

# Key organizational factors in data warehouse architecture selection

Thilini Ariyachandra<sup>a</sup>, Hugh Watson<sup>b,\*</sup>

<sup>a</sup> Xavier University, United States

<sup>b</sup> University of Georgia, United States

## ARTICLE INFO

### Article history:

Received 2 July 2007

Received in revised form 16 February 2010

Accepted 22 February 2010

Available online 4 March 2010

### Keywords:

Data warehousing

Data warehouse architecture

Architecture selection

Independent data marts

Bus architecture

Enterprise data warehouse

Hub and spoke

Federated

## ABSTRACT

Even though data warehousing has been in existence for over a decade, companies are still uncertain about a critical decision — which data warehouse architecture to implement? Based on the existing literature, theory, and interviews with experts, a research model was created that identifies the various contextual factors that affect the selection decision. The results from the field survey and multinomial logistic regression suggest that various combinations of organizational factors influence data warehouse architecture selection. The strategic view of the data warehouse prior to implementation emerged as a key determinant. The research suggests an overall model for predicting the data warehouse architecture selection decision.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Over the past decade, many companies have made data warehouses the foundation of their decision support infrastructures. These data repositories provide a solution to a recurring problem that has plagued companies' decision support initiatives — the lack of clean, accurate, timely, and integrated data. Whether the data is used for queries and reporting, decision support systems (DSS), executive information systems (EIS), online analytical processing (OLAP), or data mining, data warehouses provide the data that “fuels” these applications [36]. Data warehouses are also critical enablers of current strategic initiatives such as customer relationship management (CRM), business performance management (BPM), and supply chain management [18,23,35,85].

Despite the importance and growing experience with data warehousing, there is still a considerable discussion and disagreement over which architecture to use. Bill Inmon, commonly referred to as “the father of data warehousing,” advocates what is variously called the hub and spoke architecture (i.e., a centralized data warehouse with dependent data marts), Corporate Information Factory, DW 2.0 (i.e., Inmon's current term), or enterprise data warehouse architecture [23]. Another architecture alternative is the data mart bus architecture with linked dimensional data marts (i.e., bus architecture), advocated by Ralph Kimball, the other preeminent figure in data warehousing

[52].<sup>1</sup> These and other architectures (e.g., independent data marts and federated) have fundamental differences and strong advocates.

The selection of an appropriate architecture is an important decision. A study by the Meta Group found that architecture selection is one of the key factors influencing data warehousing success [54]. A Gartner report identified the architecture selection decision as one of the five problem areas associated with data warehouse projects [72]. A poor architecture selection decision can lead to problems such as lack of scalability, performance difficulties, and no “single version of the truth.”

The data warehouse architecture selection decision is a subset of IT infrastructure (ITI) design. Currently, the academic literature contains only limited research on ITI design [21,87] and typically presents findings from case studies or offers anecdotal recommendations [10,85]. A recent article [68], one of the few empirical studies on ITI infrastructure, analyzes current practices in data warehouse implementation methodologies but does not explicitly describe factors that influence architecture selection. Another exploratory study [58] on data warehousing identifies a list of organizational factors that influence data warehousing refresh policies but does not examine how these factors influence data warehouse design. In contrast, the IS field is rich with research on the design of IT applications [8,16]. This latter substantial body of literature does suggest factors that

\* Corresponding author.

E-mail addresses: [ariyachandrat@xavier.edu](mailto:ariyachandrat@xavier.edu) (T. Ariyachandra), [hwatson@uga.edu](mailto:hwatson@uga.edu) (H. Watson).

<sup>1</sup> Kimball takes exception with the use of the enterprise data warehouse term to describe the Inmon architecture, arguing that his approach also results in an enterprise-wide solution when fully and correctly implemented. Despite this objection, the term is widely used and is also used in this paper.

potentially influence ITI design. Many of these factors are theory-based and draw from a variety of rational and social/political theories. However, it is unclear how these factors affect ITI design decisions, and more specifically, how they influence decision support infrastructure design choices.

Considering the millions of dollars that companies typically spend on data warehouses, it is surprising that there is no empirical, theory-based research on warehouse architecture selection. This lack of rigorous academic research motivated our study. As part of our research, we were interested in understanding the factors that affect the architecture selection decision. More specifically, we wanted to explore the following research questions: “What factors are most important to the architecture selection decision?” “What factors influence the selection of a particular architecture?” And finally, we were interested in: “Do the findings on the data warehouse architecture selection decision confirm and extend current understandings about ITI decisions?” In order to investigate these questions, a multi-phase study was conducted using both qualitative and quantitative methods.

This paper is organized as follows. First, the major data warehouse architectures are described. Next, the relevant literature, theory, and expert input used to develop the research model are described. The research model and the hypotheses are presented next, followed by a detailed description of the research methodology that was used. The data analysis section discusses the validity of the research model, testing of hypotheses, and additional analyses. In the concluding sections, the research results are discussed, an overarching model for data warehouse architecture selection is proposed, and the implications for research and practice are presented.

## 2. Data warehouse architectures

A data warehouse is a specially prepared data repository that is used to support decision making [36]. Data is extracted from data sources, transformed, and loaded (ETL) into data stores. The data is then made available for end user access and decision support applications. It can be considered as decision support data infrastructure that is used for multiple, varied decision support purposes.

Architecture is the structure of something [80]. While the term is usually associated with the kind or style of a building, it is used more broadly, including for example, the structure of human societies or the solar system. A *reference* architecture is a particular kind of architecture. It helps in understanding or explaining what currently exists or serves as a guide for creating something new, such as a building or a data warehouse. In this study we refer to four different data warehouse reference architectures. The reference architectures identify the alternative ways that data is extracted, transformed, loaded, and stored in a data warehouse.

The data warehousing literature provides discussions and examples of different architectures [47,52]. Vendors and consultants promote a variety of reference architectures to guide data warehouse implementation efforts [82]. Let us consider how the major architectures have evolved over time and how they were described in the study. As discussed later, the descriptions were based on the data warehousing literature and interviews with leading experts in the field. Other architectures are also discussed in the literature, but they tend to be variations on those that were studied [63].

The early 1970s ushered in decision support systems, which were fundamentally different from operational or transactional systems. A popular way of conceptualizing a DSS was through a dialog-data-model (DDM) paradigm Sprague et al. [71]. For the data component, it was recognized that a separate data repository was needed that drew data from operational systems and other data sources. In response, independent data marts were developed as the first decision support data infrastructure. This was an application-centric approach to data

management because the repositories were designed to support only a single or a few applications [83].

This evolution of decision support management can also be viewed from a maturity model perspective. In Eckerson's [23] six-stage model, which uses a human evolution metaphor, prenatal is the first stage. It features production reporting from operational systems. In the infant stage, there are “spreadmarts” in the form of Excel spreadsheets. Next is the child stage and it features independent data marts.

### 2.1. Independent data marts (IDM)

Just as companies have legacy operational systems, they also have legacy independent data marts.<sup>2</sup> In addition, some companies choose to create new marts. These marts are independent of other data stores, and while they may meet localized needs, they do not provide “a single version of the truth” for the entire organization. They typically have inconsistent data definitions and use different dimensions and measures (i.e., non-conformed) that make it difficult to analyze data across the marts [41]. Fig. 1 shows the architecture for each independent data mart.

In the late 1980s, a number of companies, especially in the financial services, telecommunications, and retail industries, developed the first data warehouses. The warehouses emerged, in part, because of the siloed nature of independent data marts. Companies wanted an enterprise-wide data repository (often focusing on all interactions with customers) to support a variety of analytical applications (e.g., queries, OLAP, and data mining). This approach represented a data-centric approach to decision support data management. In Eckerson's maturity model, it was a movement to the teenager stage.

Two competing architectures for data warehousing quickly emerged, each advocated by one of the two luminaries in the field: Ralph Kimball for the data mart bus architecture and Bill Inmon for the enterprise data warehouse.

### 2.2. Data mart bus architecture with linked dimensional data marts (DBA)

The development of the data mart bus architecture begins with the identification of the business requirements for a specific business process, such as orders, deliveries, customer calls, or billing. The first mart is built for a single business process using dimensions and measures (i.e., conformed dimensions) that will be used with other marts [52]. Additional marts are developed using these conformed dimensions, which result in logically integrated marts and an enterprise view of the data. Atomic and summarized data are maintained in the marts and are organized in star schemas to provide a dimensional view of the data. This architecture is illustrated in Fig. 2.

### 2.3. Enterprise data warehouse architecture (EDW)

An extensive enterprise-level analysis of data requirements provides the basis for this architecture [47]. Attention is focused on building a scalable infrastructure. Using the enterprise view of the data, the architecture is developed in an iterative manner, subject area by subject area. Atomic level data is typically maintained in the warehouse in 3rd normal form. Dependent data marts are created that

<sup>2</sup> Although we include independent data marts as a data warehouse architecture, it is not a formally advocated architecture in the industry. As discussed, independent data marts are present day stovepipe solutions that result from historical organizational efforts to build decision support solutions. Yet they exist and are used in organizations as data warehouse solutions. They are also commonly considered in discussions and surveys of the various data warehouse architectures. Consequently, we discuss independent data marts as an architecture in this study.

