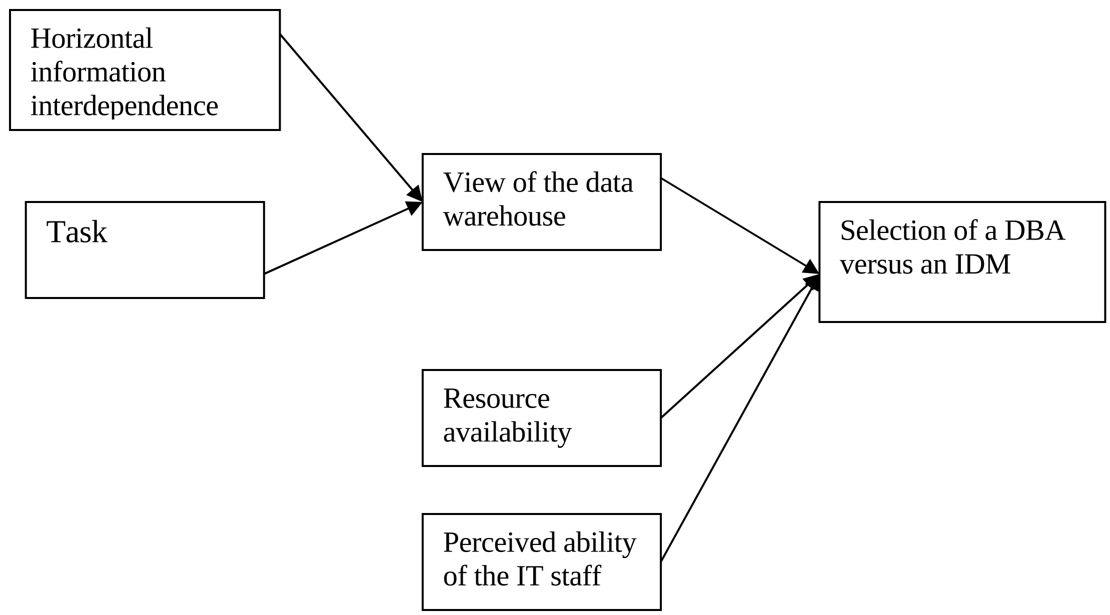


Figure 8.3: Model for factors that influence the selection of a DBA versus an IDM

8.2.2 Specific hypotheses on architecture selection

Four of the specific hypotheses, which described the influence of a selection factor on the likelihood of selecting one of two architectures, were not supported. Table 7.28 presents the results of the hypothesis testing. Two of them described the influence of a selection factor on the likelihood of selecting an IDM when compared to an EDW architecture.

The first looked at the influence of urgency. The findings from the data analyses showed that this relationship was not significant. As suggested previously, the history of the IDM architecture may provide a possible reason. Historically, IDM architecture implementations were driven by the need for data to support specific task needs that legacy systems could not satisfy. While it is a relatively quick implementation with fast project turnaround, it is possible that the majority of IDM implementations were not driven by urgency.

The second hypothesis described the influence of vertical information flow. The hypothesis suggested that as upper management's needs for detailed information from lower levels of the organization increases, a firm is more likely to implement an EDW rather than an IDM architecture. As discussed previously in the overall results on architecture selection, vertical information flow may not have been accurately captured in this study. As such, it is difficult to determine the influence of this variable on architecture selection.

The third and fourth non-significant hypotheses described the influence of a selection factor on the likelihood of choosing a DBA over an EDW architecture. The third expressed the influence of the perceived ability of the IT staff. According to the hypothesis, organizations that have an IT staff with low perceived ability are more likely to implement a DBA architecture. Although the hypothesis describing this relationship was not significant, the mean of perceived ability for DBA was lower than the mean for EDW, suggesting the outcome described in the hypothesis.

The fourth hypothesis described the influence of resource availability. The hypothesis stated that when the resource availability is low, organizations are more likely to implement a DBA architecture. However, the findings suggest that the perceived availability of resources needed to implement an EDW or a DBA may not be very different. However, as described before, the mean difference of resource availability between the two was as expected, thus providing some initial support for the hypothesis.

8.2.3 Summary of conclusions on architecture selection

The overall results on data warehouse architecture selection indicate that various combinations of selection factors influence the selection of a particular data warehouse architecture when compared to another. Some variables are more important in influencing the selection decision than others. As expected, many factors that were based on rational OIPT theory emerged as being important to architecture selection. One of the social political factors, source of sponsorship, was also identified as impacting architecture selection by influencing the view of the warehouse. In addition to source of sponsorship, strategic view seems to influence the relationship between interdependence, task, and the selection of data warehouse architectures. These findings offer some insight to the understanding of the IT implementation process, specifically IT design.

In the past few years there have been several calls for academic research that examines the impact of turbulent, uncertain environments on different phases of the IT implementation process (Cooper et al. 1990; Karimi et al. 2004). In a recent study, Karimi et al (2004) stress the importance of identifying the influence of environmental uncertainty (i.e., rational factors) on the design of IT systems. The authors indicate that in order to design and build systems that meet organizational needs, designers must investigate the level of uncertainty surrounding the IT project. While they specifically examine task routineness and interdependence in their study, Karimi et al (2004) acknowledge that political factors and other contextual variables influence IT design, implementation, and use.

This study identified other factors that could influence IT design in addition to task routineness and interdependence. Furthermore, the findings reveal that while uncertainty plays a key role, rational factors alone are not the primary drivers of IT design choice. Both rational factors and political influences affect the manner in which the implementation of an IT solution is viewed. The view of the IT solution prior to implementation along with the resource constraints, staff ability and urgency to implement the solution helps determine the selection of the appropriate IT design. Therefore, this study indicates that a combination of factors based on information processing theories and social political theories together influence IT design.

The three models for the selection of one architecture when compared to another presented in this chapter suggest the development of an overall model for architecture selection. Such a model is proposed in the next chapter under suggestions for future research. Next, the findings related to architecture success are discussed.

8.3 Architecture success

The research question on architecture success required an examination of how successful the various data warehouse architectures are. Chapter seven reported the results of testing specific hypotheses and the exploratory hypothesis on architecture success (See Table 7.40). Prior to describing possible explanations for the unsupported hypotheses, the overall findings on architecture success are presented.

8.3.1 Overall findings on architecture success

The overall findings on architecture success can be described in terms of the end product system success and the development process success. First, the results regarding the influence of architecture on system product success are discussed.

8.3.1.1 System product success

System product success was measured in terms of the dimensions of information quality, system quality, and net benefits. Overall, the findings obtained from both MANOVA and SEM indicated that the system product success of the IDM architecture is significantly lower than the other two data warehouse architectures. This finding confirmed expert opinions stated during the interviews. It has become common wisdom in industry that the IDM architecture provides limited capabilities in terms of information quality, system quality, and net benefits. For instance, according to Armstrong (1996), the IDM architecture propagates the creation of data islands as no effort is made to establish data consistency across the independent marts. The author suggests that an IDM architecture does not provide a data warehouse solution that has the high integration and consistency that an EDW or DBA architecture provides. However, organizational information processing theory suggests a different perspective on considering the results.

If the information requirements for which the IDM architecture was implemented are satisfied by the information processing capacity provided by the resulting data warehouse solution, then the IDM architecture implementation can be considered an appropriate warehouse solution for that particular organizational circumstance. As such, the success of the different architecture implementations is dependent on whether the resulting solution meets the data integration requirements for which it was implemented (Goodhue et al. 1992). Nevertheless, the findings provide some measure of the degree of information quality, system quality, and net benefits that can be provided by an IDM architecture as compared to the other two architectures.

Additionally, the data indicated that the DBA and EDW architectures cannot be significantly differentiated in terms of any of the success measures. This finding is interesting because the proponents of the two architectures hail the superiority of their favored approach. For example, the proponents of the DBA architecture state that the DBA architecture leads to the creation of a data warehouse that provides greater decision-support capabilities. While the mean of individual impacts for DBA was higher than for EDW architecture, it was not statistically significant.

Furthermore, the results from the SEM analysis, the more powerful statistical technique when compared to MANOVA, indicated that DBA was not significantly different from EDW in terms of system quality or information quality. The reduced form equations from the output further suggested that the success of the two architectures cannot be distinguished in terms of individual impacts, organizational impacts, or any of the dimensions of information and system quality.

One explanation for the lack of difference between the success measures may be due to biases of the survey respondents. According to Pinsonneault et al. (1993), people that function in different roles have different experiences and perceptions of the technology and its impacts on the organization. The data warehouse manager or the person closely involved in the selection of a data warehouse architecture was the primary respondent of this survey. It is possible that as the person involved in the selection and the implementation of the data warehouse architecture, the respondent perceived or intentionally conveyed that the data warehouse implementation is successful, irrespective of reality. In order to check this possibility, at the end of the survey, each respondent was asked to forward another survey composed of the same success questions to the manager for end user tools and applications.[28] The results of t-tests between the two groups indicated that there was no significant difference between the responses from each group.

Perhaps an explanation for the results lies in the history of the evolution of the two architectures. Over the years, as demands made on the two architectures have changed, each

[28] In a recent empirical study on the factors affecting data warehousing success, the researchers chose “the managers of end-user computing or people responsible for an application that uses data from the warehouse” as the primary respondent to answer questions related to data warehouse success (Wixom et al. 2001).

architecture has evolved to become more alike in many ways. For instance, in early implementations, the EDW architecture did not include the implementation of dependent data marts. Its early designs provided guidelines to implement a centralized integrated data store using relational data structures. An EDW architecture was designed to hold atomic and summarized data to satisfy decision-making needs (Inmon 1992). Over the years, based on user requests for dimensional data structures to conduct queries more effectively, dependent data marts have been incorporated in the initial design. On the other hand, the early implementations of the DBA architecture included separate data staging areas for each consecutive data mart implementation. As such, the early design lacked a strategy to reduce possible data inconsistencies between data marts, resulting from non-conformance to the master suite of dimensions and facts. In recent years, its proponents have prescribed the creation of a common staging area so that data mart developers would follow the master suite of dimensions and facts (Mimno 2002b).

Consequently, it is possible that both EDW and DBA provide the implementation of a solution with minimal differences in terms of the success of the resulting data warehouse. In fact, perhaps these findings provide a reason as to why both architectures have survived the ongoing debate in the industry on data warehouse architectures. They may have survived because it is difficult to discern which architecture provides a more successful data warehouse. However, the findings revealed that the two differed in terms of the development process success. The findings related to development process success are presented next.

8.3.1.2 Development process success

The results from the MANOVA analysis indicated that the data warehouse architectures differed in terms of development process success. Specifically, the development process time and costs of the EDW architecture is significantly higher when compared to the other two. The results further showed that the DBA and IDM architectures are indistinguishable in terms of development process time and costs.

The significant difference in EDW in terms of initial cost to roll out a business process or subject area is consistent with evidence in the data warehouse literature (Joshi 1999). However, the high development process cost and time associated with the implementation of an EDW architecture may be due to the scope of the business domain for which the architecture was implemented. Generally, an IDM is implemented to satisfy the needs of a functional area whereas an EDW is implemented in a larger business domain such as the entire organization. The time and cost associated with requirements gathering, data acquisition, and implementation of a data warehouse solution is much higher in a larger domain such as the entire company than a functional area. Furthermore, the data storage capacity and the data warehousing technology used in a larger domain would be more costly than in a smaller domain. Thus, the findings may be due to the fact that the costs of designing and implementing a data integration solution could increase rapidly with the size and scope of the data integration effort (Goodhue et al. 1992). From this viewpoint, the higher development process time and cost of an EDW architecture implementation may be justifiable.

By and large, the data analyses on data warehouse architecture success suggest that the information quality, system quality, and net benefits resulting from the implementation of an IDM architecture is considerably lower than the success resulting from the other two architectures. In terms of development process success, the findings show that EDW architecture is more costly and time consuming to implement when compared to the DBA and IDM architectures.

8.3.2 Specific hypotheses on success

Two of the specific hypotheses on architecture success were not supported. They revealed that flexibility and scalability were not significantly different between EDW and DBA architectures. Flexibility and scalability are both dimensions of system quality.

To identify how *architecture* influences system quality dimensions, the focus in creating hypotheses was to examine how technical differences between the architectures, such as the differences in data structures, impacts architecture success. As described previously in the literature review, an EDW is considered to be more flexible and scalable than a DBA in terms of the characteristics of the data structures. However, the results indicated no significant difference between the two in terms of flexibility or scalability. In fact, the overall results indicate no differences between the two architectures in terms of the dimensions of system product success. As mentioned before, the evolution of these two architectures over the years has made them more similar than different. It is possible that over time it has become difficult to discern their differences.

8.4 Conclusion

This chapter described the overall findings from this study in terms of data warehouse architecture selection and success. Possible explanations for the results were presented and, in some instances, additional analyses that were conducted were described. The next chapter presents the conclusions, limitations, and contributions from this study.

CHAPTER 9 – CONCLUSION

9.1 Summary of research

Currently there is a debate in the industry as to which data warehouse architecture should be used to build a data warehouse and how successful the different architectures are. This study was undertaken with the hope of providing insight for this debate. Specifically, the study examined the influence of organizational factors on the selection of a data warehouse architecture and assessed how successful the different architectures are. Based on a literature review and a series of expert interviews, a research model was developed. Next, a survey instrument was constructed to collect data on organizational selection factors, architecture success, and the data warehouse architecture used.

The online survey was promoted to potential respondents by multiple organizations and individuals, resulting in a total usable sample size of 401. The sample size for each category was: (1) 235 responses for the EDW architecture, (2) 105 responses for the DBA architecture, (3) 44 responses for the IDM architecture, and (4) 17 responses for the FED architecture. The FED architecture was dropped from the analysis since its sample size was inadequate to meet minimum sample size requirements for planned statistical analyses.

