Journal of Management and Educational Studies (2024) 2(1), 49-57 *ISSN: 3041-9263*

Contents lists available at Journal website

Journal of Management and Educational Studies

Designing and Implementing A Data Warehouse in Technology Companies

Behrouz Dolatshah [a], Ebrahim Nazari Farrokhi [b], Khatereh Qorbani [c]

[a] *Department of IT Management, Science and Research Branch, Islamic Azad university, Tehran, Iran*

[b] *Department of IT Management, Science and Research Branch, Islamic Azad university, Tehran, Iran*

[c] *Department of IT Management, Science and Research Branch, Islamic Azad university, Tehran, Iran (khatereghorbanii@yahoo.com)*

Article info

Article history:

Received 24 Oct 2023

Received in revised form 30 Nov 2023

Accepted 30 Dec 2023

Available online 10 May 2024

Keywords:

Data Warehouse,

Technology Companies,

Business Conditions,

Technology Utilized,

Data Type.

Abstract

A data warehouse is a comprehensive system utilized within enterprises for the examination and presentation of organized and partially organized data from various origins, including point-of-sale transactions, marketing automation, customer relationship management, and others. It is well-suited for both spontaneous analysis and customized reporting. The qualitative research method was employed in this study, focusing on technology companies in Tehran, with an estimated 80 active companies comprising the target population. From this population, four companies were selected as samples, resulting in a total research population of 260 individuals. The sample size was determined based on the availability of panel members, with 30 experts ultimately selected for participation. Initially, a researcher-designed questionnaire, based on prior studies and expert opinions, was used as the data collection tool. Descriptive and inferential statistics were employed to analyze the research questions, with a three-round Delphi method utilized in the inferential statistics to model the design and implementation challenges of data warehousing. The findings revealed that 26 indicators were presented to the participants for evaluating the dimensions, and the participants collectively endorsed and categorized the indicators into three dimensions: business conditions, technology utilized, and data type.

To cite this article: *Dolatshah Behrouz, Nazari Farrokhi Ebrahim, Qorbani Khatereh (2024) Designing and Implementing A Data Warehouse in Technology Companies, Journal of Management and Educational Studies 2(1), 49-57*

1. Introduction

Technology refers to electronic and digital software-based technologies. A technology company can operate in the digital e-business, software, and internet services sectors, including e-commerce (Heath, 2017). The increasing market competitiveness results in a permanent reduction in the capacity to effectively and promptly respond to new market trends. Companies are inundated with complex data, and those who can transform it into valuable information will gain a competitive advantage. Many major technology companies build their reputation through innovation and substantial investment in research and development. According to PricewaterhouseCoopers' ranking of 1000 innovations in 2017, technology companies led in the number of innovations. Amazon has the highest research and development (R&D) investment, followed by Alphabet and Intel. In Iran, the term "technology unit" refers to companies that have been accepted as members of a science and technology park. Undoubtedly, the use of a data warehouse can bring many benefits to technology companies. This system enables organizations to analyze large volumes of diverse data. Significant results can be derived from the analyzed information. Data warehouses have four distinctive characteristics. These four features have enhanced the advantages of using this system.

Data warehouses are not simple systems. Their natural complexity, owing to the kind of problems they are intended to solve, providing business analysts a unified view to information, is added to the lack of a model that defines which techniques should be applied and which is the framework for the development of this kind of systems (Kelly, 1997; Rodero et al, 1999).

The implementation of data warehouses in technology companies is of considerable importance and serves as a fundamental tool for managing and analyzing a large volume of information. These warehouses facilitate the integration of data from various sources, including internal systems, customer interactions, and market trends, into a single, unified platform. By doing so, companies can gain a comprehensive view of their operations, improve decision-making processes, and facilitate the identification of emerging opportunities. Data warehouses provide technology companies with valuable insights into their customers' preferences and behaviors, enabling personalized marketing campaigns and targeted product development initiatives. Additionally, these robust databases enhance operational efficiency by streamlining reporting processes and reducing the time needed for data retrieval. The capability to extract structured and unstructured data from data warehouses allows for real-time analytics, empowering technology companies to maintain a competitive edge, drive innovation, and enhance overall business performance. Based on this, conducting this research has numerous theoretical and practical benefits for technology companies, as outlined below:

- Enhancing the environment for government, organizations, and the business community to improve planning and decision-making
- Assisting in the formation of policy decisions.
- Implementation of a business intelligence and data mining dashboard
- The formulation of the data warehouse and data model is essential to enable these activities.