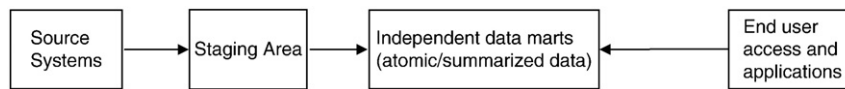


Fig. 1. The independent data marts architecture.

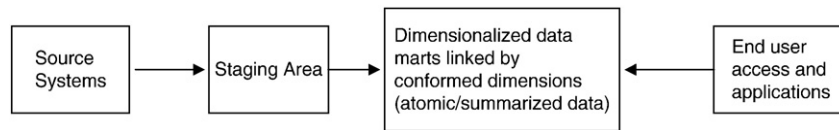


Fig. 2. The data mart bus architecture with linked dimensional data marts.

source data from the warehouse.<sup>3</sup> The dependent marts may be developed for departmental, functional area, or special purposes (e.g., data mining) and may have normalized, denormalized, or summarized/atomic dimensional data structures based on user needs. Most users query the dependent marts. Fig. 3 shows this architecture.

It should be mentioned that advocates of the DBA strongly believe that it provides an enterprise-wide solution, and they have a point. If the various marts are developed over time in a planned way with conformed dimensions, the end result is an enterprise data warehouse. By way of contrast, rather than starting with a single mart and growing, the Inmon approach (both in terms of architecture and methodology) is more enterprise-wide oriented at the outset. In terms of Eckerson's maturity model, the adult stage, which features an enterprise data warehouse, is the potential result of evolution (from the teenager stage) using the Kimball model, whereas it is the initial target of the Inmon approach.

#### 2.4. Federated architecture (FED)

The federated architecture was the final one studied. It leaves existing decision support structures (e.g., operational systems, data marts, and warehouses) in place [40]. It does not strive to integrate the existing complex decision support environment into a single integrated solution, often due to the political and implementation challenges faced by the organization [22]. It is often the result of acquisitions, mergers, and reorganizations. The data is either logically or physically integrated using shared keys, global metadata, distributed queries, enterprise information integration (EII), or other methods. This architecture is advocated as a practical 'realistic' solution for firms that have a preexisting, complex decision support environment and do not want to rebuild or plan to rebuild at a later date [49]. Firms with a federated architecture are typically in the teenager or adult stage of Eckerson's maturity model. This architecture is shown in Fig. 4.

### 3. Developing the research model

To ensure that the research model was relevant, comprehensive, and theory-based, multiple sources were used. First, expert interviews were conducted to identify the most salient factors that affect architecture selection. Using the results of the interviews as guidance, relevant theory and literature were explored to develop factors that potentially affect data warehouse architecture selection.

#### 3.1. Expert input

Interviews were conducted with 20 leading data warehousing experts. They were selected on the basis of being recognized authorities and spokespersons for the major architectures, leading authors and consultants, award winning data warehouse managers, and Fellows of The Data Warehousing Institute. The experts included Bill Inmon and Ralph Kimball, the proponents of the two architectures that are at the heart of the architecture controversy. Based on their experiences, the experts were asked to identify and discuss the factors that affect the data warehouse architecture decision. Specifically, their input was loosely structured around three questions on data warehouse architectures<sup>4</sup>. Depending on the expert's response, appropriate follow up questions were asked. The interviews were tape-recorded and typically took 30 min. Content analysis of the interviews surfaced seven selection factors that affect the architecture selection decision.

#### 3.2. Theoretical perspectives

There are two competing theories on organizational decision making that were helpful in framing and developing the research model. They are the rational and the social/political schools of organizational theory [55,65].

The proponents of the rational school argue for the use of comprehensive information processing when dealing with uncertainty to attain an organizational outcome-maximizing goal [34,61]. It promotes the performance of information systems (IS) design and development tasks in a way that minimizes implementation costs to the organization [14]. 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.

According to OIPT, variations in the degree of uncertainty faced by organizations drive the need for differing levels of information processing. Tushman and Nadler [79] identify three distinct sources of uncertainty: complex or non-routine tasks, unstable subunit task environment, and interdependence between subunits. These sources of uncertainty suggest the characteristics of the organizational context that may influence the choice of complex information processing mechanism used to satisfy information processing needs.

According to Galbraith [29], when confronted with greater levels of uncertainty, organizations may choose to use some combination of four complex mechanisms to reduce uncertainty. A data warehouse is an information system, which is one of the complex mechanisms prescribed by Galbraith [29]. The architectural design of a system ties organizational information requirements to the physical components

<sup>3</sup> The literature and some vendors make reference to the centralized architecture. This architecture is similar to the enterprise data warehouse architecture except that the dependent data marts are logical rather than physical. Because of the similarity of these two architectures in all other ways, they were treated as one for data analysis purposes.

<sup>4</sup> The three questions were: (1) What are the major data warehouse architectures? (2) What factors do you think affect architecture selection? (3) What factors help assess architecture success?

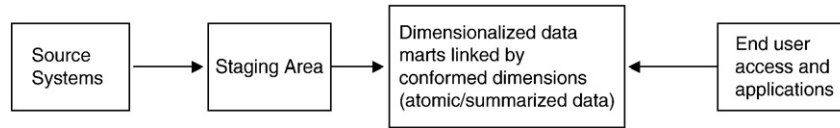


Fig. 3. The enterprise data warehouse architecture.

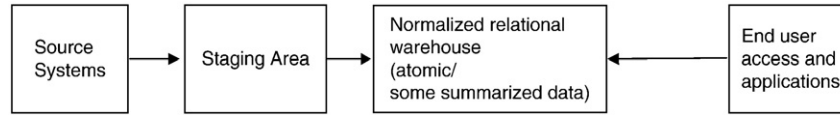


Fig. 4. The federated architecture.

necessary to satisfy those information processing requirements [28]. As such, organizational information requirements (i.e., sources of uncertainty) affect data warehouse architecture, which influences the information processing capacity of the organization. For instance, high interdependence creates the need to share information between organizational units. This information requirement needs a data warehouse architecture that provides enterprise-wide data integration and distribution capability. A hub and spoke architecture is one example of such an architecture.

The social/political school of organizational theory views decision making as a process of negotiation and coalition building in which multiple ambiguous goals exist [25]. This view contradicts the perception of the rational school that organizational action is determined by the need to accomplish an overarching organizational goal. In the context of data warehouse architecture selection, social/political theories challenge the notion that an overall company goal determines the architecture selection decision. Consequently as both the rational and political schools of organizational theory affect the selection process, factors from both schools should be considered.

The key concept underlying the social political theories is power. The literature on the structural perspective of power identifies many sources of power such as the ownership of scarce resources vital to the operation of a firm, or the possession of skills necessary to perform a critical function within an organization that enables one or more groups to acquire or possess power (e.g., [14,66]). The behavioral perspective describes the exercise of power or political activity by which groups influence actions and obtain favorable outcomes by virtue of their power [25].

Within IS, power is mostly discussed in terms of political activity that affects different aspects of system implementation. For instance, groups threatened by a system implementation may resist or hinder system development [60]. The process of developing an information system can be investigated as a series of political activities that may shape the final implementation and alter existing organizational structures. Consequently, data warehouse architecture selection can be perceived as a social/political change process in which organization members motivated by group interest use sources of power to influence architecture choice.

### 3.3. Research model and hypotheses

The insights gained from the expert interviews, the relevant theoretical foundations, and the literature were used to develop the research model; see Fig. 5. A summary of how these sources contributed to the research model is presented in Appendix A. The contextual factors and the hypotheses that relate them to the various architectures are discussed next. Hypotheses were created only when there was strong support for the statements from theory, literature, and the experts.

### 3.4. Information interdependence

Interdependence refers to the extent to which organizational units need to exchange information or material to accomplish their tasks [76]. It is a source of uncertainty when the information or actions taken by one unit affect the actions and work outcomes of other units [4]. According to OIPT, greater interdependence between units requires greater coordination/communication mechanisms to promote information sharing [34,77].

Implementing an information system, such as a data warehouse, is a mechanism that increases information processing capacity [29]. A data warehouse can increase the information processing capacity by creating a source of high quality, integrated data for decision support [20]. The architecture of a data warehouse influences the information processing capacity provided. 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.

When faced with the need for high information sharing among business units, an architecture solution that provides greater integration of data across units provide higher information processing capacity. An EDW architecture guides the implementation of such a solution; whereas the IDM architecture provides less integration across business units. Accordingly, the need for greater information between organizational units leads to the selection of an architecture that accommodates greater integration of data across units.