The multinomial logistic regression technique was applied on the data set to test the research model on architecture selection. The results suggested that specific combinations of some organizational factors influence the selection of one data warehouse architecture when compared to another. To assess the success of data warehouse architectures, SEM and MANOVA were used.

The outcome of the analyses indicated that some data warehouse architectures differed in terms of system product success and development process success. Overall, the results of the study provide some evidence as to what organizational factors are important to the selection of each major data warehouse architecture and an assessment of each architecture in terms of several measures of success.

In the sections that follow, the research findings are summarized. Next, the academic and practitioner contributions are described. Finally, the study limitations are described and directions for future research are suggested.

9.2 Research findings

The findings from this study are presenting in two parts. First, the outcomes relevant to architecture selection are presented. Second, the outcomes related to architecture success are reviewed. Part of the data analyses for both architecture selection and success was more exploratory in nature while the other part involved testing specific hypotheses. Several findings were consistent with expectations. In situations where findings differed from expectations, the results were inspected and possible explanations were drawn based on theory, the data warehouse literature, expert interview data, and the statistics themselves. The research findings for architecture selection are presented next.

9.2.1 Architecture selection

One of the key findings on architecture selection is that all the selection factors, which were identified in the first phase, do not impact the selection of each data warehouse architecture. Some variables are more important than others in the selection of a particular architecture.

The view of the warehouse prior to implementation, resource availability, and the perceived ability of the IT staff emerged as the key factors that influence architecture selection. Furthermore, the view of the warehouse seems to mediate the influence of horizontal information interdependence, task routineness, and source of sponsorship on the selection of a data warehouse architecture. Urgency to implement a data warehouse surfaced as a factor that only influences the selection between an EDW and a DBA architecture. Unlike the other selection factors examined in this study, the influence of both compatibility with existing systems and expert influence on the selection of a particular architecture could not be directly examined. They were not measured in a manner that would support such an analysis. However, the outcomes obtained from a separate analysis indicated that data warehousing literature makes a difference in the selection decision. Of the remaining eight, different combinations of factors influenced the choice of one architecture when compared to another.

The multinomial logistic regression analyses revealed that when considering the choice of architecture between EDW and IDM, the view of the warehouse, resource availability, and perceived ability of the IT staff influence the selection decision. Interdependence, task routineness, and source of sponsorship influence the choice of architecture by affecting the view of the warehouse. Except for the source of sponsorship, the same factors were identified as being important to the selection of DBA when compared to the IDM architecture. Finally, urgency and the view of the warehouse emerged as being important when deciding between an EDW and a DBA architecture. The source of sponsorship also influenced the choice between EDW and DBA by affecting the view of the warehouse. The combinations of factors influencing the selection of one architecture when compared to another are presented in Figures 8.1, 8.2, and 8.3.

9.2.2 Architecture success

Architecture success was captured in terms of system product success and development process success. System product success was measured in terms of information quality, system quality, and net benefits, while development process success captured the time and cost to implement an architecture. The specific success factors used to measure the dimensions of both types of success are presented in the research model for architecture success in Figure 4.6.

Both MANOVA and SEM were used to assess the system product success of the architectures. The outcomes obtained suggested that the system quality, information quality, and net benefits attained from an IDM architecture were considerably lower than the other two architectures. Both statistical techniques found no difference between the EDW and DBA architectures in terms of any of the factors measuring system product success. The research suggests that the data warehouse solution (i.e., system product success) resulting from a EDW or DBA architecture maybe similar.

An examination of development process success using MANOVA revealed that EDW architecture is more expensive and time consuming to build than the rest. No differences in development process time or costs were found between the IDM and DBA architectures.

9.3 Research contributions

The contributions from this research are described in terms of measurement contribution, contribution to academia, and contribution to practice.

9.3.1 Measurement contribution

The items used to measure the selection factors and success measures were developed from expert interviews, existing items in the IS literature, and the data warehousing literature. These questions were refined through an iterative process of review and feedback from 20 experts in the field, the literature, and academic theory. The final set of questions was crafted using industry terms and phrases to reflect terminology familiar to the respondents.

The measurement models for the variables of interest displayed good fit and possessed goodness-of-fit statistics well above the recommended thresholds for model fit. The reliability of the variables exceeded the requirement for exploratory research. In fact, the success measures exhibited reliabilities well above that suggested for confirmatory studies.

These validated measures, developed through an extensive process of review from field experts to reflect the true nature of variables in the real world, should be useful to other IS researchers conducting studies in data warehousing, as well as other areas of IS.

9.3.2 Contribution to academia

This dissertation makes several contributions to the academic literature. First, it extends the existing knowledge on IT infrastructure decision making by identifying the importance and the influence of organizational factors on data warehouse architecture selection. Broadbent et al (1997) describe how organizations choose IT infrastructure options based on their view of IT infrastructure, which is founded on business needs. This research provides some empirical support to these authors' findings that an organization's view of infrastructure influences the infrastructure capability attained. Further, this work applies Weill et al. (1998) findings from multiple case studies on IT infrastructure investment decisions to the current context and provides empirical support to confirm that business needs stemming from interdependence, task routineness, as well as the source of sponsorship influences the view of an IT infrastructure. Thus, this work enriches the existing sparse research on IT infrastructure.

It also furthers the existing literature by identifying other organizational factors that may be important to and influence the choice of IT infrastructure. With respect to the existing IS literature on data management in general, and data integration in particular, the study provides empirical support for the importance of many of the organizational factors identified as affecting data management practices and choice of integration in organizations.

There is limited previous research on IT infrastructure design. This is partly due to the difficulty in defining and measuring IT infrastructure design. This research contributes to IT infrastructure design research in several ways. The study illustrates that data warehouse architecture offers an opportunity to examine issues relevant to IT infrastructure design. It also identifies measurable constructs that can be used in future IT infrastructure design research. The study suggests that many of the factors that influence IT application design influences IT infrastructure design. Additionally, it offers empirical support for new variables, such as view of the data warehouse, that influence IT infrastructure design.

This research also provides support for organizational information processing theory by recognizing the importance of information processing requirements to the selection of a data warehouse architecture. The findings show that the amount of uncertainty in the organization, based on the extent of interdependence and task non-routineness, influences the choice of information processing capacity provided by the various data warehouse architectures.

Often when assessing the success of IS, such as investigating the success of enterprise systems, researchers have focused on the success of the end product system. This research introduces the importance of evaluating the development process success as a measure of the overall IS success. Based on a theoretical foundation and expert interviews, the study developed a model to assess data warehouse architecture success, which incorporated both system product success and development process success. It provides a more comprehensive picture of IS success.

Finally, this study addressed a general need for more behavioral research in data warehousing. Only a few empirical studies have focused on behavioral research in data warehousing. A valuable aspect of this study is the incorporation of existing theory and literature with expert interview findings to develop factors specific to data warehouse architecture selection and success. Both the IS infrastructure design and the data warehousing research areas should benefit from the insights gained through this dissertation. It hopefully provides further guidance for new research in data warehousing and infrastructure design.

9.3.3 Contribution to practice

This dissertation was primarily motivated by the need to shed light on two burning questions in the data warehouse industry, "what architecture should be used to implement a data warehouse architecture and how successful are the different architectures?" Companies are currently spending millions of dollars implementing data warehouse architectures without empirical evidence about the architecture alternatives. The research findings provide some insight with regard to both questions.

First, the findings identified the most salient factors that appear to influence the selection of each data warehouse architecture when compared to another. The factors provide practitioners with a basis to evaluate the current organizational circumstances within their organizations. They allow an organization to better understand or even control their data warehouse architecture selection decision by making them more aware of how their organizational situation drives choice. It also gives practitioners the opportunity to manipulate their organizational setting to implement an architecture they desire. For instance, when faced with an in-house IT staff with low perceived ability, the organization might send the development staff to training programs to raise their actual and perceived ability. Thereby the organization facilitates the selection of an EDW or DBA architecture when compared to an IDM architecture. As such, the findings on architecture selection can provide practitioners with some guidance on how to approach the data warehouse architecture selection decision.

Next, the findings also provide some evidence of the success of each architecture. The proponents of the two major data warehouse architectures (i.e., EDW and DBA) make various claims about the superiority of each architecture when compared to the other. This study provides an assessment of architecture success in terms of the different dimensions of the resulting system product, and in terms of the development process.

The results indicated that there is no significant difference between EDW and DBA architecture in terms of system product success. The findings suggest that there might not be one superior data warehouse architecture, and that both can lead to the implementation of a successful data warehouse.

9.4 Limitations

As with any research, there were a number of limitations in this study. One of the limitations was the inability to identify the influence of the organizational factors on the selection of a federated architecture and to evaluate the success of this architecture. Due to the small number of observations collected for the FED architecture, it was omitted from the analysis. However, according to past surveys describing the distribution of data warehouse architectures in industry, the FED architecture appears to be infrequently used.

When assessing data warehouse architecture success, the current study did not capture how organizational circumstances can influence the development process time and costs, as well as the success of the end product system. For example, a development team with low skills might develop a poor quality data warehouse. As another example, although it provides considerably lower integration than the other architectures, it is possible that the selection of an IDM architecture can meet the acceptable costs and benefits requirements of a company. This point is discussed further in the future research section.

From a methodological standpoint, there were several limitations in the study. This research was conducted using a cross-sectional field survey. Respondents that currently had a data warehouse in place were asked to recall the nature of organizational factors at the time of implementation. A longitudinal survey would have enabled the collection of information on architecture selection at the time the selection decision was made, followed by the assessment of its success at a later time. Alternatively, organizations currently involved in selecting a data warehouse architecture could have been the target respondent for questions related to architecture selection. Unfortunately, longitudinal studies or conducting two cross-sectional surveys demand a higher investment in time and resources than a single cross-sectional design. As such, a cross-sectional survey was conducted.

In addition, the study utilized perceptual measures to investigate architecture selection and success. This is a common practice in IS research, but a respondent's perceptions of phenomena are not completely accurate reflections of reality. However, collecting objective responses to measure desired factors would be too costly to investigate the phenomenon under study. And there may not be objective measures for the variables of interests. An attempt was made to operationalize variables with multiple items so as to gain confidence in the reliability and validity of the variables.

The data collection for this study was conducted using purposive sample selection. While generalizability is potentially a key benefit of survey research, the results are only generalizable to the population that the sample represents. The sample used in this study was not randomly selected. As such, its representativeness of the overall data warehousing population is uncertain. However, this sample was drawn from multiple sources and every effort was made to gain participation from a diverse set of organizations. As mentioned previously, recent industry surveys confirmed that the distribution of data warehouse architectures in this study was similar to that found in industry. Furthermore, the examination of the demographic information of the organizations that responded to the survey revealed that they represent a wide variety of industries and come from a wide variety of countries and throughout the United States. This demographic information further confirmed the representativeness of the data set.

9.5 Future research

The study suggests a number of avenues for future research. Here, three possibilities are presented. First, the selection factors that emerged as salient in influencing the selection of one architecture versus another suggests the creation of one overall model for architecture selection. Figures 8.1, 8.2, and 8.3 graphically present the common themes of selection factors that emerge across the selection of one architecture versus another. The overall model can be described as follows.

First, based on OIPT, both interdependence and task routineness can be combined to represent the information processing needs that influence the selection of a data warehouse (Tushman et al. 1978). Along with the source of sponsorship, information processing needs influence the creation of a view of the warehouse as a strategic infrastructure initiative. Next, resource availability, perceived ability of the IT staff, and urgency can be described as the facilitating conditions[29] within an organization that influences architecture selection. Together these factors suggest the possibility of developing a more parsimonious causal model for the selection of a particular architecture based on information processing requirements, source of sponsorship, view of the warehouse, and facilitating conditions. This new model is graphically presented in Figure 9.1, with the more parsimonious model presented in Figure 9.2. One possible opportunity for research is to empirically test this parsimonious model for architecture selection. It may also be useful to conduct a series of case studies to enrich this model.

The current research does not help understand what architecture is best for a given organizational circumstance. It identifies the salient organizational factors that lead to the selection of a data warehouse architecture but stops short of assessing if it is the best fit for the organization. Furthermore, when considering architecture success, the findings do not help identify how the success attained relates to organizational requirements for which it was implemented. For instance, while the scalability of an IDM architecture implementation might be low, it maybe sufficient to meet the organizational needs for which the architecture was implemented. Testing the full contingency model for architecture fit would provide a more comprehensive understanding of the

[29] In individual level theory, Triandis (1979) describes facilitating conditions as the conditions of the environment which facilitate a behavior. It is the perception of the ease or difficulty of performing a behavior based on the environment. In a similar manner, in the context of this study, resource availability, the perceived ability of the IT staff, and urgency creates a perception of the ease or difficulty of implementing a data warehouse infrastructure initiative. For instance, high resource availability, high perceived ability, and low urgency would results in the total assessment of facilitating conditions as assisting the implementation of a data warehouse infrastructure initiative.

architecture best suited for an organization. Additionally, it would help provide a richer understanding of architecture success by assessing if the architecture met the requirements for which it was implemented.