2. Data Warehouse Background Work

A data warehouse is a repository of data from an organization's operational systems and other sources that supports analytics applications to help drive business decision-making. Data warehousing is a key part of an overall data management strategy: The data stored in data warehouses is processed and organized for analysis by business analysts, executives, data scientists and other users (Pratt, 2023). In computing, a data warehouse (DW or DWH), also known as an enterprise data warehouse (EDW), is a system used for reporting and data analysis and is considered a core component of business intelligence. Data warehouses are central repositories of integrated data from one or more disparate sources. They store current and historical data in one single place. that are used for creating analytical

reports for workers throughout the enterprise. This is beneficial for companies as it enables them to interrogate and draw insights from their data and make decisions (Nedim & Clare, 2016).
Inmon et al, 1994 says that a data warehouse is defined as “a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management’s decision-making process”. Elmasri & Navathe (2000) says that “data warehouses store huge amount of information from multiple data sources which is used for query and analysis. Therefore, the data is stored in the multidimensional (M D) structure”. A multidimensional model stores information into facts and dimensions. A fact contains the interesting concepts or measures (fact attributes) of a business process (sales, deliveries, etc.), whereas a dimension represents the perspective or view for analyzing a fact (product, customer, time, etc.) using hierarchically organized dimension attributes. Multidimensional modelling requires specialized design techniques that resemble the traditional database design (Rizzi et al, 2006; Jindal1 and Taneja, 2012) as shown in table 1.

Table 1. Database Design Methods (Jindal1 and Taneja, 2012)

Step	Input	Output
Analysis of Operational System	Information regarding the operational systems	Database schemes
Requirements Elicitation	Database scheme	Specifications for data warehouse
Conceptual design	Database scheme and Specifications	Conceptual schema
Logical design	Conceptual schema	Logical schema
Physical design	Logical schema	Physical schema

The pioneer author in the field of data warehouse design is Juan Trujilio (2001). He has made a major contribution. He proposed the use of UML for the design of data warehouse. Another author who also has a significant role in the design of data warehouse is Stefano Rizzi (2006).The author in proposed a graphical conceptual model for data warehouses, called Dimensional Fact model, and gave a semi-automated methodology to build it from the pre-existing (conceptual or logical) schemes. Then in based on the Dimensional Fact Model (DFM), he gave a general methodological framework for data warehouse design. Then he discussed some issues in Multidimensional modelling for the design of data ware house in. After that different authors gave different techniques and models for the design of data warehouse which we have discussed and compared in the next section (Jindal1 and Taneja, 2012).
One of the main functions of the project director consists in setting the developing framework, evaluating the chosen solution, especially with regard to the availability of technology and data needed. That is how the implementation of solutions for which there is neither a technological nor organizational background is avoided (Gardner, 1998; Rodero et al, 1999).

3. Research Method

The research method used in this study is qualitative. The target population for this research consists of technology companies in Tehran, with approximately 80 active companies. Four of these companies were selected as samples, resulting in a total research population of 260 individuals.
The sample size was determined by the availability of panel members. In this research, 30 experts were selected based on their willingness to participate.

The data collection tool in the initial stage is a researcher-designed questionnaire, which is based on previous studies and expert opinions.
The validity of the questionnaire will be confirmed using the Delphi method. To assess the reliability of the questionnaire, the Cronbach's alpha test was utilized. Since the Cronbach's alpha of the variables is greater than 0.70, it can be concluded that the questionnaire has the necessary reliability. Table 2 shows the test results.

Table 2: Questionnaire Reliability

Row	Criterion Name	Alpha
1	Business conditions	0.78
2	Utilized technology	0.77
3	Data type	0.82

Descriptive and inferential statistics have been utilized to analyze research questions. In inferential statistics, a three-round Delphi method has been utilized to model the design and implementation challenges of data warehousing.

4. Findings

The Delphi method is a systematic and interactive approach for gathering opinions and predictions from a panel of independent experts over two or more rounds. It is a type of consensus method that does not require in-person meetings.
The established methods of consensus are:
• Delphi process
• Nominal Group Technique
• Consensus Development Conference
The purpose of the Delphi process is to assess the level of agreement among experts or the general public on a specific topic and to establish a consensus opinion, while also identifying areas of disagreement. The Delphi technique typically involves the use of a questionnaire. While focus groups intentionally use group dynamics to generate discussion about a topic, Delphi methods maintain the anonymity of participants even after the study.
After selecting the panel members according to the three steps outlined by Schmidt et al. (2001), the selection and Delphi process commenced. The questionnaires were distributed and collected in person. Table (3) displays the distribution and collection dates of questionnaires for each round, along with their respective numbers.
Table 3: Questionnaire Distribution and Compilation

Rounds	Questionnaires Compilation		Questionnaire Distribution	
	Distribution Date	Number	Last date	Number
First	1 Sep 2023	30	19 Sep 2023	28
Second	23 Sep 2023	28	2 Oct 2023	25
Third	4 Oct 2023	25	22 Oct 2023	24

The research question is: What are the indicators of data warehouse design and implementation in information technology companies?