**Hypothesis 1 (H1).** Organizations with high interdependence are more likely to select an EDW than an IDM architecture.

### 3.5. Urgency

A dynamic environment is a major source of uncertainty that can be reduced through mechanisms that increase information processing [79]. Consequently, business environments with greater turbulence will experience greater urgency for information systems that provide accurate information.

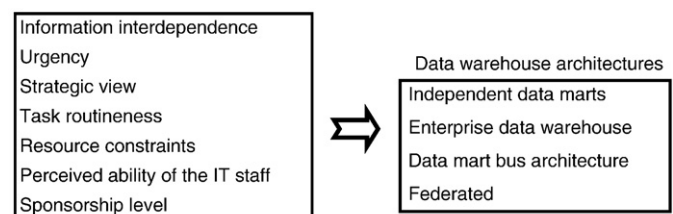


Fig. 5. Research model for data warehouse architecture selection.



Long wait times for systems to become operational are unacceptable in dynamic, uncertain markets [67]. In Allen et al.'s [2] investigation of IS architectures, they conclude that a dispersed rather than a centralized information system architecture is key to the swift implementation of new systems. Likewise, when the environment dictates greater urgency for information, it creates a need for a decentralized data warehouse solution. The IDM or DBA architectures are decentralized architectures that satisfy the urgent need for a warehousing solution; whereas, an EDW architecture, a centralized architecture, requires more time to implement.

**Hypothesis 2A (H2A).** Organizations with high urgency are more likely to select an IDM than an EDW architecture.

**Hypothesis 2B (H2B).** Organizations with high urgency are more likely to select a DBA than an EDW architecture.

### 3.6. Task routineness

Past IS research has focused on many characteristics of tasks that indicate greater or less need for information systems, such as task complexity [70], task analyzability [3], and task routineness [32]. Goodhue [32] describes tasks characterized by the analysis of ad hoc situations in new ways as non-routine tasks and 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. When potential users require the ability to do tasks characterized by high non-routineness, the information processing demands made on the data warehouse will be high.

Some architectures are better able to support information processing requirements dictated by task routineness than others. For example, the FED architecture relies on EAI/EII technologies much more so or exclusively so than others and these technologies limit the ability to do complex queries. In contrast, the EDW architecture provides integrated data in various levels of detail from data sources across the organization. Consequently, the EDW architecture is better able to support unstructured non-routine tasks.

**Hypothesis 3 (H3).** Organizations with tasks that are relatively less routine are more likely to select an EDW than a FED architecture.<sup>5</sup>

### 3.7. Strategic view of the data warehouse

Based on IT's impact on core operations and business strategy, firms may perceive IT to be an enabler of strategic transformation or merely a source of backend support for operations [62]. Information processing theory suggests that organizations may differ in their view of IT depending on the source of uncertainty they face. Duncan [21] asserts that organizations faced with uncertainty from a turbulent environment may perceive IT infrastructure as a mechanism that supports long-term competitive advantage rather than a short-term fix for immediate information pains. In their study of IT infrastructure investment decisions in organizations, Broadbent and Weill [11] noted that depending on their business and IT strategic objectives, organizations may differ in their view of IT infrastructure. Furthermore, the organizational view of IT suggests the need for different IT capabilities provides a high-level view of the specific IT infrastructure desired by an organization [87].

Depending on their motivation, organizations may differ in the manner in which they view the implementation of a data warehouse [15]. According to Griffin [37], expressing the future state or view of the desired data warehouse architecture is necessary in order to select

a suitable one. For example, if an organization views implementing decision support data infrastructure as a point solution to meet a functional area need, they may build an IDM rather than an enterprise data warehouse [51]. In contrast, if the warehouse is viewed as a means for bringing about organization-wide change, the organization might implement an EDW architecture [86]. Thus, the manner in which an organization views the data warehouse affects the data warehouse architecture selected.

**Hypothesis 4 (H4).** 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 than an EDW architecture.

### 3.8. Resource constraints

Organizational slack gives companies the opportunity to meet their need for increased information processing capacity by allocating resources for the development of information systems [63]. Conversely, the information processing capacity obtained from systems development may be constrained due to the lack of availability of resources. The information processing capacity provided by a data warehousing solution is influenced by its architecture. Thus, the choice of architecture is constrained by the availability of slack resources.

When faced with greater constraints on resources, an organization may select an architecture that requires fewer resources. For instance, many organizations with a limited budget may choose an IDM over an EDW architecture because it requires a smaller investment of resources [5]. Kimball et al. [52] and Mimno [64] state that the ability to build a data warehouse with relatively fewer resources leads firms to implement a DBA rather than an EDW architecture. Thus, resource constraints influence the data warehouse architecture selected.

**Hypothesis 5A (H5A).** Organizations with low resource availability are more likely to select an IDM than an EDW architecture.

**Hypothesis 5B (H5B).** Organizations with low resource availability are more likely to select a DBA than an EDW architecture.

### 3.9. The perceived ability of the in-house IT staff

Research in self-efficacy and organizational learning provide insights into how the perceived ability of the in-house IT staff affects the selection of an architecture. In a team context, specific self-efficacy is described as the team's belief about a task-specific team capability [39]. 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 [57,59]. In the context of this study, existing perceptions of their ability to develop a data warehouse can influence a data warehousing staff's decision about which architecture to use.

Organizational learning research suggests that there are many knowledge barriers that impede the adoption, development, and infusion of complex technologies such as a data warehouse [66]. Teams with experience related to a complex technology such as a data warehouse are more able to successfully adopt and sustain the complex technology [26]. Consequently, an IT staff may have low knowledge barriers to the adoption of a particular data warehouse architecture as a result of their existing experience and skills.

Breslin's [10] examination of data warehouse architectures suggests that developing an EDW architecture has high knowledge barriers in regard to technical skills and the experience of specialists, while the IDM and DBA architectures have relatively low knowledge barriers. Thus, the perceived ability of the in-house IT staff affects the architecture selected.

<sup>5</sup> This hypothesis was ultimately not testable due to the small number of responses for the federated architecture.

**Hypothesis 6A (H6A).** Organizations that have an IT staff with low perceived ability are more likely to select an IDM than an EDW architecture.

**Hypothesis 6B (H6B).** Organizations that have an IT staff with low perceived ability are more likely to select a DBA than an EDW architecture.

### 3.10. Sponsorship level

A sponsor is an individual or group that allocates resources for a systems development project, fights political resistance, and promotes the benefits of the project within the organization [88]. According to social/political theories, many sponsors wield power over other organizational members by controlling scarce resources [9,74,84]. With control over resources, the sponsor can influence decisions related to the project [48].

A sponsor may impose a system design that is most beneficial to the organizational level represented by the sponsor, or enforce a design that meets the sponsor's personal goals. For instance, a social/political view suggests that upper management may sponsor a data warehouse project to gain more control of the organization, such as an EDW solution. In contrast, when sponsorship for the data warehouse comes from a specific functional area, a DBA architecture may be selected [85]<sup>6</sup>. The practitioner literature suggests that project sponsorship for an IDM architecture often comes from business units [41]. Consequently, the architecture selected may be influenced by the source of sponsorship.

**Hypothesis 7A (H7A).** Organizations with upper management sponsorship are more likely to select an EDW than an IDM architecture.

**Hypothesis 7B (H7B).** Organizations with upper management sponsorship are more likely to select an EDW than a DBA architecture.

## 4. Research methodology

The research study had three phases. We previously described the first phase in which experts, the data warehousing literature, and theory were used to determine the architectures to study, the factors that affect the data warehouse architecture selection decision, the research model, and specific hypotheses to test. In the study's second phase, which is discussed next, survey instruments were developed and tested, and data was collected and analyzed to test the hypotheses and conduct additional analyses. In the study's third phase, selected experts and survey respondents were interviewed in order to help interpret the study's findings.

### 4.1. Instrument development

One of the challenges of positivist quantitative research is accurately capturing and measuring the social entities under investigation [73]. Many researchers use existing items when possible. However, according to Swanson [74], the context of existing research measures and questionnaires may not apply to a researcher's current project because the measures are deeply embedded in the research project that they pertain to. It is necessary to exercise caution

when adopting existing measures and should serve only as useful starting point in operationalizing variables of interest [90].

While theory provided some help in developing the questionnaire items, more assistance was gained from the experts and data warehousing literature. Whenever possible, existing, validated questions were used. However, when appropriate items were unavailable, a deductive, iterative approach to item development was used [44,69]. Multiple items were generated for the constructs and individual questions were constructed according to the recommendations of Babbie [6]. The items were refined through multiple iterations of review and feedback from the data warehousing experts. Finally, twelve data warehouse managers (i.e., potential survey respondents) were asked to pretest the online questionnaire. After several minor modifications were made, the resulting questionnaire contained seven constructs and 28 items, and demographic information questions. Most of the questions employed a seven-point Likert scale anchored at strongly disagree (1) and strongly agree (7).<sup>7</sup> The questions were phrased to ask about organizational conditions at the time the reported architecture was selected.