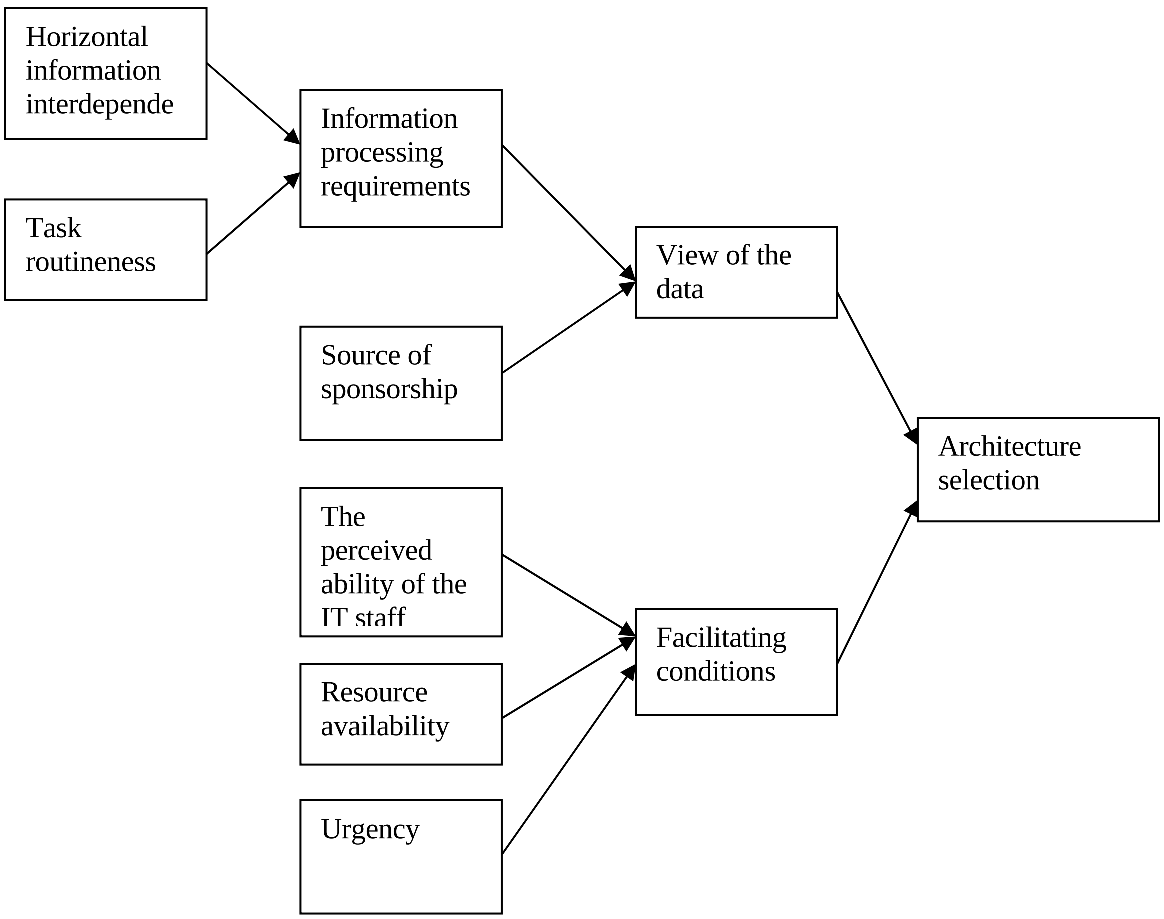


Figure 9.1: A suggested model for data warehouse architecture selection

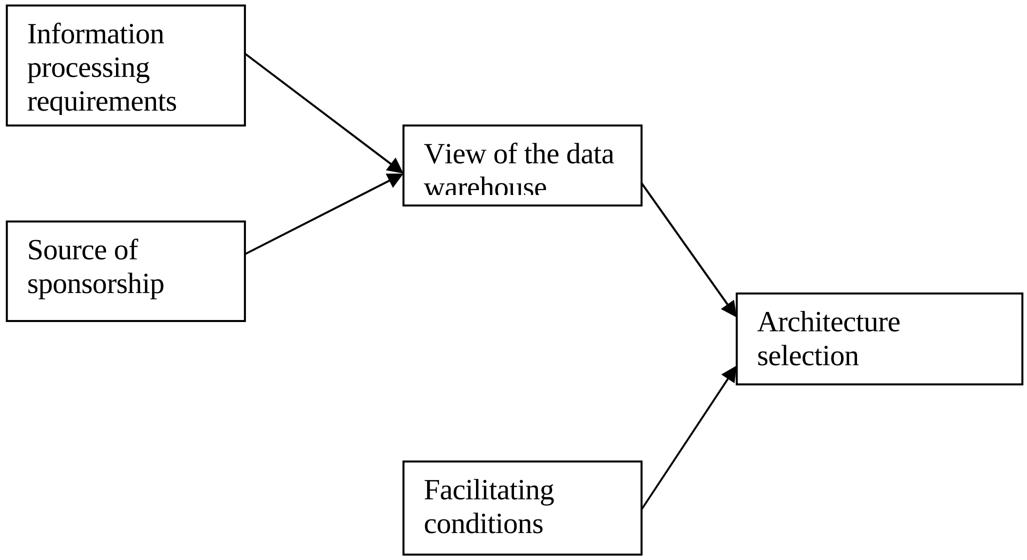


Figure 9.2: A parsimonious model for data warehouse architecture selection

Finally, domain of the business was identified as an additional variable that influences architecture success. It is not clear how domain influences other selection factors, as well as architecture selection. For instance, the domain or scope of the data warehouse architecture implementation may impact the source of sponsorship for a data warehousing initiative. It seems that domain may moderate the relationship between certain selection variables and architecture selection. Further research is needed to understand how domain influences architecture selection.

9.6 Conclusion

This chapter first presented a summary of the research conducted in this dissertation. It highlighted the findings in terms of architecture selection and success. Next, it described the contributions of this research to both academia and practice. Finally, the chapter concluded by presenting the limitations, as well as the future research opportunities of this study.

REFERENCES

Abrahamson, E. "Managerial fads and fashions: the diffusion and rejection of innovations," *The academy of management review* (16:3), July 9 1991, pp 586-612.

Agarwal, R., Sambamurthy, V., and Stair, R.M. "The Evolving Relationship between General and Specific Computer Self-Efficacy: An Empirical Assessment Decision and Information Technologies," *Information Systems Research* (11:4), December 2000, pp 418-430.

Agosta, L. "Data warehouse volume growth continues," RIB-062001-00035, Giga Information Group, p. 2.

Allen, B.R.B., Andrew C. "Information Architecture: In Search of Efficient Flexibility," *MIS Quarterly* (15:4), Dec., 1991, pp 435-445.

Andres, H.P., and Zmud, R.W. "A contingency approach to software project coordination," *Journal of Management Information Systems* (18:3), Winter 2001-2002, pp 41-70.

Armstrong, R. "Data warehousing: the fallacy of data mart centric strategies," NCR, Denver, p. 9.

Arnold, S.E. "Information manufacturing: the road to database quality," *Database* (15:5) 1992, pp 32.

Avison, D. "The search for the discipline of information systems," in: *Understanding social research: perspectives on methodology and practice*, G. McKenzie, J. Powell and R. Usher (eds.), Falmer Press, London, 1997, pp. 87-100.

Babbie, E. *Survey research methods* Wadsworth Pub, Belmont, 1990.

Babbie, E.R. *The practice of social research*, (9 ed.) Wadsworth/Thompson learning, Belmont, CA, 2000, p. 116.

Badami, V. "Sarbanes-Oxley: The Role of Business Intelligence and Data Management," The Data Warehousing Institute, 2003.

Bajaj, A. "A study of senior information systems managers' decision models in adopting new computing architectures," *Journal of the Association for Information Systems* (1), June 2000, pp 1-56.

Bajwa, D.S., Rai, A., and Ramaprasad, A. "The structural context of executive information systems adoption," *Information Resources Management Journal* (11:3), Summer 1998, pp 28-39.

Bandura, A. "Self-efficacy: Toward a unifying theory of behavioral change.," *Psychological Review* (84) 1977, pp 191-215.

Bandura, A., and Wood, R. "Effect of Perceived Controllability and Performance Standards on Self-Regulation of Complex Decision Making," *Journal of Personality and Social Psychology*. (56:5), May 1989, pp 805-814.

Barbusinski, L., Bischoff, J., and Howard, S. "In using a true DW design, if data integration techniques and standards are created and followed, are there other things that make it difficult to add additional data design (tables, fields, etc.) in later stages?" in: *DM Review Online*, Thomson Media, 2001.

Baron, R.M., and Kenny, D. "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration.," *Journal of Personality and Social Psychology* (51) 1986, pp 1173-1182.

Begg, C.B., and Gray, R. "Calculation of polychotomous logistic regression parameters using individualized regressions," *Biometrika*:71) 1984, pp 11-18.

Benbasat, I., and Taylor, R.N. "The Impact of Cognitive Styles on Information System Design," *MIS Quarterly* (2:2), Jun., 1978, pp 43-54.

Benbasat, I., and Weber, R. "Research Commentary: Rethinking "Diversity" in Information Systems Research.," *Information Systems Research* (7:4), December 1996, pp 389-400.

Benbasat, I., and Zmud, R.W. "Empirical Research in Information Systems: The Practice of Relevance," *MIS Quarterly* (23:1), Mar99 1999, pp 3-17.

Bhend, M. "Make Your Data Warehouse a Real Success: Start With a Scalable Data Warehouse Pilot Project," in: *DM Review*, 1999, pp. 34-37.

Bloomfield, B.P., and Best, A. "Management consultants: Systems development, power and the translation of problems.," *Sociological Review* (40:3), Aug92 1992, pp 533-561.

Bloomfield, B.P., and Danieli, A. "The role of management consultants in the development of information technology: The indissoluble.," *Journal of Management Studies* (32:1), Jan95 1995, pp 23-47.

Bollen, K.A. *Structural Equations with Latent Variables* Wiley, New York, 1989.

Bontempo, C., and Zagelow, G. "The IBM data warehouse architecture," *Association for Computing Machinery. Communications of the ACM* (41:9), Sep 1998, pp 38.

Boudreau, M.-C., Gefen, D., and Straub, D.W. "Validation in information systems research," *MIS Quarterly* (25:1), March 2001, pp 1-14.

Bowman, C.M., Danzig, P.B., Manber, U., and Schwartz, M.F. "Scalable Internet resource discovery: Research problems and approaches," *Communications of the ACM* (37:8), August 1994, pp 98-109.

Bradshaw-Camball, P., and Murray, V.V. "Illusions and Other Games: A Trifocal View of Organizational Politics," *Organization Science* (2:4), Nov1991 1991, pp 379-398.

Brancheau, J.C., Janz, B.D., and Wetherbe, J.C. "Key issues in information systems management: 1994-95 SIM delphi results," *MIS Quarterly* (20:2), Jun 1996, pp 225-242.

Broadbent, M., and Weill, P. "Management by maxim: how business and IT managers can create IT infrastructures," *Sloan Management Review* (38:3), Spring 1997, pp 77-93.

Brohman, M.K. "Explaining variation in data warehouse usage: an interpretation perspective," in: *Richard Ivey School of Business*, The University of Western Ontario, London, Ontario, 2000, p. 320.

Byrd, T.A., and Turner, D.E. "An exploratory analysis of the value of the skills of IT personnel: Their relationship to IS infrastructure and competitive advantage," *Decision Sciences* (32:1), Winter 2001, pp 21-55.

Byrd, T.A.T., Douglas E "Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a Construct.," *Journal of Management Information Systems* (17:1), Summer2000 2000, pp 167-209.

Calvanese, D.d.G., Giuseppe; Lenzerini, Maurizio; Nardi, Daniele; Rosati, Riccardo "Data Integration in Data Warehousing.," *International Journal of Cooperative Information Systems* (10:3), Sep 2001, pp 237-272.

Cavaye, A., and Christiansen, J. "Understanding IS implementation by estimating power of subunits," *European Journal of Information Systems* (5:4), Dec 1996, pp 222-232.

Chatterjee, D., Pacini, C., and Sambamurthy, V. "The Shareholder-Wealth and Trading-Volume Effects of Information-Technology Infrastructure Investments," *Journal of Management Information Systems* (19:2) 2002, pp 7-42.

Chau, P.Y.K., and Tam, K.Y. "Factors affecting the adoption of open systems: An exploratory study," *MIS Quarterly* (21:1), March 1997, pp 1-24.

Chein, I. *Appendix: An introduction to sampling* Holt, Rinehart and Winston, Austin, 1981.

Chen, H.-M., and Sheldon, P.J. "Destination information systems: Design issues and directions," *Journal of Management Information Systems* (14:2), Fall 1997, pp 151.

Chen, L., and Gillenson, M. "A framework for data warehousing research," Americas Conference on Information Systems, Boston, 2000.

Chwelos, P., Benbasat, I., and Dexter, A.S. "Research Report: Empirical Test of an EDI Adoption Model," *information Systems Research* (12:3), Sep 2001, pp 304-321.

Ciol "Enterprise special: data warehousing," http://www.ciol.com/, 2002.

Cohen, G. "The unquenchable thirst for information," in: *InformationWeek*, 2003, p. 17.

Cohen, J. *Statistical power analysis for the behavioral sciences* Erlbaum, Hillsdale, 1988.

Comrey, A.L., and Lee, H.B. *A first course in factor analysis* Erlbaum, New Jersey, 1992.

Conte, J.M., Landy, F.J., and Mathieu, J.E. "Time Urgency: Conceptual and Construct Development," *Journal of Applied Psychology* (80:1) 1995, pp 178-185.

Cook, C., Heath, F., and Thompson, R. "A meta analysis of response rates in Web or Internet based surveys," *Educational and psychological measurement* (60:6), December 2000, pp 821-836.

Cooper, B.L., Watson, H.J., Wixom, B.H., and Goodhue, D.L. "Data warehousing supports corporate strategy at First American Corporation," *MIS Quarterly* (24:4), Dec 2000, pp 547-567.

Couper, M. "A review of issues and approaches," *Public opinion quarterly* (64:4) 2000, pp 464-494.

Damanpour, F. "Organizational innovation: a meta analysis of effects of determinants and moderators," *Academy of Management Journal* (34:3) 1991, pp 555-590.

Darrow, B. "Making the right choice: solution providers are evaluating a plethora of options as they puzzle over the future of business intelligence," in: *Computer reseller news*, 2003, p. 16.