4.1.First Round: Generating Ideas

At this panel members participants were members the at of companies. They were asked to specify the indicators related to the pathology of data warehouse design and implementation in technology companies. Thirty questionnaires were distributed in the first round, and ultimately 28 questionnaires were returned. In this section, 18 indicators have been endorsed based on their interpretation of data warehouse design and implementation pathology, while 9 indicators have been proposed. Table (4) presents the initial and recommended indicators.

Table (4) Primary Indicators List

Answers provided in the initial round	Yes	No
Construction method based on previous techniques	23	5
Analyzing existing information systems and collecting user requirements	26	2
Use constraints and stereotypes to represent Model-Driven (MD) modeling features in UML	27	1
Implementation of the WAND tool	23	5
Utilizing an object-oriented approach based on an integrated modeling language	24	4
Various object-oriented model structures are utilized, including nodes, detail levels, arcs, and stereotypes	24	4
Design guidelines are provided.	25	3
Master of MD modeling, MD databases, and OLAP support	23	5
From computer modeling to final implementation	24	4
Utilizing UML-based packages	25	3
Attention to unresolved issues in warehouse modeling and design	26	2
Utilizing UML to organize components	26	2
Translation of UML model to XML logical model	28	0
Providing various methods to counter standardization, such as increasing data redundancy and adding derived columns	23	5
Use action rules to transform the data model into a logical model	23	5
Mapping rules for converting a UML class diagram to a multidimensional model	22	6
Developing an object-oriented data model in Markdown for describing data	20	8
Using Ontology to Address Current Data Warehouse Design Limitations	18	10
Indicators suggested by experts		
Use of Decision Support System (DSS) technologies, applications and methods for collecting, integrating and analyzing information or data.		
Considering the business process in the design		
Consider the actual document(s) when designing the data warehouse		
Attention to different users of the system in design and implementation		
Attention to the location and server management		
Paying attention to the server security		
Provide an architecture diagram for each location and/or a generic high-level diagram		
Use a specialized version of the data warehouse		
Existence of data exchange between different locations		

According to the results of the above table, the number of "yes" answers compared to "no" is very high and a large percentage of the respondents are assigned to "yes."

4.2.Second round: Reduction of items

At this stage, all the ideas of the members were summarized and classified, and the same words were defined for the factors, and finally, 27 ideas that could be examined and summarized were provided to all the experts and the members' opinions about the importance of the factors and indicators. received. At this stage, factors with medium and lower importance were removed and removed based on Kendall's coordination coefficient table.

Table 5: Interpretation table of various values of Kendall's coordination coefficient

Indicators	**Kendall**	**Result**
Construction method based on previous techniques	0.515	Moderate consensus
Analyzing existing information systems and collecting user requirements	0.724	Strong consensus
Use constraints and stereotypes to represent Model-Driven (MD) modeling features in UML	0.724	Strong consensus
Implementation of the WAND tool	0.778	Strong consensus
Utilizing an object-oriented approach based on an integrated modeling language	0.695	Strong consensus
Various object-oriented model structures are utilized, including nodes, detail levels, arcs, and stereotypes	0.750	Strong consensus
Design guidelines are provided.	0.726	Strong consensus
Master of MD modeling, MD databases, and OLAP support	0.726	Strong consensus
From computer modeling to final implementation	0.821	Strong consensus
Utilizing UML-based packages	0.720	Strong consensus
Attention to unresolved issues in warehouse modeling and design	0.720	Strong consensus
Utilizing UML to organize components	0.850	Strong consensus
Translation of UML model to XML logical model	0.766	Strong consensus
Providing various methods to counter standardization, such as increasing data redundancy and adding derived columns	0.756	Strong consensus
Use action rules to transform the data model into a logical model	0.623	Strong consensus
Mapping rules for converting a UML class diagram to a multidimensional model	0.756	Strong consensus
Developing an object-oriented data model in Markdown for describing data	0.784	Strong consensus
Using Ontology to Address Current Data Warehouse Design Limitations	0.656	Strong consensus
Use of Decision Support System (DSS) technologies, applications and methods for collecting, integrating and analyzing information or data.	0.826	Strong consensus
Considering the business process in the design	0.762	Strong consensus