### 4.2. Dependent variable: the data warehouse architecture

The *architectural options* (i.e., IDM, DBA, EDW, and FED) and their descriptions were created through an extensive process of review and feedback from the experts as well as the data warehousing literature. Survey respondents were given both a graphical and written description of each possible architectural choice and asked to identify the data warehouse architecture currently implemented in their organizations<sup>8</sup>. Because there are many variations of the four main architectures, the respondents were asked (on a seven-point Likert scale) to indicate how closely their architecture matched the reference architecture they selected.

### 4.3. Independent variables: the contextual variables

Measuring *interdependence* is challenging because there are multiple possible organizational domains. Five domains were used based on the academic and data warehousing literature: entire company, several business units, single business unit, functional area unit, and sub unit [22,33]. Perceptual measures of interdependence by Wybo and Goodhue [89] were modified to fit the context of this study and used to measure information interdependence (II). The presentation of the items in the online survey was dynamically customized based on the domain (i.e., the wording of the questions was changed to fit the domain in which the respondent's data warehouse architecture was implemented).

IS research recognizes that the *urgency* of the business need impacts the decision to implement systems [2]. However, no items exist to measure the extent of urgency of the business need on technology choice. The expert interviews shed some light on how to conceptualize and measure urgency. One expert described urgency as "pressure to quickly meet business needs." Three items were created to measure urgency.

Based on the work of Goodhue [32] and input from the experts, *task routineness* was operationalized as the extent to which users' jobs required non-routine data analyses. Three items were used to measure task routineness by modifying the items from Goodhue [32] and Brohman [12].

<sup>6</sup> Sponsorship can also be considered as a rational factor. For example, the upper management of an organization characterized by high interdependence between organizational units may sponsor a data warehouse to ensure that the solution satisfies the need for greater information sharing across the entire organization. However, as social/political 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 considered the influence of the source of the sponsor from a social/political perspective.

<sup>7</sup> Each construct and corresponding items are available at [http://staff.xu.edu/~ariyachandrat/constructs\\_questions.htm](http://staff.xu.edu/~ariyachandrat/constructs_questions.htm).

<sup>8</sup> The respondents were asked to describe the architecture for a particular data warehouse implementation even if there were multiple implementations in the organization and answer questions for the data warehouse of that particular architecture.

The *strategic view* of the warehouse was based on Weill and Broadbent's [87] work on an organization's perception of IT infrastructure as either an enabler of strategy, a utility, strategy dependent, or none. The literature (e.g., [81]) and expert interviews indicated that "none" and "strategy dependent" are the most applicable and influential views or perceptions of a data warehouse implementation, and three questions were created to measure these two views.

Previous data warehouse research has measured *constraints on resources* in terms of time, money, and personnel [42]. Because the constraints on time were captured through the urgency construct, the respondent's perception of resource constraints in terms of money, IT personnel, and business personnel was measured.

Three items for the *perceived ability of the IT staff* were developed. One question measured the perception of the in-house IT staff's prior experience in implementing a data warehouse. Another captured the perception of the in-house IT staff's relevant technical skills. Finally, a third estimated the in-house IT staff's confidence in their ability to develop a data warehouse. Based on the work of Watson [83] and Eckerson [22], a single question was created to measure the *level of sponsorship* for the data warehousing projects.

#### 4.4. Data collection

Various professional organizations and leading authorities in the field were asked to promote the online survey to their clients/subscribers/members. For example, The Data Warehousing Institute (TDWI) included a message about the study in an electronic newsletter sent to over 5000 members. Because of the role and reputation of these promoters, and the large number of potential respondents and companies that receive these electronic communications, it was believed that the survey respondents would be representative of companies with data warehouses and the various architectures.<sup>9</sup> The target survey respondent was the person most closely involved with the design, development, and implementation of the company's data warehouse architecture, such as a data warehouse manager.

Usable data were collected from 400 organizations.<sup>10</sup> The organizations that participated in the study had mean gross revenues of \$660 million and a mean of 33,750 employees. Organizations from every region of the United States and from a variety of industries were represented. Over half of the respondents were IS/data warehouse managers or staff members and 68% of them were actively involved in the architecture selection decision.

The distribution of the four data warehouse architectures was: Independent data mart architecture (IDM) – 44 (11%), data mart bus architecture (DBA) – 104 (26%), enterprise data warehouse architecture (EDW) – 235 (58%), and Federated architecture (FED) – 17 (5%). This distribution is consistent with other data warehouse surveys (e.g., [22]).

The sample size for the FED architecture was insufficient to conduct multinomial logistic regression. Consequently, FED was dropped from the data analysis. Data analyses were conducted on the remaining 383 usable data points. The means and standard deviations (SD) of the independent variables across the various architectures are shown in Table 1.

**Table 1**

Descriptive statistics for contextual variables for each architecture.

	For IDM		For DBA		For EDW	
	Mean	SD	Mean	SD	Mean	SD
Interdependence	4.30	1.51	4.97	1.44	4.87	1.36
Strategic view	4.19	1.72	5.04	1.43	5.35	1.28
Urgency	4.73	1.67	5.14	1.43	4.91	1.41
Resource constraints	4.67	1.41	4.16	1.57	4.27	1.46
Task	4.71	1.41	5.16	1.28	5.20	1.25
Perceived ability	3.36	1.55	3.99	1.75	4.23	1.57
Sponsorship level	2.43	1.23	2.59	1.36	2.85	1.30

#### 4.5. Measurement model

In order to assess the measurement model, exploratory factor analysis (EFA) was conducted to assess the unidimensionality of items and constructs using principal components analysis with a promax rotation.<sup>11</sup> The EFA included all the items with the number of factors specified to be the number of latent constructs. The initial factor matrix indicated that all items loaded above 0.60 [17] on the intended factor except for one item from strategic view of warehouse. This item was dropped and a confirmatory factor analysis (CFA) was conducted to examine the measurement model. The CFA model fit was assessed in terms of four measures identified by authorities: root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), Tucker Lewis index (TLI or NNFI), and the adjusted goodness-of-fit index (AGFI) [30,80]. The recommended target values for each of these fit statistics and the corresponding values for the study measurement model are presented in Table 2.

In addition to model fit, the reliability, convergent validity, and discriminant validity was assessed. As indicated in Table 3, all estimated standard loadings were significant ( $p < 0.05$ ), suggesting good convergent validity. Also, the average variance extracted (AVE) for each construct exceeded 0.50, demonstrating convergent validity [27]. Finally, the square root of AVE was examined to see if they were larger than the correlation coefficient to assess discriminant validity [27]. All the correlations exceeded the recommended criterion, indicating discriminant validity.

Cronbach alpha values were used to assess the reliability of the items associated with the latent constructs. According to Gefen et al. [30], Cronbach alphas for constructs in exploratory studies should exceed 0.60, and in confirmatory studies they should exceed 0.70. Other authorities suggest that a reliability of at least 0.6 suffices for early stages of research [64]. The internal consistency of constructs in this study surpassed the suggested guidelines for exploratory studies.

### 5. Analysis and results

The data analyses in this study involved three major steps: (1) testing the overall research model using stepwise multinomial logistic regression (MLR)<sup>12</sup> (2) testing the hypotheses using binomial logistic regression (BLR), and (3) additional analysis to explore differences between MLR and BLR results.

<sup>11</sup> The level of sponsorship was not included in the assessment of the measurement model because it was a single item factor.

<sup>12</sup> 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 was undertaken to identify which factors influence the selection between two architectures choices using binomial logistic regression (BLR): (1) 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) and (2) analyses to identify which selection factors influence the selection of one architecture when compared to another (e.g., EDW versus IDM). Both steps mostly corroborated the results obtained through the multinomial logistic regression which is more powerful than BLR.

<sup>9</sup> The individuals/organizations that promoted the survey included TDWI, Bill Inmon, Kimball Group, DM Review, Teradata, Business Objects, MicroStrategy, DAMA International, Claudia Imhoff, and William McKnight.

<sup>10</sup> There were additional respondents who did not complete all of the questions and their partial responses were not included in the data analysis. There were also 27 other organizations that were not included in the data analysis because they indicated that their architecture was less than a 4 on the 1–7 scale that asked how closely their architecture matched one of the four reference data warehouse architectures.