Delone, W.H., and Mclean, E.R. "Information systems success: quest for the dependent variable," *Information Systems Research* (3:1) 1992, pp 60-95.

Delone, W.H., and McLean, E.R. "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems* (19:4), 2003/04/01 2003, pp 9-30.

Delone, W.H., and Mclean., E.R. "Information systems success revisited," HICSS, Hawaii, 2002.

Devlin, B.A. *Data warehouse: from architecture to implementation* Addison Wesley, Reading, Massachusetts, 1997.

Dillman, D., Tortora, R., and Bowker, D. "Principles for constructing Web surveys," SESRC, Pullman, Washington, p. 16.

Dobbs, T., Stone, M., and Abbott, J. "UK data warehousing and business intelligence implementation," *Qualitative Market Research* (5:4) 2002, pp 235.

Doll, W.J. "Avenues for Top Management Involvement in Successful MIS Development (in Theory and Research)," *MIS Quarterly* (9:1), March 1985, pp 17-35.

Drury, D.H., and Farhoomand, A.F. "A hierarchical structural model of information systems success," *INFOR* (36:1/2), Feb/May 1998, pp 25-40.

Duncan, N.B. "Capturing flexibility of information technology infrastructure: A study of resource characteristics and their measure," *Journal of Management Information Systems* (12:2), Fall 1995, pp 37.

Easton, E. "Reporting Companies Beware: SOX 409 Will Knock Your Socks Off," *DM Review*:2004, March 2004, pp 16-18.

Eckerson, W. "Four ways to build a data warehouse," in: *ADT*, 2002.

Eckerson, W. "The Rise of Analytic Applications: Build or Buy," The data warehousing institute, Seattle, p. 47.

Eckerson, W. "Smart Companies in the 21st Century: The Secrets to Creating Successful Business Intelligence Solutions," The data warehousing institute, Seattle, p. 53.

Edelman, F. "The Management of Information Resources - A Challenge for American Business," *MIS Quarterly* (5:1), Mar., 1981, pp 17-27.

Ein-Dor, P., and Segev, E. "Organizational context and the success of management information systems," *Management Science* (24) 1978, pp 1067-1077.

Ein-Dor, P., and Segev, E. "Organizational Context and MIS Structure: Some Empirical Evidence," *MIS Quarterly* (6:3), Sep 1982, pp 55-69.

Eisenhardt, K.M., and Bourgeois, L.J. "Politics of Strategic Decision-Making in High-Velocity Environments - toward a Midrange Theory," *Academy of Management Journal* (31:4), Dec 1988, pp 737-770.

Fan, M., Stallaert, J., and Whinston, A.B. "The adoption and design methodologies of component-based enterprise systems," *European Journal of Information Systems* (9:1), Mar 2000, pp 25.

Fedorowicz, J., and Konsynski, B. "Organization support systems: Bridging business and decision processes," *Journal of Management Information Systems* (8:4), Spring 1992, pp 5.

Fichman, R.G., and Kemerer, C.F. "The assimilation of software process innovations: An organizational learning perspective.," *Management Science* (43:10), Oct 1997, pp 1345-1364.

Finkelstein, C. "The Enterprise: Building Successful Data Warehouses," in: *DM Review*, 1999.

Francalanci, C., and Piuri, V. "Designing information technology architectures: a cost-oriented methodology," *Journal of Information Technology* (14:4), Dec1999 1999, pp 181-192.

Galbraith, J. *Designing complex organizations* Addison-Wesley, Reading, 1973.

Galbraith, J. "Organizational design: an information processing view," *Interface* (4:3) 1974, pp 87-89.

Gardner, S.R. "Building the data warehouses," *Communications of the ACM* (41:9) 1998, pp 51-60.

Gattiker, T. "The Costs and Benefits of Enterprise Resource Planning Systems in Operations Management: A Model and Empirical Analysis," in: *Management*, University of Georgia, Athens, 2000, p. 210.

Gattiker, T., and Goodhue, D.L. "Understanding the local level costs and benefits of ERP through organizational information processing theory," *Information & Management* (Forth coming) 2003.

Gefen, D., Straub, D.W., and Boudreau, M.-C. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the association of information systems* (4:7), October 2000.

Gentry, J. "Methodology vs. architecture: the myth of the "big bang" data warehouse," TDWI Flashpoint, 2002.

Gibson, C.B. "Do They Do What They Believe They Can? Group Efficacy and Group Effectiveness Across Tasks and Cultures.," *Academy of Management Journal* (42:2) 1999, pp 138-153.

Goodhue, D., Wixom, B., and Watson, H. "Realizing Business Benefits Through CRM: Hitting the Right Target in the Right Way," *MIS Quarterly Executive* (1:2) 2002, pp 79-94.

Goodhue, D.L. "Understanding user evaluations of information systems," *Management Science* (41:12), Dec 1995, pp 1827-1844.

Goodhue, D.L. "Development and measurement validity of a task-technology fit instrument for user evaluations of information systems," *Decision Sciences* (29:1), Winter 1998, pp 105-139.

Goodhue, D.L., Quillard, J.A., and Rockart, J.F. "Managing the Data Resource - a Contingency Perspective," *MIS Quarterly* (12:3), Sep 1988, pp 373-392.

Goodhue, D.L., and Thompson, R.L. "Task-Technology Fit and Individual-Performance," *Mis Quarterly* (19:2), Jun 1995, pp 213-236.

Goodhue, D.L., Wybo, M.D., and Kirsch, L.J. "The Impact of Data Integration on the Costs and Benefits of Information-Systems," *MIS Quarterly* (16:3), Sep 1992, pp 293-311.

Gray, P., and Watson, H.J. *Decision support in the data warehouse* Prentice hall, New Jersey, 1998.

Greenley, G., and Oktemgil, M. "A comparison of slack resources in high and low performing British companies," *Journal of Management Studies* (35:3), May 1998, pp 377-398.

Griffin, J. "Information Strategy: Data Mart vs. Data Warehouse," *DM Review*: February 1998), Feb 1998.

Grover, V., Lee, C.C., Durand, D., and Durand, D. "Analyzing methodological rigor of MIS survey research from 1980-1989," *Information & Management* (24:6), 1993, pp 305-317.

Grover, V., and Saeed, K.A. "Strategic orientation and performance of internet-based businesses," *Information Systems Journal* (14:1), January 2004, pp 23-41.

Grover, V., and Segars, A.H. "The relationship between organizational characteristics and information system structure: An international survey," *International Journal of Information Management* (16:1), 1996, pp 9-25.

Gully, S.M., Incalcaterra, K.A., Joshi, A., and Beaubien, J.M. "A Meta-Analysis of Team-Efficacy, Potency, and Performance: Interdependence and Level of Analysis as Moderators of Observed Relationships," *Journal of Applied Psychology* (87:5) 2002, pp 819-832.

Hackney, D. "What is a data mart?" in: *DM Review*, 1997, pp. 22-24.

Hackney, D. "Architecture of architectures: data warehouse delivery," in: *DM Review*, 1998.

Hackney, D. "Architecture anarchy and how to survive it: God save the queen," *Enterprise Systems Journal* (15:4) 2000, pp 24-30.

Hackney, D. "BI architecture tiers," in: *Data Management Review*, 2002, pp. 12-18.

Hackney, D. "Data warehouse delivery architecture of architectures," in: *DM Review*, 2002, pp. 22-24.

Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. *Multivariate data analysis* Prentice Hall, New Jersey, 1995.

Haley, B.J. "Implementing the Decision Support Infrastructure: Key Success Factors in Data Warehousing," in: *Management*, University of Georgia, Athens, 1998, p. 190.

Heinecke, G. "Web-based data warehousing on S/390 at the Landesbank Rheinland-Pfalz," in: *IBM Bulletin*, 2000, pp. 38-39.

Hinkin, T.R. "A review of scale development practices in the study of organizations," *Journal of Management* (21:5) 1995, pp 967-988.

Hosmer, D.W., and Lemeshow, S. *Applied logistic regression* Wiley, New York, 2000.

Hoss, D. "The Bottom Line Looms, but Innovation Persists," *DM Review*, 2002.

Hsieh, F.Y., Bloch, D.A., and Larsen, M.D. "A simple method of sample size calculation for linear and logistic regression," *Statistics in Medicine* (17) 1998, pp 1623-1634.

Huff, S.L., and Munro, M.C. "Information Technology Assessment and Adoption: A Field Study," *MIS Quarterly* (9:4), Dec 1985, pp 327.

Humphries, M. *Data warehousing: architecture and implementation*, 1999.

Hwang, M., and Cappel, J.J. "Data warehouse development and management: Practices of some large companies," *Journal of Computer Information Systems* (43:1), Fall 2002, pp 3-6.

Iacovou, C.L., Benbasat, I., and Dexter, A.S. "Electronic Data Interchange and Small Organizations: Adoption and Impact of Technology," *MIS Quarterly* (19:4) 1995, pp 465-485.

Inmon, W., Imhoff, C., and Sousa, R. *Corporate Information Factory*, (Second Edition ed.) Wiley & sons, New York, 2001.

Inmon, W.H. *Building the data warehouse* Wiley & Sons, New York, 1992.

Inmon, W.H., Zachman, J.A., and Geiger, J.G. *Data Stores, Data Warehousing, and the Zachman Framework: Managing Enterprise Knowledge* McGraw-Hill, Fairfield, 1997, p. 359.

Ives, B., Hamilton, S., and Davis, G.B. "A Framework for Research in Computer-Based Management Information-Systems," *Management Science* (26:9) 1980, pp 910-934.

Jasperson, S., Carte, T., Saunders, C., Butler, B., Croes, H., Zheng, W., “Review: Power and Information Technology Research: A Metatriangulation Review,” MIS Quarterly, (26:4) Dec2002, pp397-460.

Jeusfeld, M.A. "Special Issue on Design and Management of Data Warehouses.," *International Journal of Cooperative Information Systems* (10:3), Sep 2001 2001, pp 235-237.

Jiang, J.J., Klein, G., and Carr, C.L. "Measuring information systems service quality: SERVQUAL from the other side.," *MIS Quarterly* (26:2) 2002, pp 145-166.

Johnston, H.R., and Carrico, S.R. "Developing Capabilities to Use Information Strategically," *MIS Quarterly* (12:1), Mar1988 1988, pp 37-48.

Joshi, K., and Curtis, M. "Issues in building a successful data warehouse," *Information Strategy: The executive's journal*: Winter) 1999, pp 28-35.

Karimi, J., Somers, T., Gupta, Y. "Impact of Environmental Uncertainty and Task Characteristics on User Satisfaction with Data," *Information Systems Research* (15:2), June 2004, pp 175-193.

Keil, M. "Pulling the Plug: Software Project Management and the Problem of Project Escalation (in Application)," *MIS Quarterly* (19:4), December 1995, pp 421-447.

Kelly, S. "Data marts: The latest silver bullet.," in: *Data mart review*, 1997, pp. 12-16.

Kerby, J.K. "Factor Analysis: Key Tool in Consumer Behavior Research," *Academy of Marketing Science Journal* (1979:7), 374-388 1979, pp 4.

Kerlinger, F.N. *Foundations of behavioral research* Holt, Rinehart and Winston, New York, 1986.

Kimball, R. "A practical method for planning a data warehouse," in: *Intelligent Enterprise*, 1999.

Kimball, R., Reeves, L., Ross, M., and Thornthwaite, W. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses* Wiley, New York, 2003.

King, J.L. "Centralized versus decentralized computing: organizational considerations and management options," *ACM Computing Surveys* (15) 1983, pp 319-350.

Kleinbaum, D.G. *Logistic regression: a self learning text* Springer-Verlag, New York, 1994.

Koch, C. "The Middle Ground," *CIO*: Jan 15) 1999, pp 1-12.

Landy, F.J., Rastegary, H., Thayer, J., and Colvin, C. "Time Urgency: The Construct and its Measurement," *Journal of Applied Psychology* (76:5) 1991, pp 644-701.

Laney, D., "Data warehouse factors to address for success," *HP Professional,* (14:5) 2000, pp 21-22.

Langley, A., Mintzberg, H., Pitcher, P., Posada, E., and Saintmacary, J. "Opening up decision making: the view from the black stool," *Organization science* (6:3), MAY-JUN 1995 1995, pp 260-279.

Lauter, J. "Sample size requirements for the Test of MANOVA," *Biometrical Journal* (20) 1978, pp 389-406.

Lee, K., and Koval, J.J. "Determination of the best significance level in forward stepwise logistic regression," *Communication in Statistics* (B:26) 1997, pp 559-575.

Lesar, C. "Adopting Technology for Compliance," in: *DM Review*, 2004, pp. 23-24.

Lincoln, Y.S., and Guba, E.G. *Naturalistic inquiry* Sage, Thousand oaks, 1985.

Lindsley, D.H., and Brass, D.J. "Efficacy-performance spirals: A multilevel perspective.," *Academy of Management Review* (20:3), July 1995, pp 645-679.

Lloyd, P.T.L., and Galambos, G.M. "Technical reference architectures," *IBM Systems Journal* (38:1) 1999, pp 51-76.

Ma, C., Chou, D., and Yen, D. "Data warehousing, technology assessment and management," *Industrial management and data systems* (100:3) 2000, pp 125-134.

Maier, M., Emery, D., and Hilliard, R. "Introducing IEEE 1471," *IEEE Computer* (34:4), April 2001 2001, pp 9-14.

Maier, M., and Rechtin, E. *The art of systems architecting*, (2nd Edition ed.) CRC Press, Boca Raton, 2000, p. 313.

Maierhofer, N.I., Griffin, M.A., and Sheehan, M. "Linking Manager Values and Behavior With Employee Values and Behavior: A Study of Values and Safety in the Hairdressing Industry," *Journal of Occupational Health Psychology* (5:4), Oct 2000.

Manning, I.T. "Data warehousing," http://www.manning.demon.co.uk/, 2000.

Marakas, G.M., Yi, M.Y., and Johnson, R.D. "The Multilevel and Multifaceted Character of Computer Self-Efficacy: Toward Clarification of the Construct and an Integrative Framework for Research," *Information Systems Research* (9:2), June 1998, pp 126-163.

Marco, D. "Meta Data & Knowledge Management: Independent Data Marts: Stranded on Islands of Data, Part 2," in: *DM Review*, 2003.

Markus, M.L. "Power, politics and MIS implementation," *Communications of the ACM* (26:6) 1983, pp 430-444.

Matheson, C. "Rationality and Decision-making in Australian Federal Government," *Australian Journal of Political Science* (33:1) 1998, pp 57-72.

McCarthy, R., Aronson, J., and Claffey, G. "Task technology fit in data warehousing environments: analyzing the factors that affect utilization," AMCIS, 2000.

McFarlan, F.W., and McKenney., J.L. *Corporate Information Systems Management: The Issues Facing Senior Executives* Homewood, Irwin, 1983.

Merriam, S.B. *Qualitative research and case study applications in education* Jossey-Bass Publisher, San francisco, 1998.

Might, R.J., and Fisher, W.A. "The role of structural factors in determining project management success.," *IEEE Transactions on Engineering Management* (32:2) 1985, pp 22-40.

Miller, D., and Friesen, P.H. "Innovation in Conservative and Entrepreneurial Firms: Two Models of Strategic Momentum," *Strategic Management Journal* (3:1), Jan-Mar 1982, pp 1-25.

Mimno, P. "How to Get a Fast Payoff for Your Data Warehousing Investment," TDWI Flash point, 2001.

Mimno, P. "How to avoid data mart chaos using hybrid methodology," TDWI Flash point, 2002.

Mingers, J. "Combining IS Research Methods: Towards a Pluralist Methodology.," *Information Systems Research* (12:3), Sep2001 2001, pp 240-260.

Moncla, B. "The Rise of the i-Market: The Convergence of E-Business and Business Intelligence," in: *DM Review*, 2000.

Mukherji, A. "The evolution of information systems: Their impact on organizations and structures," *Management Decision* (40:5/6) 2002, pp 497.

Nemati, H.R., Steiger, D.M., Iyer, L.S., and Herschel, R.T. "Knowledge warehouse: An architectural integration of knowledge management, decision support, artificial intelligence and data warehousing," *Decision Support Systems* (33:2), Jun 2002 2002, pp 143.

Niederman, F., Brancheau, J.C., and Wetherbe, J.C. "Information-Systems Management Issues for the 1990s," *MIS Quarterly* (15:4), Dec 1991, pp 475-495.

Nunally, J.C. *Psychometric Theory* McGraw Hill, New York, 1978.

Nutt, P.C. "The identification of solution ideas during organizational decision making," *Management Science* (39:9), Sep 1993, pp 1071-2001.

O'Donnell, D., and Gibson, M. "Data warehouse development methodologies: a comparative analysis," in: *Decision support systems laboratory*, Melbourne, Australia, 2002, p. 12.

Orlikowski, W.J., and Barley, S.R. "Technology and institutions: What can research on information technology and research on organizations learn from each other?" *Mis Quarterly* (25:2), Jun 2001, pp 145-165.

Orlikowski, W.J., and Baroudi, J.J. "Studying Information Technology in Organizations: Research Approaches and Assumptions," *Information Systems Research* (2:1), March 1991, pp 1-29.

Pfeffer, J. *Power in organizations* Pitman Publishing, Marshfield, 1981.

Pinsonneault, A., and Kraemer, K.L. "Survey research methodology in management information systems: An assessment," *Journal of Management Information Systems* (10:2), Fall 1993, pp 75-106.

Pitt, L., Watson, R.T., and Kavan, C. "Service quality: a measure of information systems effectiveness," *MIS Quarterly* (19:2) 1995, pp 249-266.

Powers, R.F., and Dickson, G.W. "MIS Project Management: Myths, Opinions, and Reality.," *California Management Review* (15:3), Spring 1973, pp 147-157.

Prahalad, C.K., and Krishnan, M.S. "The Dynamic Synchronization of Strategy and Information Technology," *MIT Sloan Management Review* (43:4), Summer2002 2002, pp 24-32.

Punch, K. *Introduction to social research* Sage, Thousand oaks, 1998.

Raghunathan, B., and Raghunathan, T.S. "Adaptation of a Planning System Success Model to Information Systems Planning," *Information Systems Research* (5:3) 1994, pp 326-340.

Rai, A., Lang, S.S., and Welker, R.B. "Assessing the validity of IS success models: An empirical test and theoretical analysis," *Information Systems Research* (13:1), Mar 2002, pp 50-69.

Ravichandran, T. "Innovation assimilation in the presence of knowledge barriers and technology uncertainty: the case of component-based software development," Academy of Management, Washington DC, 2001.

Rechtin, E. *Systems architecting, creating and building complex systems* Prentice hall, Englewood Cliffs, 1991, p. 299.

Richman, D. "Extraction tools: Better than nothing?" in: *Software magazine*, 1997, pp. 17-19.

Robey, D. "Research Commentary: Diversity in Information Systems Research: Threat, Promise, and Responsibility," *Information Systems Research* (7:4), Dec 1996, pp 400-409.

Robey, D., and Boudreau, M.C. "Accounting for the contradictory organizational consequences of information technology: Theoretical directions and methodological implications," *Information Systems Research* (10:2), Jun 1999, pp 167-185.

Robinson, M. "Business Intelligence Infrastructure," in: *BI Report*, 2002.

Rockart, J.E., Earl, M.J., and Ross, J.W. "Eight Imperatives for the New IT Organization," *Sloan Management Review* (38:1), Fall96 1996, pp 43-56.

Rockart, J.F., and Short, J.E. "IT in the 1990s: managing organizational interdependence," *Sloan Management Review* (7:Winter), Winter 1989, pp 23-34.

Ross, J.W. "Creating a Strategic IT Architecture Competency: Learning in Stages," *MIS Quarterly Executive* (2:1), March 2003, pp 31-44.

Ryan, S.D., and Prybutok, V.R. "Factors affecting the adoption of knowledge management technologies: a discriminant approach," *Journal of Computer Information Systems* (41:4), Summer2001 2001, pp 31-38.

Sambamurthy, V., and Kirsch, L.J. "An integrative framework of the information systems development process," *Decision Sciences* (31:2), Spring 2000, pp 391-411.

Seddon, P.B. "A respecification and extension of the Delone and McLean model of IS success," *Information Systems Research* (8:3), September 1997, pp 240-253.

Seddon, P.B. "Dimensions of information systems success," *Communications of the association of information systems* (2) 1999.

Sen, A., and Jacob, V.S. "Industrial-strength data warehousing," *Communications of the ACM* (41:9), Sep 1998, pp 28.

Sethi, V., and King, W.R. "Construct Measurement in Information Systems Research: An Illustration in Strategic Systems," *Decision Sciences* (22:3), Jul/Aug 1991 1991, pp 455-481.

Shin, B. "A case of data warehousing project management," *Information & Management* (39:7), Jul 2002, pp 581.

Shin, B. "An Exploratory Investigation of System Success Factors in Data Warehousing," *Journal of the Association for Information Systems* (4) 2003, pp 141-170.

Singh, H. **Interactive data warehousing*, 1999.

Sinha, S. "Point of return," in: *Intelligent Enterprise*, 2002, pp. 38-42.

Sowa, J.F., and Zachman, J.A. "Extending and Formalizing the Framework for Information Systems Architecture," *IBM Systems Journal* (31:3) 1992, pp 590.

Specht, P.H. "Job characteristics as indicants of CBIS data requirements," *MIS Quarterly* (10:3), Sep 1986, pp 271-287.

Sperley, E. *Enterprise data warehouse*, Wiley, New York, 1999.

Steinacher, S. "ETL provides the keys to keeping data relevant, available," in: *InfoWorld*, 2000, pp. 74-76.

Stewart, D.W. "The Application and Misapplication of Factor Analysis in Marketing Research," *Journal of Marketing Research* (1981:18), 51-64 1981, pp 1.

Strange, K. "Data warehouse: essential infrastructure for BI," AV-14-8985, Gartner Group.

Strange, K.F., T "Making BI and data warehousing strategic: The key issues," LE-19-4691, Gartner Group.

Straub, D.W. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2) 1989, pp 146-160.

Straub, D.W., Boudreau, M.-C., and Gefen, D. "Validation Guidelines for IS Positivist Research," *Communications of the association of information systems* (13:24), April 2004.

Swan, J., Newell, S., and Robertson, M. "The illusion of 'best practice' in information systems for operations management," *European Journal of Information Systems* (8:4) 1999, pp 284-293.

Swanson, E.B. "Information Systems Innovation among Organizations," *Management Science* (40:9), Sep1994 1994, pp 1069-1092.

Tait, P., and Vessey, I. "The Effect of User Involvement on System Success: A Contingency Approach," *MIS Quarterly* (12:1), March 1988, pp 91-108.

Teo, H.H., Tan, B.C.Y., and Wei., K.K. "Innovation diffusion theory as a predictor of adoption intention for financial EDI.," International Conference of Information Systems, Amsterdam, 1995, p. 15.

Timo, S. "An expanded instrument for evaluating information system success," *Information & Management* (31:2), November 1996, pp 103-119.

Torkzadeh, G., and Doll, W.J. “The development of a tool for measuring the perceived impact of information technology on work.," *The International Journal of Management Science* (27:3) 1999, pp 327-339.

Tractinsky, N., and Jarvenpaa, S. "Information systems design decisions in a global versus domestic context," *MIS Quarterly* (19:4), Dec1995 1995, pp 507-534.

Tushman, M.L., and Nadler, D.A. "Information processing as an integrating concept in organizational design," *Academy of Management Review* (3:3) 1978, pp 613-624.

Vaduva, A., and Vetterli, T. "Metadata Management for Data Warehousing: An Overview.," *International Journal of Cooperative Information Systems* (10:3), Sep 2001, pp 273-299.

Vail, E.F. "Causal Architecture: Bringing the Zachman Framework to life," *Information Systems Management.* (19:3), Summer 2002, pp 8-21.

Vandenberg, R.J., and Lance, C.E. "A review and synthesis of the measurement invariance literature: suggestions, practices and recommendations for organizational research," *Organizational research methods* (3:1) 2000, pp 4-69.

Varadarajan, P.R., Jayachandran, S., and White, J.C. "Strategic interdependence in organizations: Deconglomeration and marketing strategy," *Journal of Marketing* (65:1), Jan 2001, pp 15-29.

Volonino, L., Watson, H.J., and Robinson, S. "Using EIS to respond to dynamic business conditions," *Decision Support Systems* (14) 1995, pp 105-116.

Wang, R.W., and Strong, D.M. "Beyond accuracy: what data quality means to data consumers," *Journal of Management Information Systems* (12:4), Spring 1996.

Wateridge, J. "IT projects: a basis for success," *International Journal of Project Management* (13:3 SU -), 1995/6 1995, pp 173-176.

Wateridge, J. "How can IS/IT projects be measured for success?" *International Journal of Project Management* (16:1 SU -), 1998/2 1998, pp 59-63.

Watson, H.J. "Recent developments in data warehousing," *Communications of the association of information systems* (8) 2001, pp 1-25.

Watson, H.J., Goodhue, D.L., and Wixom, B.H. "The benefits of data warehousing: why some organizations realize exceptional payoffs," *Information & Management* (39:6), May 2002, pp 491-502.

Watson, H.J., Fuller, C., and Ariyachandra, T. "Data warehousing governance: Best Practices at Blue Cross Blue Shield of North Carolina," *Decision Support Systems* forthcoming.

Watson, H.J., and Haley, B.J. "Managerial considerations," *Communications of the ACM* (41:9) 1998, pp 32-37.

Watson, H.J., and Swift, R.S. "Data warehousing around the world," *Journal of Global Information Technology Management* (5:2) 2002, pp 1.

Watson, H.J., Watson, R.T., Singh, S., and Holmes, D. "Development Practices for Executive Information-Systems - Findings of a Field-Study," *Decision Support Systems* (14:2), Jun 1995, pp 171-184.

Watson, H.J., Wixom, B.H., Buonamici, J.D., and Revak, J.R. "Strategy: creating intelligence across supply chain," *Communications of the association of information systems* (5:9) 2001.

Watson, H.J., Wixom, B.H., and Goodhue, D.L. "The effects of technology-enabled business strategy at first American corporation," *Organizational Dynamics* (31:3) 2002, pp 313-324.

Weill, P., and Broadbent, M. *Leveraging the new infrastructure* Harvard business school, Boston, 1998.

Wells, D. "Choosing the right data warehouse approach," TDWI Flashpoint, 2003.

Wells, J.D., and Hess, T.J. "Understanding decision-making in data warehousing and related decision support systems: An explanatory study of a customer relationship management application," *Information Resources Management Journal* (15:4), Oct-Dec 2002, pp 16.

White, C. "Intelligent Business Strategies: Understanding Performance Management," in: *DM Review*, 2003.

Whiting, R. "Tower Of Power," *InformationWeek*, Feb. 11, 2002.

Whiting, R. "The Data Warehouse Advantage," in: *InformationWeek*, 2003.

Willcocks, L.P., and Sykes, R. "Enterprise resource planning: the role of the CIO and IT function in ERP," *Communications of the ACM* (43:4), April 2000, pp 32-38.

Winsberg, P. "Modeling the data warehouse and the data mart," in: *INFODB*, 1996, pp. 1-10.