**Table 2**  
The goodness-of-fit statistics for architecture selection measurement model.

Measure	Target value	Architecture selection
RMSEA	<0.08	0.043
SRMR	<0.10	0.042
NNFI	>0.90	0.96
AGFI	>0.80	0.92

### 5.1. Testing the overall research model

Multinomial logistic regression (MLR) was used to test the research model in Fig. 5 in order to identify and understand the impact of independent variables [53]. In presenting guidelines for logistic regression, Hosmer and Lemeshow [45] suggest the use of the stepwise procedure when the phenomenon being studied is relatively new, the important covariates may not be known, and associations with the outcome are not well understood. Based on Lee and Koval [56], the forward Wald method using an alpha level of 0.15 for entry was applied to conduct the stepwise analyses.<sup>13</sup>

The research model was tested using stepwise multinomial logistic regression with the seven contextual factors as independent variables<sup>14</sup> and the data warehouse architecture as the dependent variable. The EDW architecture was treated as the reference category (i.e., the category against which the other categories are compared). The main model was analyzed using the likelihood ratio test and model fitting information, including chi-square and *R*-squared values.

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 a 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 indicating a small effect size despite the significance of the overall model. According to Hosmer and Lemeshow [45] 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 results of the stepwise MLR process are presented in Table 4. The model results show the selection variable parameter estimates and odds ratios for the IDM and DBA architectures with EDW as the reference architecture. The table reveals that strategic view, resource constraints, and perceived ability influence the selection of a data warehouse architecture when comparing the IDM and EDW architectures. Also, a 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 architecture when compared to selecting an EDW architecture decreased by 19.5% as strategic view increased by 1 unit. The analysis indicated that perceived ability and resource constraints did not influence the chance of selecting a DBA when compared to the EDW architecture.

<sup>13</sup> 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 and Koval [57] 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 0.15 for entry was applied to conduct the stepwise analyses.

<sup>14</sup> The multiple items that survived the measurement model assessment were averaged to create each selection variable.

**Table 3**  
The measurement model statistics.

Construct (reliability)	ITEM	Loadings	t-Value
Information interdependence (0.8625)	INT1	0.81	
	INT2	0.85	16.95
	INT3	0.81	16.47
Strategic view (0.6326)	VIEW2	0.54	
	VIEW3	0.86	7.00
Urgency (0.9049)	URG1	0.85	
	URG2	0.92	21.88
	URG3	0.86	20.56
Resource constraints (0.7536)	RES1	0.73	
	RES2	0.81	10.61
	RES3	0.61	10.01
Task routineness (0.7781)	TSK1	0.60	
	TSK2	0.84	10.81
	TSK3	0.78	10.85
Perceived ability (0.8492)	SKL1	0.88	
	SKL2	0.76	15.53
	SKL3	0.78	15.80

The multinomial model was also assessed for its discriminating power. Assessing a multinomial model's predictive power is problematic when most of the responses are in a single category [53]. More specifically, for the choice of data warehouse architectures, it is difficult to create a model that outperforms using the dominant category (EDW was used 61% of the time) as the prediction for every case. In response, a two-step assessment approach was used.

First, the model's predictive power was tested against a "naïve model." With a naïve model, every response category (i.e., the various architectures) is assumed to be equally likely. Applying the naïve model, one of the three architectures is randomly selected, with any of the three architectures as likely to be predicted as the others. With this approach, the expected value is that the prediction is correct 33.3% of the time. For comparison purposes, when the multinomial model was used, it predicted the correct architecture 62.8% of the time.

The second approach was to compare the multinomial model's predictions for IDM and DBA against the percentages of responses in those architecture categories. In other words, can the model predict which architecture would be used when the largest category is removed? There were 104 DBA and 44 IDM responses for these architectures, and the model predicted the correct architecture 76.8% of the time.

**Table 4**  
MLR output: parameter estimates.

Architecture category <sup>a</sup>	Variables	Beta	Odds ratio [Exp(Beta)]	Wald	Degrees of freedom	Significance
IDM	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	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

<sup>a</sup> The reference architecture category is: EDW.

\*\* =  $p < 0.05$ .



Statistical power analysis was also carried out in order to identify if sample size may affect prediction. The power analysis indicated that when conducting multinomial logistic regression, in order to detect a medium effect with 0.80 power, a total sample size of 1500 responses would be required [46]. Therefore, it is possible that low predictive power is due to the study's sample size.

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 [53], when the goal of the model is to identify and understand the impact of independent variables rather than prediction, the predictive accuracy of a model is less of a concern.

## 5.2. Hypotheses testing

Each of the specific hypotheses, except for those that referred to the FED architecture,<sup>15</sup> was tested using separate binary logistic regression analyses. The results of the hypotheses testing are presented in Table 5. As illustrated by a strong support for H1, H4, H5A, H6A, and H7A, when considering the choice between IDM and EDW architectures, information interdependence, strategic view, resource constraints, perceived ability of the in-house IT staff, and level of sponsorship each influence architecture selection. Several hypotheses were not supported: H2A, H2B, H5B, H6B, and H7B. As multiple tests were being conducted on the same sample, in order to control for family-wise and comparison-wise errors the *p*-values were adjusted for multiple testing using the Benjamini et al. [19] procedure. As the adjusted significance levels for the hypotheses tests indicate (see Table 5), the previous hypotheses testing results were significant except for H7A. As the initial hypothesis test results indicate, H7A was marginally significant at a *p*-value of 0.049. It is possible that low sample size may have affected the testing of the influence of sponsorship on the selection between IDM and EDW.

## 5.3. Additional analysis

Contrary to strong evidence from the experts and literature, the MLR data analyses suggests that information interdependence (II), task routineness (TR) and level of sponsorship (SP) do not influence architecture selection. In addition, the BLR analysis, conducted to test hypothesis H1, suggests that II directly influences the choice between IDM and EDW architecture, while the stepwise MLR analyses conclude otherwise. These contrary findings led to further data analyses. Additional investigation suggests that when considering the choice between IDM and EDW or IDM and DBA architectures, strategic view may mediate the influence of II, TR, and SP<sup>16</sup> on data warehouse architecture selection. In order to test if these factors influence strategic view, structural equation modeling was used. The fit statistics for the model were all higher than the recommended thresholds. The model indicated that all the factors have a significant, positive influence on strategic view, explaining 39% of the variance.<sup>17</sup>

<sup>15</sup> Because FED was omitted from the data set due to the small sample size (i.e., 17 data points), H3 could not be tested.

<sup>16</sup> Even though the adjusted hypothesis H7A, which captures the influence of sponsorship on IDM compared to EDW, was not significant at the 0.05 level, we chose to include it in this additional analysis. The rationale is that it is significant at the 0.1 level, and there is strong support in the literature for the inclusion of sponsorship in the selection of a data warehouse architecture. We also believe that had our sample size been larger, sponsorship would have emerged as a significant architecture selection factor.

<sup>17</sup> Each of the remaining four contextual variables was also further investigated to explore if other strong mediation effects exists. The results suggest that the most powerful mediation effects are displayed between strategic view and II, TR, and SP.

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 and Kenny [7], were conducted to assess whether strategic view mediates the influence of II on selection of an IDM versus EDW (see Fig. 6). The first three causal steps of the mediation test (i.e., steps a, b, and c in Fig. 6) for II provided significant path coefficients. The fourth step, the path from II to architecture selection while controlling for strategic view, was zero, suggesting that the view of the warehouse fully mediates the relationship. Similar findings were obtained through the mediation tests for the influence of TR and SP. The mediation tests suggest that increases in information interdependence, task non-routineness, and sponsorship level 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. Mediation tests confirmed that the influence of the three variables on the selection of an IDM when compared to a DBA is also mediated by the strategic view of the warehouse. Similar tests of mediation suggest that strategic view does not mediate the influence of II, TR, and SP on the choice between EDW and DBA architectures.

## 6. Discussion

Based on the hypotheses testing, the multinomial logistic regression analyses used to explore the full research model, and further analyses employed to test mediation effects, some factors appear to be more important than others in the selection of a particular architecture. The overall findings suggest that different combinations of specific factors affect the likelihood of selecting one architecture over the others.

Three selection factors emerged as being important in choosing between the IDM and EDW architectures. They are the perceived ability of the IT staff, resource constraints, and the strategic view of the warehouse. In addition, further analysis revealed that interdependence, task routineness, and the level of sponsorship influence the selection of an EDW through strategic view. When interdependence is high, task routineness is low, and the sponsorship level is high, organizations perceive the implementation of a data warehouse as a strategic initiative.

Comments from an expert helped put these findings into perspective. He said, "The implementation of an enterprise data warehouse is all about the business, rather than a functional area. If the warehouse is perceived as strategic, the executives are behind it. And if the executives are behind it, the resources will be available. If the executive support and resources aren't there, you are likely to end up with independent data marts." He also commented on the perceived ability of the IT staff, "If the IT staff lacks confidence in building an enterprise data warehouse, it may 'down scope' and 'dumb down' the requirements to the point of only providing simple reports and meeting departmental objectives."