Wixom, B.H., and Watson, H.J. "An empirical investigation of the factors affecting data warehousing success," *MIS Quarterly* (25:1), Mar 2001, pp 17.

Wybo, M.D., and Goodhue, D.L. "Using interdependence as a predictor of data standards - Theoretical and measurement issues," *Information & Management* (29:6), Dec 1995, pp 317-329.

Wynns, J., Van Brussel, H., Valckenaers, P., and Bongaerts, L. "Workstation Architecture in holonic manufacturing systems," *Cirp Journal on Manufacturing Systems* (26:4) 1997, pp 220-231.

Zachman, J.A. "A Framework for Information Systems Architecture," *IBM Systems Journal* (26:3) 1987, pp 276.

Zmud, R.W., and Boynton, A.C. "Survey measures and instruments in MIS: inventory and appraisal," in: *The Information Systems Research Challenge: Survey Research Methods,* K.L. Kraemer (ed.), Harvard Business School Press, Cambridge, 1991.

APPENDIX A: THE EXPERT PANEL

Data warehouse Expert	Accomplishments
Rob Armstrong	Director of Technical Marketing, Teradata A consultant and author of two books on data warehousing. Helped build the first Terabyte data warehouse. Been involved in all aspects of the industry from developing code for relational database systems, to leading implementations for some of the most successful data warehouses in the world.
Karolyn Duncan	A TDWI certified instructor, a TDWI Fellow & a consultant Has more than 10 years of practical experience in all elements of data warehousing development. Well published, has articles and several TDWI Ten Mistakes booklets to her credit. She has taught or consulted on data warehousing in eight other countries.
Wayne Eckerson	The director of education and research for TDWI. Oversees TDWI's education curriculum, member publications, and various research and consulting services. Has written and spoken extensively on data warehousing and business intelligence subjects since 1994. Prior to joining TDWI in 2000, He was director of the Patricia Seybold Group's Business Intelligence & Data Warehouse Service and coordinator of the group's business-to-business e-business coverage.

Jane Griffin	A managing director in the Enterprise Solutions Group at BearingPoint, Inc. Has designed and built business intelligence solutions and data warehouses for clients in numerous industries. Well published, has articles on numerous topics on data warehousing.
Douglas Hackney	The authority and spokesperson for the FED architecture. The president of Enterprise Group Ltd., a consulting company specializing in business intelligence (BI). Has over 20 years of experience in business management and in designing and implementing business intelligence solutions. A recognized BI industry leader, with a recent market research study showing him ranking #3 in name recognition in the BI space. A frequent and highly rated speaker at conferences around the globe. A frequent lecturer at leading MBA programs across the United States. Author of Understanding and Implementing Successful Data Marts. A contributing editor and writes a monthly column for DM Review.
Bill Inman	The authority and spokesperson for the EDW architecture. World-renowned expert, speaker and author on data warehousing, is widely recognized as the “father of data warehousing.” He is co-creator of the Corporate Information Factory and more recently, creator of the Government Information Factory. He has over 35 years of experience in database technology management and data warehouse design, and he is known globally for his seminars on developing data warehouses. He has been a keynote speaker for every major computing association and many industry

	conferences, seminars, and tradeshows. Has written more than 650 articles on a variety of topics about building, using, and maintaining the data warehouse and the Corporate Information Factory. His works have been published in major computing journals including Data Management Review. He has written 46 books, many of which have been translated into nine languages; one has sold over one-half million copies.
Jim Revak	IT manager at The Sherwin-Williams Co. Worked as a leader of data warehousing projects and technologies for eight years. He managed three enterprise data warehouse projects, two of which were recognized as Best Practice Award Winners by The Data Warehousing Institute.
Don Stoller	Director, Decision Services, Owens & Minor Has more than 24 years of experience in the IT field, concentrating in data warehousing/decision support and applications development for the mainframe and client/server environments. Owens & Minor is a two-time winner of TDWI's Best Practices in Data Warehousing Award, and was the 2000 Leadership in Data Warehousing Award winner. He was named one of the top 100 Premier IT Leaders by Computerworld magazine in 2000.
Ron Swift	Vice President of Strategic Customer Relationships for Teradata Internationally known consultant, and author in the areas of CRM, analytical marketing, customer knowledge systems, data warehousing, decision support. Has more than 30 years of experience with hundreds of clients on six continents to achieve their business strategies and goals. He

	also is a frequent lecturer at major universities, conferences, symposiums, and executive forums worldwide.
Jim Thomann	A TDWI Fellow, Ph.D. and consultant with Web Data Access. A widely recognized as a leading authority on data warehousing, object-oriented methods, and business process analysis. Has nearly 30 years experience in the information technology field. He has provided data warehousing consulting and training services to numerous organizations worldwide in the areas of data warehouse organizational readiness, project management, development methodologies, data modeling, tool evaluation, and implementation.

APPENDIX B – THE ONLINE QUESTIONNAIRE[30]

Architecture Survey: Marts and Warehouses

Please assist in this research effort by completing the online questionnaire based on the data mart(s) or warehouse in your organization (or if you are a consultant/vendor, refer to an architecture implementation in an organization you know well). Click the following link to access the questionnaire.

THE QUESTIONNAIRE

If you have any questions, you can contact either of the researchers listed below. Thank you very much for participating in this important research study! We look forward to your responses.

Hugh Watson
Professor of MIS
Terry College of Business
University of Georgia
Athens, Georgia 30602
Tel: (706)542-3744
E-mail: hwatson@uga.edu

Thilini Ariyachandra
Doctoral Candidate
Terry College of Business
University of Georgia
Athens, Georgia 30602
Tel: (706)542-4665
E-mail: thilinia@uga.edu

[30] The questionnaire contains branching logic based on choices made by respondents. The questionnaire also contains few more questions than described in the dissertation. These questions were included at the request of the panel of experts to gain more insight into architecture selection and success. They were not analyzed as part of the dissertation.

A **data warehouse** is an integrated data repository that supports decision making.
A **data mart** is smaller in scope and stores data for a limited number of business processes or subject areas.
A **data warehouse architecture** describes the main components of a data warehouse. It provides an overall framework for detailed design and development of a data warehouse.
In the survey that follows**, the term "data warehouse" is used to refer to both warehouses and marts**.

The Status of Data Warehousing in Your Company*:

Please answer the following questions about the data warehouse in your organization.
*If you are an independent data warehouse consultant or vendor, please answer the entire questionnaire by referring to a particular data warehouse implementation in a specific client organization.

1. Do you have a decision support architecture (i.e., data marts or data warehouse) currently in place?
   a. Yes
   b. No

   [If Yes, ask question 1a, If No, take respondent to section "To receive survey findings" On page 24.]

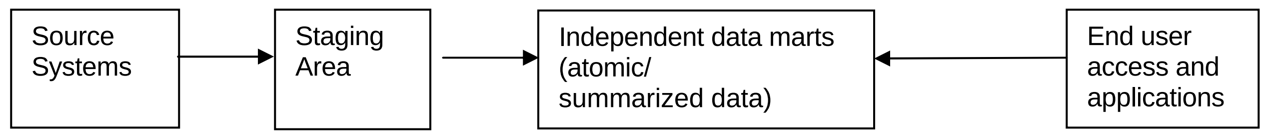
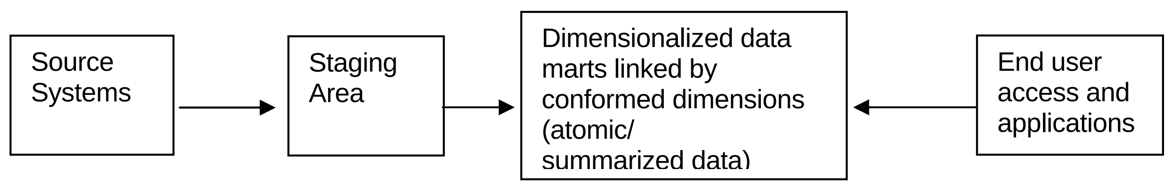
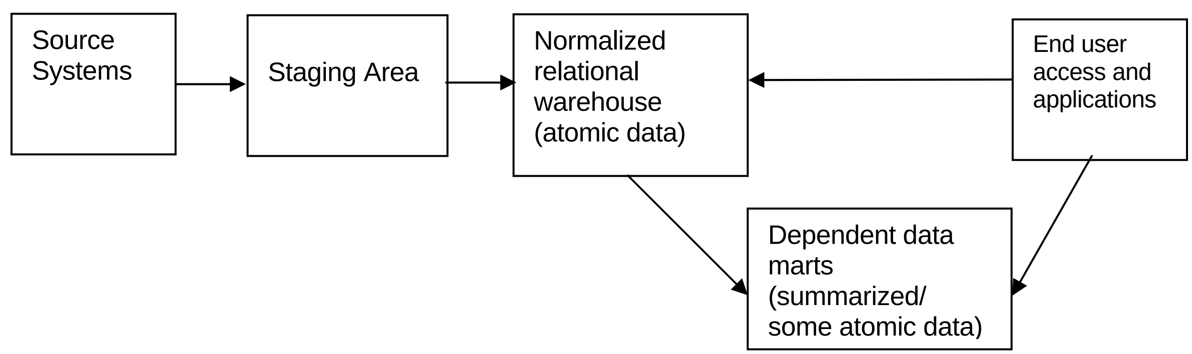
1a. It has been in use for ____________ years and ___________ months.

2. In answering the above question, what is the **domain of the business** for which the data warehouse architecture was implemented. (i.e., the business area or scope in the company for which the architecture was implemented)?
   a. Entire company (e.g., General Motors Corporation)
   b. Several but not all business units within the company (e.g., the Saturn and Chevrolet automobile divisions at General Motors)
   c. A single business unit with several functional areas (e.g., the Saturn automobile division at General Motors)
   d. A single functional area unit within a business unit (e.g., the marketing area within Saturn automobile division)
   e. A single subunit within a business unit (e.g., the used car fleet within Saturn automobile division)

[If the answer to question 2 is a, skip to question 4]
[If answer to question 2 is b, c, d, or e answer question 3]

3. Approximately what percentage of the entire company is in this domain? ______%

4. Please indicate which of the following best describes **your current data warehouse architecture**?
The figures represent the various **physical** architectures. You can click each architecture for a brief description.

DATA WAREHOUSE ARCHITECTURE	**COMPONENTS OF THE ARCHITECTURE**
1. Independent data marts	
2. Data mart bus architecture with linked dimensional data marts	
3. Hub and spoke architecture, Corporate Information Factory	

4. Centralized data warehouse (**no dependent data marts)**	Source Systems → Staging Area → Normalized relational warehouse (atomic/ some summarized data) ← End user access and applications
5. Federated	Existing Data Warehouses, Data Marts, and Legacy Systems → Logical/physical integration of common data elements ← End user access and applications
6. I do not know/ None of the above	

[if answer to question #4 was 6, please skip to question #7].

4a. How closely does your data warehouse architecture match the one you selected in the previous question?

Slight match			Moderate match			Exact match
1	2	3	4	5	6	7

5. Did you initially start out with a different data warehouse architecture but at some point switch to the current architecture?
   a. Yes
   b. No

[If Yes, ask following question 6, if No, skip to question 7]

6. Which of the following data warehouse architectures best describes the **architecture you started with**?
[Next skip to question 7]
The figures represent the various **physical** architectures. You can click each architecture for a brief description.

DATA WAREHOUSE ARCHITECTURE	**COMPONENTS OF THE ARCHITECTURE**
1. Independent data marts	Source Systems → Staging Area → Independent data marts (atomic/ summarized data) ← End user access and applications
2. Data mart bus architecture with linked dimensional data marts	Source Systems → Staging Area → Dimensionalized data marts linked by conformed dimensions (atomic/ summarized data) ← End user access and applications
3. Hub and spoke architecture, Corporate Information Factory	Source Systems → Staging Area → Normalized relational warehouse (atomic data) ← End user access and applications; Normalized relational warehouse (atomic data) → Dependent data marts (summarized/ some atomic data) ← End user access and applications

4. Centralized data warehouse (**no dependent data marts)**	Source Systems → Staging Area → Normalized relational warehouse (atomic/ some summarized data) ← End user access and applications
5. Federated	Existing Data Warehouses, Data Marts, and Legacy Systems → Logical/physical integration of common data elements ← End user access and applications
6. I do not know/ None of the above	

Please answer the rest of the questions in terms of how they relate to the *architecture currently implemented****. Many of the questions that follow ask about the organizational conditions*** *at the time that the current architecture was selected.*

Some of the questions may appear similar. Please understand that we need your answers to all the questions in order to thoroughly understand your opinions.

[If the answer to question #2 is a, then answer the following, if not skip to page 8]

Factors that Affected the Selection of Your Data Warehouse Architecture:

Please indicate the extent to which the following factors affected the selection of your company's data warehouse architecture.

Interdependence between Business Units within the Company:	Strongly Disagree			Neither/neutral			Strongly Agree
7. Close coordination among business units within the company was essential for them to successfully do their work.	1	2	3	4	5	6	7
8. The decisions and actions of every business unit had important implications for the operation of the other units within the company.	1	2	3	4	5	6	7
9. Information provided by other business units within the company was critical to each unit's performance.	1	2	3	4	5	6	7

The Upper Management of the Entire Company's Needs for Information:	Strongly			Neither/neu			Strongly
10. Senior management needed the ability to drill down into data from lower organizational levels in order to carry out their job responsibilities	1	2	3	4	5	6	7
11. Upper management needed the capability to drill through to detailed atomic level data.	1	2	3	4	5	6	7
12. Senior management needed an aggregated view of data from lower levels of the organization.	1	2	3	4	5	6	7
13. In order to comply with regulations (e.g., Sarbanes-Oxley Act), upper management needed a complete, accurate view of company information.	1	2	3	4	5	6	7

[Now, Skip to page 14]

[If the answer to question #2 is b, then answer the following, if not skip to page 10]

Factors that Affected the Selection of Your Data Warehouse Architecture:

Please indicate the extent to which the following factors affected the selection of your business units' data warehouse architecture.

Interdependence between Business Units within the Company:	Strongly Disagree			Neither/neutral			Strongly Agree
7. Close coordination among business units within the company was essential for them to successfully do their work.	1	2	3	4	5	6	7
8. The decisions and actions of every business unit had important implications for the operation of the other units within the company.	1	2	3	4	5	6	7
9. Information provided by other business units within the company was critical to each unit's performance.	1	2	3	4	5	6	7

Interdependence between Functional Area Units within a Business Unit:	Strongly Disagree			Neither/neutral			Strongly Agree
10. Close coordination among functional area units within a business unit was essential for them to successfully do their work.	1	2	3	4	5	6	7
11. The decisions and actions of every functional area unit had important implications for the operation of the other subunits or functional units within a business unit.	1	2	3	4	5	6	7
12. Information provided by other functional area units within a business unit was critical to each subunit's or functional area unit's performance.	1	2	3	4	5	6	7

The Upper Management of the Entire Company's Needs for Information:	Strongly Disagree			Neither/neutral			Strongly Agree
13. Senior management needed the ability to drill down into data from lower organizational levels in order to carry out their job responsibilities	1	2	3	4	5	6	7
14. Upper management needed the capability to drill through to detailed atomic level data.	1	2	3	4	5	6	7
15. Senior management needed an aggregated view of data from lower levels of the organization.	1	2	3	4	5	6	7
16. In order to comply with regulations (e.g., Sarbanes-Oxley Act), upper management needed a complete, accurate view of company information.	1	2	3	4	5	6	7

[Now, Skip to page 14]

[If the answer to question #2 is c, then answer the following, if not skip to page 12]

<u>Factors that Affected the Selection of Your Data Warehouse Architecture:</u>

Please indicate the extent to which the following factors affected the selection of your business unit's data warehouse architecture.

	Strongly Disagree			Neither/neutral			Strongly Agree
Interdependence between Business Units within the Company:							
7. Close coordination among business units within the company was essential for them to successfully do their work.	1	2	3	4	5	6	7
8. The decisions and actions of every business unit had important implications for the operation of the other units within the company.	1	2	3	4	5	6	7
9. Information provided by other business units within the company was critical to each unit's performance.	1	2	3	4	5	6	7

	Strongly Disagree			Neither/neutral			Strongly Agree
Interdependence between Functional Area Units within a Business Unit:							
10. Close coordination among functional area units within a business unit was essential for them to successfully do their work.	1	2	3	4	5	6	7
11. The decisions and actions of every functional area unit had important implications for the operation of the other subunits or functional units within a business unit.	1	2	3	4	5	6	7
12. Information provided by other functional area units within a business unit was critical to each subunit's or functional area unit's performance.	1	2	3	4	5	6	7

The Upper Management of the Business Unit's Needs for Information:

	Strongly Disagree			Neither/neutral			Strongly Agree
13. Senior management of the business unit needed the ability to drill down into data from lower organizational levels in order to carry out their job responsibilities	1	2	3	4	5	6	7
14. Upper management of the business unit needed the capability to drill through to detailed atomic level data.	1	2	3	4	5	6	7
15. Senior management of the business unit needed an aggregated view of data from lower levels of the organization.	1	2	3	4	5	6	7
16. In order to comply with regulations (e.g., Sarbanes-Oxley Act), upper management of the business unit needed a complete, accurate view of company information.	1	2	3	4	5	6	7

[Now, Skip to page 14]

[If the answer to question #2 is d, then answer the following, if not skip to page 13]

Factors that Affected the Selection of Your Data Warehouse Architecture:

Please indicate the extent to which the following factors affected the selection of your functional area unit's data warehouse architecture.

Interdependence between Functional Area Units within a Business Unit:	Strongly Disagree			Neither/neutral			Strongly Agree
7. Close coordination among functional area units within a business unit was essential for them to successfully do their work.	1	2	3	4	5	6	7
8. The decisions and actions of every functional area unit had important implications for the operation of the other functional units within a business unit.	1	2	3	4	5	6	7
9. Information provided by other functional area units within a business unit was critical to each functional area unit's performance.	1	2	3	4	5	6	7

The Upper Management of the Business Unit's Needs for Information:	Strongly Disagree			Neither/neutral			Strongly Agree
10. Senior management of the business unit needed the ability to drill down into data from lower organizational levels in order to carry out their job responsibilities	1	2	3	4	5	6	7
11. Upper management of the business unit needed the capability to drill through to detailed atomic level data.	1	2	3	4	5	6	7
12. Senior management of the business unit needed an aggregated view of data from lower levels of the organization.	1	2	3	4	5	6	7
13. In order to comply with regulations (e.g., Sarbanes-Oxley Act), upper management of the business unit needed a complete, accurate view of company information.	1	2	3	4	5	6	7

[Now, Skip to page 14]

[If the answer to question #2 is e, then answer the following]

Factors that Affected the Selection of Your Data Warehouse Architecture:

Please indicate the extent to which the following factors affected the selection of your subunit's data warehouse architecture.

Interdependence between Subunits within a Business Unit:	Strongly Disagree			Neither/neutral			Strongly Agree
7. Close coordination among subunits within a business unit was essential for them to successfully do their work.	1	2	3	4	5	6	7
8. The decisions and actions of every subunit had important implications for the operation of the other subunits within a business unit.	1	2	3	4	5	6	7
9. Information provided by other subunits within a business unit was critical to each subunit's performance.	1	2	3	4	5	6	7

The Upper Management of the Business Unit's Needs for Information:	Strongly Disagree			Neither/neutral			Strongly Agree
10. Senior management of the business unit needed the ability to drill down into data from lower organizational levels in order to carry out their job responsibilities	1	2	3	4	5	6	7
11. Upper management of the business unit needed the capability to drill through to detailed atomic level data.	1	2	3	4	5	6	7
12. Senior management of the business unit needed an aggregated view of data from lower levels of the organization.	1	2	3	4	5	6	7
13. In order to comply with regulations (e.g., Sarbanes-Oxley Act), upper management of the business unit needed a complete, accurate view of company information.	1	2	3	4	5	6	7

[Now, Skip to page 14]

[The numbering may change if the respondent selected option b or c for question #2]

Source and Support from Sponsor:

14. The **main source** of sponsorship for the data warehouse:
    a. was from a single functional area unit or a subunit.
    b. was from a single business unit.
    c. was from IT.
    d. was from multiple business units.
    e. had senior business management (i.e., CXO) support.
    f. sponsorship permeated the entire company.

15. The main sponsor(s) provided strong and sustained support for the data warehouse initiative.

Strongly Disagree						Strongly Agree
1	2	3	4	5	6	7

Strategic View of the Data Warehouse Prior to Implementation:	Strongly Disagree			Neither/neutral			Strongly Agree
16. The implementation of a data warehouse was viewed as a point solution to meet a functional area unit(s) or subunit(s) need.	1	2	3	4	5	6	7
17. The implementation of a data warehouse was viewed as an infrastructure project to support a range of applications.	1	2	3	4	5	6	7
18. The implementation of a data warehouse was viewed as a solution to support strategic objectives.	1	2	3	4	5	6	7

Urgency of Need for a Data Warehouse:	Strongly			Neither/neutral			Strongly Agree
19. There was pressure to build the data warehouse quickly.	1	2	3	4	5	6	7
20. The business needs demanded a fast implementation of the data warehouse.	1	2	3	4	5	6	7
21. A fast project turnaround time was critical in developing the data warehouse.	1	2	3	4	5	6	7

Availability of Resources:	Strongly			Neither/neutral			Strongly Agree
22. The availability of business personnel constrained the choice of architecture.	1	2	3	4	5	6	7
23. The availability of IT personnel constrained the choice of architecture.	1	2	3	4	5	6	7
24. The availability of monetary resources constrained the choice of architecture.	1	2	3	4	5	6	7
25. The anticipated annual operating monetary budget for the data warehouse constrained the choice of architecture.	1	2	3	4	5	6	7

Expert Influence:	Strongly			Neither/neutral			Strongly Agree
26. The choice of data warehouse architecture was strongly influenced by data warehousing consultants.	1	2	3	4	5	6	7
27. The choice of data warehouse architecture was strongly influenced by the data warehouse literature.	1	2	3	4	5	6	7
28. The choice of data warehouse architecture was strongly influenced by attendance at data warehousing seminars and conferences.	1	2	3	4	5	6	7
29. The choice of data warehouse architecture was strongly influenced by internal data warehousing experts.	1	2	3	4	5	6	7
30. End users strongly influenced the data warehouse architecture that was selected.	1	2	3	4	5	6	7

Nature of End User Tasks:	Strongly Disagree			Neither/neutral			Strongly Agree
31. Users often had questions that could not be addressed by structured queries and standard reports.	1	2	3	4	5	6	7
32. Users often had to answer questions that were novel and unique.	1	2	3	4	5	6	7
33. Users often faced questions that they had never answered before.	1	2	3	4	5	6	7

The Perceived Ability of the In-house IT Staff:	Strongly Disagree			Neither/neutral			Strongly Agree
34. The in-house IT staff had the necessary technical skills for developing the data warehouse.	1	2	3	4	5	6	7
35. The in-house IT staff had prior experience successfully implementing a data warehouse.	1	2	3	4	5	6	7
36. The IT staff was confident that it could successfully implement the data warehouse.	1	2	3	4	5	6	7

Compatibility with Existing Systems:	Strongly			Neither/neut			Strongly
37. Being compatible with the systems (e.g., applications, data bases, ERP systems) that were already in place influenced the architecture that was implemented.	1	2	3	4	5	6	7

Additional Technical Factors: Please indicate the extent to which the following technical factors affected the selection of your company's data warehouse architecture.	Strongly			Neither/neutral			Strongly Agree
38. Meta data integration was an important consideration in selecting the data warehouse architecture.	1	2	3	4	5	6	7
39. Scalability (e.g., amount of data, number of users, query performance) was an important consideration in selecting the data warehouse architecture.	1	2	3	4	5	6	7
40. The ability to maintain historical data was an important consideration in selecting the data warehouse architecture.	1	2	3	4	5	6	7
41. The ability to adapt to technical changes (e.g., volatile source systems) was an important consideration in selecting the data warehouse architecture.	1	2	3	4	5	6	7

The Importance of Factors that Affect the Selection of Your Data Warehouse Architecture:

Please indicate the importance of each of the following factors on the selection of your data warehouse architecture.

	Not Important			Neither/neutral			Very Important
42. Interdependence between Organizational Units: The extent to which the work of one organizational unit depended upon information from one or more other organizational units	1	2	3	4	5	6	7
43. Upper Management's Information Needs: The extent to which senior management's activities and performance were dependent on information from lower organizational levels.	1	2	3	4	5	6	7
44. Strategic View of the Data Warehouse Prior to Implementation: The extent to which implementing a data warehouse was viewed as being important to supporting strategic objectives.	1	2	3	4	5	6	7
45. Urgency of Need for a Data Warehouse: The extent to which there was an urgent need to build the data warehouse.	1	2	3	4	5	6	7
46. Availability of Resources: The extent to which IT personnel, business unit personnel, and monetary resources were available for building the data warehouse.	1	2	3	4	5	6	7
47. Expert Influence: The extent to which the organization's choice of data warehouse architecture was influenced by sources of expertise.	1	2	3	4	5	6	7
48. Nature of End User Tasks: The extent to which users' jobs required non-routine data analyses.	1	2	3	4	5	6	7
49. The Perceived Ability of the In-house IT Staff: The extent of the perceived ability of the in-house IT staff in terms of the relevant technical skills, successful experiences, and confidence in developing a data warehouse.	1	2	3	4	5	6	7
50. Compatibility with Existing Systems: The extent to which the data warehouse architecture was compatible with existing systems.	1	2	3	4	5	6	7
51. Technical issues: The extent to which technical issues (e.g., scalability, meta data integration) affected the data warehouse architecture.	1	2	3	4	5	6	7

Success Indicators

Please indicate the extent to which the data warehouse architecture impacts the following success measures for your data warehouse.