Strategic view appears to be significant when choosing between the DBA and EDW architectures. When the implementation of a data warehouse is perceived as a strategic initiative, organizations are more likely to implement an EDW than a DBA architecture. Against expectations, resources, the perceived ability of the IT staff, and urgency were not significant in determining which architecture is selected.<sup>18</sup>

<sup>18</sup> Urgency was significant at the 0.1 level.

**Table 5**  
Results of hypothesis testing.

Hypothesis	Beta	Significance	Adjusted Significance
H1: Organizations with high interdependence are more likely to select an EDW than an IDM architecture.	−0.280**	0.014	0.035
H2A: Organizations with high urgency are more likely to select an IDM than an EDW architecture.	−0.027	0.818	0.818
H2B: Organizations with high urgency are more likely to select a DBA than an EDW architecture	0.152	0.088	0.126
H4: 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 than an EDW architecture.	−0.449**	0.001	0.010
H5A: Organizations with low resource availability are more likely to select an IDM than an EDW architecture.	0.370**	0.011	0.037
H5B: Organizations with low resource availability are more likely to select a DBA than an EDW architecture.	−0.102	0.228	0.285
H6A: Organizations that have an IT staff with low perceived ability are more likely to select an IDM than an EDW architecture.	−0.331**	0.004	0.020
H6B: Organizations that have an IT staff with low perceived ability are more likely to select a DBA than an EDW architecture.	−0.083	0.276	0.307
H7A: Organizations with upper management sponsorship are more likely to select an EDW than an IDM architecture.	−0.265	0.049	0.098
H7B: Organizations with upper management sponsorship are more likely to select an EDW than a DBA architecture.	−0.184	0.059	0.098

\*\* =  $p < 0.05$ .

The combined observations from several experts and a data warehouse manager help to understand these findings. One expert observed, “To do an enterprise data warehouse, you have to step back and take an enterprise view. The bus architecture doesn’t require this. And people equate enterprise with strategic.” Another expert commented, “An enterprise data warehouse is more expensive. However, because it is also more strategic, the resources will be available.”

Even though enterprise data warehouses have many pieces (e.g., the central repository and dependent data marts) and require considerable up-front planning (e.g., enterprise data model), the skills required to do this may not be perceived as being more daunting than those needed to implement a DBA. Most IT professionals are familiar with the methodologies and data modeling approaches associated with EDW and may feel up to the challenge. To illustrate, one data warehouse manager who selected the EDW architecture said that she was strongly influenced by her college professor’s teaching of the merits of E–R modeling techniques and normalization principles (both strongly associated with EDW). She said, “In building the warehouse, I honored sound data design methodology. All of the data is maintained in a centralized architecture in 3rd normal form and users are given logical views of the data.” For this data warehouse manager, she was knowledgeable and comfortable with the skills need to build an EDW. An expert commented, “The skills needed to build either a bus or a hub and spoke architecture are much more common than they used to be,” further suggesting why perceived ability is not an important consideration when choosing between the DBA and EDW architectures.

Finally, three factors surfaced as being important to selecting between a DBA and an IDM. They are perceived ability of the IT staff, strategic view, and resource constraints. 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.<sup>19</sup>

Therefore, overall, strategic view, the perceived ability of the IT staff, and resource availability are the factors with the greatest predictive power when selecting a particular data warehouse architecture. Based on case studies of 50 organizations, Broadbent

and Weill [11] suggest that when IT is viewed as being strategic, organizations invest in extensive firm-wide IT infrastructure services in a centralized way. The results of this study corroborate Broadbent and Weill’s [11] findings. Furthermore, the study revealed that strategic view mediates the influence of information interdependence, task routineness, and sponsorship level on architecture choice.

The data analyses showed that urgency has only limited impact on the selection of a data warehouse architecture. Somewhat surprising, since IDMs are sometimes described as a “fast and dirty” solution, urgency did not have a significant influence on the selection of the IDM architecture. The history of IDM implementations suggests a possible explanation for this finding. For many companies, IDMs were their initial attempts to provide decision support data [23]. Over time, organizations have recognized that while an IDM might be an expedient, point solution, it does not provide a robust decision support environment. Consequently, though there may be some IDM architecture implementations that were driven by urgency for a quick solution, most implementations in organizations today may have been created because they were the prevalent architecture at the time they were developed.

In our follow up interviews, one expert made an observation that supports the legacy view of IDM. She said that, “No matter what reference architecture is primarily used, there are always some independent data marts, because they have been in place for a long time and their sponsors have the political clout to ensure their existence.”

The influence of urgency on the likelihood of selecting a DBA architecture over EDW was only marginally significant (i.e., at  $p < 0.1$ ). This result confirms the majority opinion of authors in the data warehousing literature (e.g., [41,50]) and the opinion of experts that were interviewed that the DBA architecture can be implemented quickly. The finding also provides support for the OIPT theory [78] by showing that as urgency increases, organizations are more likely to implement an architecture that, at least initially, provides lower information processing capacity. For example, when market pressure creates an immediate need for a data warehouse, a DBA architecture can be implemented more quickly and provide an initial data mart solution, but with lower information processing capacity compared to an EDW architecture.

As expected, many factors that were based on rational OIPT theory emerged as being important to architecture selection. One of the social/political factors, sponsorship level, was also identified as impacting architecture selection by influencing the strategic view of the warehouse, which in turn influences architecture choice. These findings contribute to better understanding the broader topic of IT infrastructure design.

<sup>19</sup> 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.

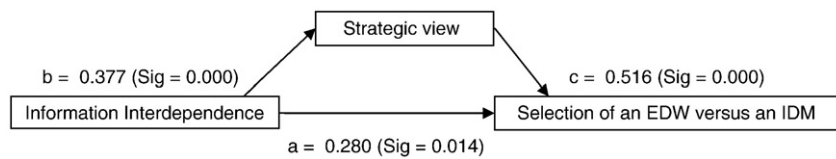


Fig. 6. Mediation test.

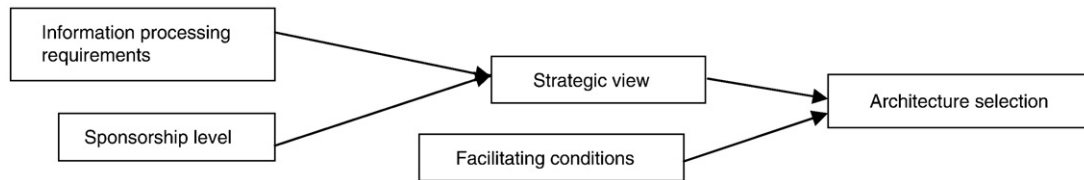


Fig. 7. An integrated model for data warehouse architecture selection.

## 7. Conclusion

This study examined the factors that affect the selection of a data warehouse architecture by using a research model that was developed using rational and social/political theories, contingency and adoption literature, data warehousing literature, and expert interviews. Contextual variables that affect architecture choice were identified to discover the relative influence of each factor on data warehouse architecture selection.

There are several academic contributions from this research. It extends the existing knowledge about IT infrastructure decision making by identifying the importance and the influence of organizational factors on data warehouse architecture selection. Broadbent and Weill [11] describe how organizations choose IT infrastructure options based on their view of IT infrastructure, which is founded on business needs. It provides some empirical support for these findings that an organization's view of infrastructure influences the infrastructure capability attained. Further, this work applies Weill and Broadbent [87] 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 sponsorship level influences the view of an IT infrastructure. Thus, this work enriches the existing sparse research on IT infrastructure and IT infrastructure design.

There is limited previous research on IT infrastructure design partly due to the difficulty of defining and measuring IT infrastructure design. This research 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. It also offers empirical support for new variables, such as strategic view.

The world of practice also benefits from this research by shedding light on an important question in the data warehouse industry. Companies are currently spending millions of dollars implementing data warehouse architectures. The findings identified the most salient factors that appear to influence the selection decision and provide practitioners with a basis to evaluate the current organizational circumstances within their firms. The factors 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. Consequently, practitioners are provided with some guidance on how to approach the data warehouse architecture selection decision.

The study also suggests numerous avenues for future research. One such opportunity is described next. The selection factors that emerged as most salient suggest the creation of one overall model for architecture selection. 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 [79]. Along with the sponsorship level, information processing needs influence the creation of a strategic view of the warehouse. Next, resource constraints, the perceived ability of the IT staff, and urgency can be described as the facilitating conditions<sup>20</sup> 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 (Fig. 7). Additional long-term research is needed to develop a refined understanding of the relationships proposed in the integrated model. One short-term future research possibility is to investigate the data in other interesting ways. The data analysis method used in this study, multinomial logistic regression, is a standard statistical technique that helps identify the influence of selection factors on data warehouse architectures. It is possible to analyze the data in other interesting and useful ways such as using three class ROC analysis [43].

When thinking about data warehouse architectures, it is important to recognize that the reference architectures provide only guidelines or blueprints to follow. For a variety of reasons – historical, political, business need – variations on the reference architectures is the norm. As one expert said, “Nothing is pristine in terms of architecture.”