Flexibility:

	Strongly Disagree			Neither/neutral			Strongly Agree
52. Your data warehouse architecture makes it is easy to add new business processes and subject areas.	1	2	3	4	5	6	7
53. Your data warehouse architecture provides the capability to satisfy new requirements quickly.	1	2	3	4	5	6	7
54. Your data warehouse architecture provides the capability to easily support future application needs.	1	2	3	4	5	6	7

Scalability:

	Strongly			Neither/neutra			Strongly
55. Your data warehouse architecture is scalable to handle increases in the number of users without negatively impacting system performance.	1	2	3	4	5	6	7
56. Your data warehouse architecture is scalable to handle increases in the complexity and number of simultaneous queries without degrading system performance.	1	2	3	4	5	6	7
57. Your data warehouse architecture is easily scalable to handle increases in the volume of data.	1	2	3	4	5	6	7

Integration:

	Strongly			Neither/neutral			Strongly Agree
58. Your data warehouse architecture supports and facilitates the integration of data from multiple systems.	1	2	3	4	5	6	7
59. Your data warehouse architecture supports and facilitates the integration of internal and external data sources.	1	2	3	4	5	6	7
60. Your data warehouse architecture supports and facilitates the integration of all needed data around primary keys.	1	2	3	4	5	6	7

Accuracy:

	Strongly Disagree			Neither/neutral			Strongly Agree
61. Data warehouse structured queries and reports contain few data errors.	1	2	3	4	5	6	7
62. Your data warehouse provides the level of data correctness needed for its intended purpose.	1	2	3	4	5	6	7
63. The data in the data warehouse correctly represent the real world objects and events.	1	2	3	4	5	6	7

Completeness:

	Strongly Disagree			Neither/neutral			Strongly Agree
64. Your data warehouse architecture includes data about all the business processes and subject areas that are required by users and applications.	1	2	3	4	5	6	7
65. All necessary decision support data is available within the data warehouse.	1	2	3	4	5	6	7
66. Your data warehouse architecture provides all the data needed by users and applications.	1	2	3	4	5	6	7

Consistency:

	Strongly Disagree			Neither/neutral			Strongly Agree
67. Your data warehouse architecture provides "a single version of the truth."	1	2	3	4	5	6	7
68. Your data warehouse architecture reduces data inconsistencies.	1	2	3	4	5	6	7
69. Your data warehouse architecture provides "a single system of record" for decision-support data.	1	2	3	4	5	6	7

Individual Impact:	Strongly Disagree			Neither/neutral			Strongly Agree
70. End users are using the data warehouse that was implemented.	1	2	3	4	5	6	7
71. Users can access the data more easily and quickly because of the data warehouse.	1	2	3	4	5	6	7
72. The data in the data warehouse is easy and intuitive for users to understand and use.	1	2	3	4	5	6	7
73. Your data warehouse enables users to think about, ask questions, and explore issues in ways that were previously not possible.	1	2	3	4	5	6	7
74. Your data warehouse has improved the decision-making capabilities of end users.	1	2	3	4	5	6	7
Organizational Impact:	Strongly Disagree			Neither/neutral			Strongly Agree
75. Your data warehouse architecture has met the business requirements for which it was implemented.	1	2	3	4	5	6	7
76. Your data warehouse architecture has greatly facilitated the use of Business Intelligence.	1	2	3	4	5	6	7
77. Your data warehouse architecture has enabled improvements in business processes.	1	2	3	4	5	6	7
78. Your data warehouse architecture has supported the achievement of strategic business objectives.	1	2	3	4	5	6	7
79. Your data warehouse architecture has led to high and measurable ROI.	1	2	3	4	5	6	7
80. Your data warehouse architecture has improved communication and cooperation across organizational units.	1	2	3	4	5	6	7

81. Which of the following best describes your data warehousing initiative?
    a. It is potentially in trouble.
    b. It is an up and coming system.
    c. It is a runaway success.

About Your Company:

Please answer the following questions about your organization.

82. Please indicate the name of your company: ________________

83. Please indicate the **state/province and country** in which the data warehouse is located: _________

84. Please check the activity that best describes the **primary business** of your company.

- Computer manufacturing
- Consulting/professional services
- Education/publishing
- Financial services/banking
- Government
- Healthcare
- Insurance
- Manufacturing (non-computer)
- Retail/Wholesale/distribution
- Software/Internet
- Telecommunications
- Transportation/logistics
- Utilities
- Other; please specify: ______________________

87. Please estimate the approximate 2004 **gross revenues or operating budget** (U.S. $) of the company:

b. Less than 10 million
c. 10 million – 100 million
d. 100 million -500 million
e. 500 billion - 1 billion
f. 1 billion – 10 billion
g. Above 10 billion
h. Not sure

88. Please estimate the approximate 2004 **number of employees** in your company:

a. 1 - 99
b. 100 - 499
c. 500 - 2,499
d. 2,500 – 4,999
e. 5,000 – 9,999
f. Above 10,000
g. Not sure

About Your Data Warehouse:

Please answer the following questions about your data warehouse.

89. Cost of the data warehouse

With reference to the data warehouse architecture implemented in your organization:

a. What was the **cost** (US $) (e.g., hardware/software, personnel) of developing **the first business process(es) or subject area(s)** ?
    a. Less than 100,000
    b. 100,000 – 500,000
    c. 500,000 – 1 million
    d. 1 million – 5 million
    e. 5 million – 10 million
    f. Above 10 million
    g. Not sure

b. The cost of developing the first business process(es) or subject area(s) was:
    a. Over budget
    b. On budget
    c. Under budget

c. What is **the annual cost** (US $) (e.g., hardware/software, personnel) of **maintaining** (not enhancements) the architecture?
    a. Less than 100,000
    b. 100,000 – 500,000
    c. 500,000 – 1 million
    d. 1 million – 5 million
    e. 5 million – 10 million
    f. Above 10 million
    g. Not sure

d. What is the **cost to date** (US $) (e.g., hardware/software, personnel) of **developing** (both maintenance and enhancements) the overall architecture?
    a. Less than 100,000
    b. 100,000 – 500,000
    c. 500,000 – 1 million
    d. 1 million – 5 million
    e. 5 million – 10 million
    f. Above 10 million
    g. Not sure

e. The cost of developing the overall architecture was:
    a. Over budget
    b. On budget
    c. Under budget

90. Development time of the data warehouse.
With reference to the data warehouse architecture implemented in your organization:

a. How much **time was required to develop and roll out the first business process(es) or subject area(s)** in the architecture?

a. 3 months or less
b. 4 – 6 months
c. 7 – 12 months
d. 13 – 24 months
e. Over 24 months
f. Not sure

b. The time required to develop and roll out the first business process(es) or subject area(s) was:
   a. Behind schedule
   b. On schedule
   c. Ahead of schedule

91. The data warehouse has _______________ active users (e.g., access reports at least once a month).

92. How much raw data is in your data warehouse?
   a. 0-1GB
   b. 1-100GB
   c. 100-250GB
   d. 250-500GB
   e. 500-1TB
   f. 1-5TB
   g. 5-10TB
   h. 10TB+
   i. Not sure

93. Please indicate the vendor database platform for each of the following (Please check all that apply):
   a. Your physically separate data marts (if applicable):
      a. IBM
      b. Microsoft
      c. Oracle
      d. Teradata
      e. Sybase
      f. Other

   b. Your data warehouse (if applicable):
      a. IBM
      b. Microsoft
      c. Oracle
      d. Teradata
      e. Sybase
      f. Other

About Yourself:

Please answer the following questions about **yourself**.
94. Which of the following best describes your position in the organization?

a. DW manager
b. DW staff member
c. IS manager or professional
d. Independent consultant or systems integrator
e. Vendor (sales, service, support, or development)
f. Other; please specify ____________

95. Were you actively involved in the selection of the data warehouse architecture in your organization?

a. Yes
b. No

To Receive Survey Findings

Thank you for completing the questionnaire. To receive a copy of the study findings, please supply:

- Your name: ______________________________
- Your email address: ______________________
- Your phone number: (optional)______________

SUBMIT

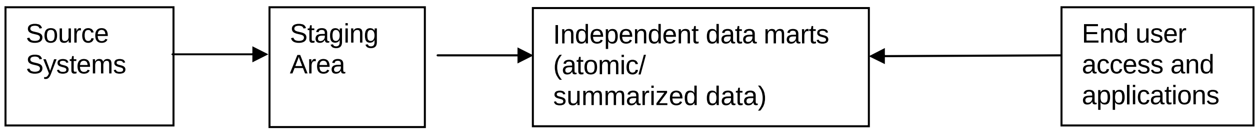
Thank you for your participation.

We would also like to get reactions from the person who works most closely with end users and warehouse applications. This person may hold the role of "Decision support/BI consultant, developer or, manager."

Please ask this person to complete the survey by linking to___________________. They will only be asked to respond to 27 questions that relate to the use of the data warehouse. If you prefer, please specify the email address of this person so that we can send them email with the Web link to the survey.
Email address: _______________________

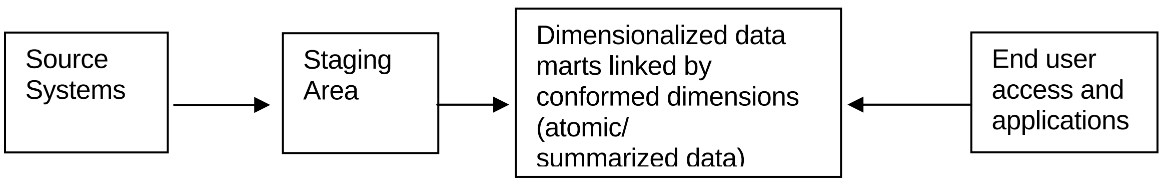
[Take email address and send a Web link to a separate survey on architecture success]

Independent data mart architecture



- Each data mart is developed independently.
- Data marts do not have conformed dimensions (i.e., do not have the same dimensions and measures for the data in data marts, which would allow the data across the data marts to be combined).
- Each mart is built to meet the needs of a separate business unit, functional area, or department.

Data mart bus architecture with linked dimensional data marts

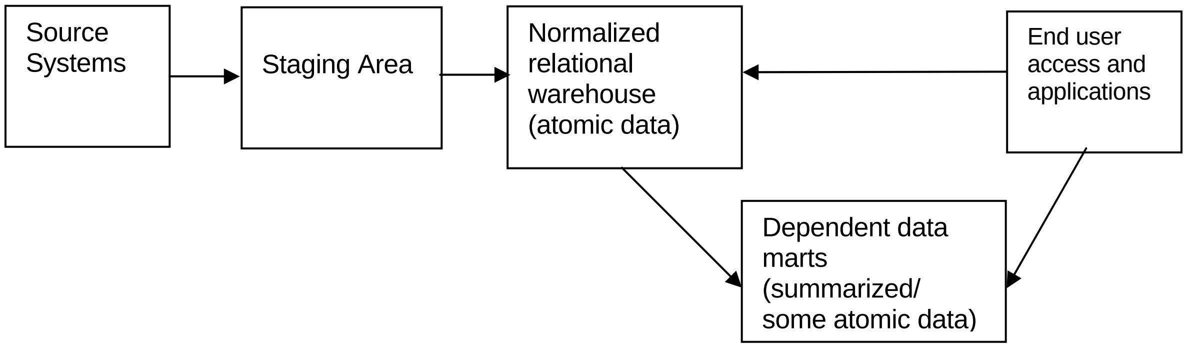


- The architecture is created through a business requirements analysis, which leads to creating the preliminary data bus architecture with common business dimensions.
- The first data mart is built to support a specific business process such as orders, deliveries, customer calls, or billing.
- Additional data marts are built sequentially using a master suite of conformed dimensions and fact tables, resulting in logically integrated marts and an enterprise view of the data.
- The staging area includes flat files, normalized, XML data sets, and dimensional schemas. The presentation area is composed of dimensionalized data marts.
- Atomic data is accessed from the data mart(s).
- As described in

Kimball, R., Reeves, L., Ross, M., and Thorthwaite, W. *The Data Warehouse Lifecycle Toolkit*, Wiley, New York, 1998.

Kimball, R., and Ross, M. *The Data Warehouse Toolkit*, Wiley, New York, 2002.

Hub and spoke architecture, Corporate Information Factory



- The architecture is created through an enterprise level analysis of data requirements to gain a detailed enterprise view.
- Using the enterprise view of data, the architecture is developed in an iterative manner, subject area by subject area.
- Based on user needs for decision support data, dependent data marts are developed that source data from the normalized warehouse. *The marts may have normalized, denormalized, flat files,*

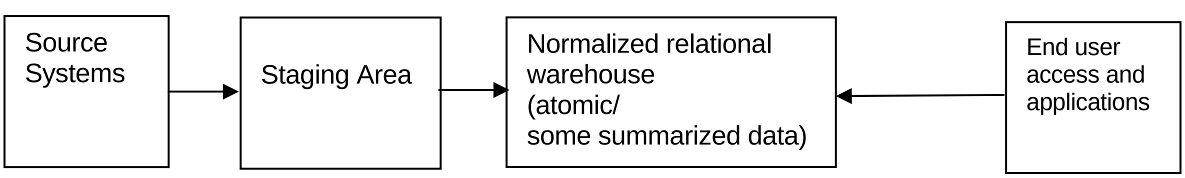
or summary dimensional data structures based on user needs.

- Dependent data marts may be developed for departmental, functional area, or special purposes (e.g., data mining).
- Most users query the dependent data marts, not the data warehouse.
- As described in

Inmon, W.H., Imhoff, C., and Sousa, R. *Corporate Information Factory, Second Edition*, Wiley, New York, 2001.

Imhoff, C., Galemmo, N., and Geiger, J. *Mastering Data Warehouse Design*, Wiley, New York, 2003.

Centralized data warehouse (no dependent data marts)

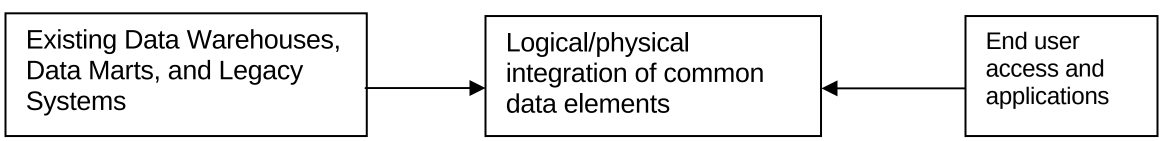


- The architecture is created through an enterprise level analysis of data requirements to gain a detailed enterprise view.
- Using the enterprise view of data, the architecture is developed in an iterative manner, subject area by subject area.
- The data warehouse contains atomic data, some summarized data, and logical dimensional views of data.
- All queries and applications can access data from both the relational data and dimensional views.
- It maybe a logical rather than a physical implementation of the Corporate Information Factory.
- As described in

Inmon, W.H., Imhoff, C., and Sousa, R. *Corporate Information Factory, Second Edition,* Wiley, New York, 2001.

Imhoff, C., Galemmo, N., and Geiger, J. *Mastering Data Warehouse Design*, Wiley, New York, 2003.

Federated architecture



- The architecture leaves existing analytic structures in place and involves combining data in an organization's existing data environment.
- It is characterized by either logically or physically combing key metrics and measures to some degree using shared keys, shared columns, global metadata, distributed queries, or some other method.