It is also worth noting that changes in the technologies that are available can affect architecture design decisions. For example, recent advances in enterprise information integration (EII) software make it increasingly feasible to leave data in source systems, run queries against these distributed data sources, and integrate the data “on the fly.” With this approach, a company might integrate most of its data in a data warehouse, but leave other data where it is, either as a short or long-term solution.

<sup>20</sup> In individual level theory, Triandis [78] 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 constraints, 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, low resource constraints, high perceived ability, and low urgency would result in the total assessment of facilitating conditions as assisting the implementation of a data warehouse infrastructure initiative.

## Appendix A

Summary of theory, literature, and interview comments leading to the development of the research model.

Factor	Description	Theory	Literature	Sample interview comments
1. Information interdependence	The extent to which tasks and their outcomes were contingent upon information from one or more other organizational units.	Rational/information processing (e.g., Galbraith [29], Goodhue et al. [35], Thompson [76], and Tushman and Nadler [79])	Andres and Zmud [3], Armstrong [5], Devlin [20], and Inmon et al. [47].	The need to share information across units within the entire enterprise often leads organizations to select an EDW architecture compared to IDM architecture solutions.
2. Urgency	The extent to which there was an urgent need to build the data warehouse.	Rational/information processing (e.g., Galbraith [29] and Tushman and Nadler [79])	Allen and Andrew [2], Griffin [37], Grover and Segars [38], Mimno [64], and Rockart et al. [67].	The need to build a data warehouse solution to satisfy an immediate <i>business pain</i> leads most organizations to select an IDM architecture compared to an EDW architecture solution.
3. Task routineness	The extent to which users' jobs required non-routine data analyses.	Rational/information processing (e.g., Galbraith [29])	Specht [70], Anandarajan and Arinze [4], Goodhue [32], and Campbell [13].	The need to accomplish complex, ad hoc data analysis may require the implementation of an EDW architecture solution. When task requirements are routine and predictable, organizations often move towards less complex architecture solutions than an EDW architecture.
4. Strategic view	The extent to which implementing a data warehouse was viewed as being important to supporting strategic initiatives.	Rational/information processing (e.g., Tushman and Nadler [79])	McFarlan and McKenney [62], Duncan [21], Broadbent and Weill [11], Weill and Broadbent [87], Chen and Sheldon [15], Griffin [37], and Watson et al. [86].	An organization may see the data warehouse as a way of achieving organizational strategies, and might implement a robust data warehouse architecture solution such as an EDW architecture.
5. Resource constraints	The extent to which IT personnel, business unit personnel, and monetary resources were unavailable for building the data	Rational/information processing (e.g., Galbraith [29], Tushman and Nadler [79])	Miller and Friesen [63], Ein-dor and Segev [24], Tait and Vessey [75], Armstrong [5], Kimball et al. [52], Mimno [64], and Hackney [41].	Limited resource availability often leads organizations to develop data mart architecture solutions to satisfy their information needs.
6. Perceived ability of the IT staff	The extent of the perceived ability of the IT staff in terms of technical skills, successful experiences and confidence in developing data warehouses.	Rational/information processing (e.g., Galbraith [29] and Goodhue et al. [35])	Ravichandran [66], Fichman and Kemerer [26], Agarwal et al. [1], Marakas et al. [58], Lindsley and Brass [57], Gibson [31], and Breslin [10].	Specialized IT skills and successful past experiences implementing data warehouses would influence a development team to choose an EDW architecture.
7. Sponsorship level	Level of sponsorship for the data warehousing initiative.	Social political (e.g., Pfeffer [65], Bradshaw-Camball and Murray [9], and Jasperson et al. [48])	Tractinsky and Jarvenpaa [77], Watson et al. [85], and Watson et al. [86].	
Hackney [41].	The sponsor may use political influence to affect the data warehouse implemented. When sponsorship comes from a particular department or functional area, the sponsor may select an architecture that specifically accommodates that department or functional area (e.g., an IDM architecture).			

## References

- [1] R. Agarwal, V. Sambamurthy, R. Stair, The evolving relationship between general and specific computer self-efficacy, *Information Systems Research* 11 (4) (2000).
- [2] B. Allen, C. Andrew, Information architecture: in search of efficient flexibility, *MIS Quarterly* 15 (4) (1991).
- [3] M. Anandarajan, B. Arinze, Matching client/server processing architectures with information processing requirements, *Information & Management* 34 (5) (1998).
- [4] H.P. Andres, R. Zmud, A contingency approach to software project coordination, *Journal of MIS* 18 (3) (2001).
- [5] R. Armstrong, Data warehousing, The fallacy of data mart strategies, NCR, 2003.
- [6] E. Babbie, *The Practice of Social Research*, Thompson Learning, Belmont, CA, 2000.
- [7] R. Baron, D. Kenny, The moderator–mediator variable distinction in social psychological research, *Journal of Personality & Social Psychology* 51 (1986).
- [8] I. Benbasat, R. Taylor, The impact of cognitive styles on information systems design, *MIS Quarterly* 2 (2) (1978).
- [9] P. Bradshaw-Camball, V.V. Murray, Illusions and other games: a trifocal view of organizational politics, *Organization Science* 2 (4) (1991).
- [10] M. Breslin, Data warehousing battle of giants: comparing the basics of the Kimball and Inmon models, *Business Intelligence Journal* 9 (4) (2004).
- [11] M.P. Broadbent, P. Weill, Management by maxim: how business and IT managers can create IT infrastructures, *Sloan Management Review* 38 (3) (1997).
- [12] M.K. Brohman, Explaining Variation in Data Warehouse Usage, Richard Ivey School of Business, The University of Western Ontario, London, Ontario, 2000.
- [13] D. Campbell, Task complexity: a review and analysis, *Academy of Management Review* 13 (1988).
- [14] A. Cavaye, J. Christiansen, Understanding IS implementation by estimating power of subunits, *European Journal of Information Systems* 4 (1996).
- [15] L. Chen, M. Gillenson, A Framework for Data Warehousing Research, Americas Conference on Information Systems, Boston, 2000.
- [16] H. Chen, P. Sheldon, Destination information systems: design issues and directions, *Journal of Management Information Systems* 14 (2) (1997).
- [17] A.L. Comrey, H. Lee, *A First Course in Factor Analysis*, Erlbaum, New Jersey, 1992.
- [18] B. Cooper, H.J. Watson, B. Wixom, D.L. Goodhue, Data warehousing supports strategy at First American Corporation, *MIS Quarterly* 24 (4) (2000).
- [19] F. Damanpour, Organizational innovation: a meta analysis of effects of determinants and moderators, *Academy of Management Journal* 34 (3) (1991).
- [20] B.A. Devlin, *Data Warehouse: From Architecture to Implementation*, Addison Wesley, Reading, Massachusetts, 1997.
- [21] N. Duncan, Capturing flexibility of information technology infrastructure, *Journal of Management Information Systems* 12 (2) (1995).
- [22] W. Eckerson, Smart companies in the 21st century, TDWI Research Report, 2003.
- [23] W. Eckerson, Gauge your data warehouse maturity, *DM Review* 14 (11) (2004).
- [24] P. Ein-Dor, E. Segev, Organizational context and the success of management information systems, *Management Science* 24 (1978).
- [25] K. Eisenhardt, L. Bourgeois, Politics of strategic decision-making in high-velocity environments, *Academy of Management Journal* 31 (4) (1988).
- [26] R.G. Fichman, C. Kemerer, The assimilation of software process innovations: an organizational learning perspective, *Management Science* 43 (10) (1997).
- [27] C. Fornell, D. Larcker, Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research* 18 (1981).
- [28] C. Francalanci, V. Piuri, Designing information technology architectures: a cost-oriented methodology, *Journal of Information Technology* 14 (4) (1991).
- [29] J. Galbraith, Organizational design: an information processing view, *Interface* 4 (3) (1974).
- [30] D. Gefen, D. Straub, M. Boudreau, Structural equation modeling and regression: guidelines for research practice, *CAIS* 4 (7) (2000).
- [31] C. Gibson, Do they do what they believe they can? Group efficacy and group effectiveness across tasks, *Academy of Management Journal* 42 (2) (1999).
- [32] D. Goodhue, R. Thompson, Task-technology fit and individual-performance, *MIS Quarterly* 19 (2) (1995).



- [33] D. Goodhue, J. Quillard, J. Rockart, Managing the data resource – a contingency perspective, *MIS Quarterly* 12 (3) (1988).
- [34] D. Goodhue, M. Wybo, Kirsch, L., The impact of data integration on the costs and benefits of information-systems, *MIS Quarterly* 16 (3) (1992).
- [35] D. Goodhue, B. Wixom, H.J. Watson, Realizing business benefits through CRM: hitting the target in the right way, *MIS Quarterly Executive* 1 (2) (2002).
- [36] P. Gray, H.J. Watson, *Decision Support in the Data Warehouse*, Prentice-Hall, New Jersey, 1998.
- [37] J. Griffin, Information strategy: mart vs. data warehouse, *DM Review* 12 (3) (2001).
- [38] V. Grover, A. Segars, The relationship between organizational characteristics and information system structure, *International Journal of Information Management* 16 (1996).
- [39] S. Gully, K. Incalcaterra, A. Joshi, J. Beaubien, A meta-analysis of team-efficacy and performance, *Journal of Applied Psychology* 87 (5) (2002).
- [40] D. Hackney, Architecture of architectures: warehouse delivery, *DM Review* 12 (4) (1998).
- [41] D. Hackney, Architecture anarchy and how to survive it: God save the queen, *Enterprise Systems Journal* 15 (4) (2000).
- [42] B. Haley, Implementing the Decision Support Infrastructure: Key Success Factors in Data Warehousing, in: Management, University of Georgia, Athens, 1988.
- [43] X. He, C. Frey, Three class ROC analysis: a decision theoretic approach under the ideal observer framework, *IEEE Transactions on Medical Imaging* 25 (2006).
- [44] T. Hinkin, A review of scale development practices in the study of organizations, *Journal of Management* 21 (5) (1995).
- [45] D. Hosmer, S. Lemeshow, *Applied Logistic Regression*, Wiley, New York, 2000.
- [46] F.Y. Hsieh, D.A. Block, M.D. Larsen, Simple method of sample size calculation for linear and logistic regression, *Statistics in Medicine* 17 (1998).
- [47] V. Inmon, C. Imhoff, R. Sousa, *Corporate Information Factory*, Wiley, New York, 2001.
- [48] S. Jaspersen, T. Carte, C. Saunders, B. Butler, H. Croes, W. Zheng, Review: power and IT research, *MIS Quarterly* 26 (4) (2002).
- [49] R. Jindal, A. Acharya, Federated data warehouse architecture, White Paper for ITtoolbox Data Warehouse, 2008 [<http://research.ittoolbox.com/white-papers/datamgt/dw/federated-data-warehouse-architecture-3309> accessed 5/14/2008].
- [50] K. Joshi, M. Curtis, Issues in building a successful data warehouse. *Information Strategy, The Executive's Journal* 4 (1) (1999).
- [51] S. Kelly, Data marts: the latest silver bullet, *DM Review* 7 (2) (1997).
- [52] R. Kimball, L. Reeves, M. Ross, W. Thornthwaite, The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses, Wiley, New York, 2003.
- [53] D. Kleinbaum, *Logistic Regression: A Self Learning Text*, Springer-Verlag, New York, 1994.
- [54] D. Laney, Warehouse factors to address for success, *HP Professional* 14 (5) (2000).
- [55] A. Langley, H. Mintzberg, P. Pitcher, E. Posada, J. Saintmacary, Opening up decision making: the view from the stool, *Organization Science* 6 (3) (1995).
- [56] K. Lee, J. Koval, Determination of the best significance level in forward stepwise logistic regression, *Communication in Statistics*, B 26 (1997).
- [57] D. Lindsley, D. Brass, Efficacy-performance spirals: a multilevel perspective, *Academy of Management Review* 20 (3) (1995).
- [58] M. Mannino, Z. Walter, A framework for data warehouse refresh policies, *Decision Support Systems* 42 (2006).
- [59] G.M. Marakas, M. Yi, R. Johnson, Information Systems Research, The multilevel and multifaceted character of computer self-efficacy, *Information Systems Research* 9 (2) (1998).
- [60] M.L. Markus, Power, politics and MIS implementation, *Communications of the ACM* 26 (6) (1983).
- [61] C. Matheson, Rationality and decision-making in Australian federal government, *Australian Journal of Political Science* 33 (1) (1998).
- [62] W. McFarlan, J. McKenney, Corporate information systems management: the issues facing senior executives, Homewood, Irwin, 1983.
- [63] D. Miller, P. Friesen, Innovation in conservative and entrepreneurial firms: two models of strategic momentum, *Strategic Management Journal* 3 (1) (1983).
- [64] P. Mimmo, How to avoid mart chaos using hybrid methodology, TDWI Flash point, 2002.
- [65] J. Pfeffer, *Power in Organizations*, Pitman Publishing, Marshfield, 1981.
- [66] T. Ravichandran, Innovation Assimilation in the Presence of Knowledge Barriers and Technology Uncertainty, Academy of Management, Washington DC, 2001.
- [67] J.E. Rockart, M. Earl, J. Ross, Eight imperatives for the new IT organization, *Sloan Management Review* 38 (1) (1996).
- [68] A. Sen, S. Atish, Toward developing data warehousing process standards: an ontology-based review of existing methodologies, *IEEE Transactions on Systems, Man and Cybernetics* 37 (1) (2007).
- [69] V. Sethi, W. King, Construct measurement in information systems research: an illustration in strategic systems, *Decision Sciences* 22 (3) (1991).
- [70] P.H. Specht, Job characteristics as indicants of CBIS data requirements, *MIS Quarterly* 10 (3) (1986).
- [71] R.H. Sprague, E.D. Carlson, *Building Effective Decision Support Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1982.
- [72] K.F. Strange, Making BI strategic, The key issues, IE-19-4691, Gartner Group, 2003.
- [73] D. Straub, M. Boudreau, D. Gefen, Validation guidelines for IS positivist research, *Communications of the AIS* 13 (24) (2004).
- [74] E.B. Swanson, Information systems innovation among organizations, *Management Science* 40 (9) (1991).
- [75] P. Tait, I. Vessey, The effect of user involvement on system success: a contingency approach, *MIS Quarterly* 12 (1) (1988).
- [76] J.D. Thompson, *Organizations in Action*, McGraw-Hill, 1967.
- [77] N. Tractinsky, S. Jarvenpaa, Information systems design decisions in a global versus domestic context, *MIS Quarterly* 19 (4) (1995).
- [78] H. Triandis, Values, attitudes, and interpersonal behavior, Nebraska Symposium on Motivation: Beliefs, Attitude, and Values, University of Nebraska Press, 1979.
- [79] M.L. Tushman, D. Nadler, Information processing as an integrating concept in organizational design, *Academy of Management Review* 3 (3) (1978).
- [80] E.F. Vail, Causal architecture: bringing the Zachman framework to life, *Information Systems Management* 19 (3) (2002).
- [81] L. Volonino, H.J. Watson, S. Robinson, Using EIS to respond to dynamic business conditions, *Decision Support Systems* 14 (1995).
- [82] T. Walter, Teradata's real-time enterprise reference architecture: a blueprint for the future of IT, *Teradata Magazine* 5 (1) (2005).
- [83] H.J. Watson, Real time: the next generation of decision-support data management, *Business Intelligence Journal* 10 (3) (2005).
- [84] H.J. Watson, R. Watson, S. Singh, D. Holmes, Development practices for EIS: findings of a field-study, *Decision Support Systems* 14 (2) (1995).
- [85] H.J. Watson, B. Wixom, J. Buonomi, Creating intelligence across the supply chain, *Communications of the AIS* 5 (9) (2001).
- [86] H.J. Watson, D.L. Goodhue, B. Wixom, The benefits of warehousing: why some firms realize exceptional payoffs, *Information & Management* 39 (6) (2002).
- [87] P. Weill, M. Broadbent, *Leveraging the New Infrastructure*, Harvard Business School, Boston, 1998.
- [88] L. Willcocks, R. Sykes, Enterprise resource planning: the role of the CIO and IT function in ERP, *Communications of the ACM* 43 (4) (2000).
- [89] M.D. Wybo, D.L. Goodhue, Using interdependence as a predictor of data standards: theoretical and measurement issues, *Information & Management* 29 (6) (1995).
- [90] R. Zmud, A. Boynton, Survey measures and instruments in MIS, in: K.L. Kraemer (Ed.), *The Information Systems Research Challenge*, Harvard Business School Press, Cambridge, 1991.

**Thilini Ariyachandra** is an Assistant Professor of MIS in the Williams College of Business at Xavier University in Cincinnati. Her research is focused on the selection, design, and implementation of decision support systems in organizations. She has published in journals such as *Information Systems Management*, *Business Intelligence Journal*, *Communications of the ACM* and *Decision Support Systems* and in national and international conference proceedings.

**Hugh J. Watson** is a Professor of MIS and a holder of a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia. Hugh is a leading scholar and authority on decision support, having authored 22 books and over 100 scholarly journal articles. Hugh is a Fellow of the Association for Information Systems and The Data Warehousing Institute and is the Senior Editor of the *Business Intelligence Journal*. He is also the Senior Director of the Teradata University Network, a free portal for faculty who teach and research data warehousing, BI/DSS, and database.