Consider the actual document(s) when designing the data warehouse	0.628	Strong consensus
Attention to different users of the system in design and implementation	0.820	Strong consensus
Attention to the location and server management	0.870	Strong consensus
Paying attention to the server security	0.825	Strong consensus
Provide an architecture diagram for each location and/or a generic high-level diagram	0.720	Strong consensus
Use a specialized version of the data warehouse	0.832	Strong consensus
Existence of data exchange between different locations	0.679	Strong consensus

Hence, in accordance with the findings presented in the aforementioned table, factors of moderate and low significance were excluded using Kendall's coefficient of concordance table. Additionally, one indicator (Construction method based on previous techniques) was eliminated from the initial set of 27 indicators due to its weak and moderate importance coefficients. Consequently, based on the Kendall's coefficient of concordance, 26 indicators were ultimately retained, as indicated in the subsequent table.

The indicators were classified in 3 dimensions of business conditions, technology used, and data type, which are shown in Figure 1. By removing the indicators of low importance, the final operating model of this research was approved according to the following figure based on the opinion of experts.

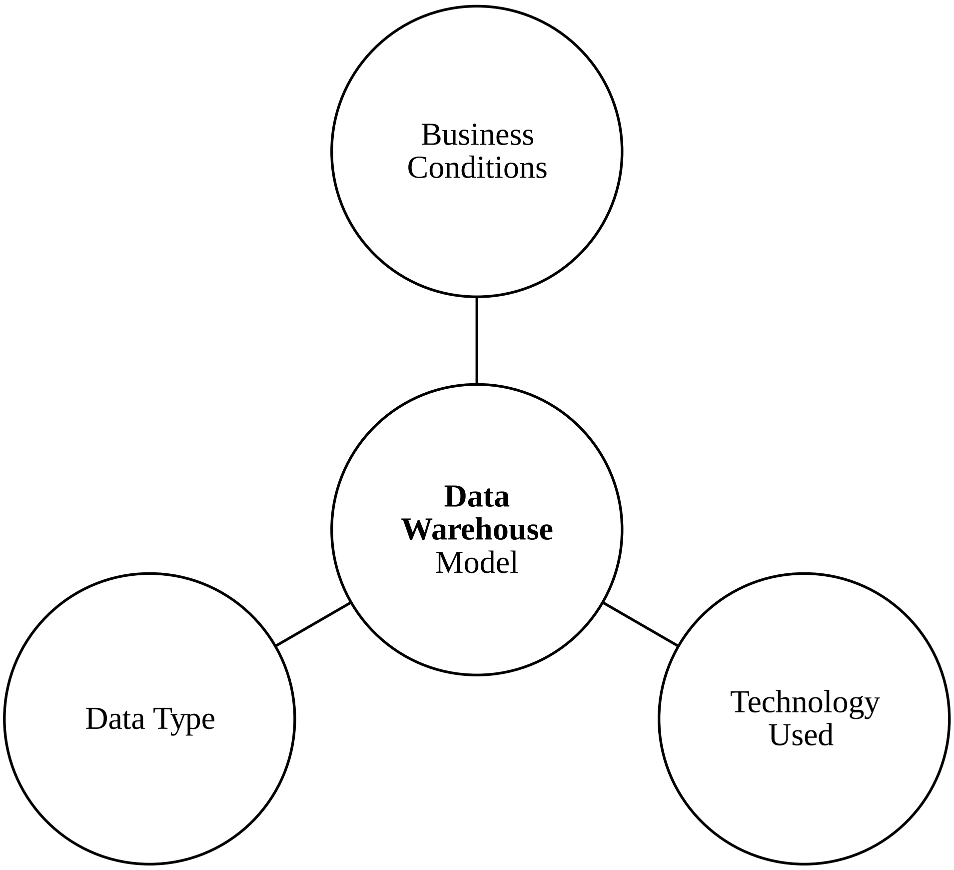

Figure No.1: The final research model

4.3.The third round: determining the combination of indicators and dimensions

26 indicators were presented to the participants for assessing the dimensions, and the participants collectively endorsed and categorized the indicators into three dimensions: business conditions, technology utilized, and data type, as depicted in Table 6.

The final model was approved based on the opinion of the elites with 26 indicators and 3 item dimensions as described in Table 6.

Table 6: dimensions and indicators of the model

Indicators	Dimensions
Using Ontologies to Overcome Current Data Warehouse Design Deficiencies	Business Conditions
Covering computer modeling to final implementation	
Considering the business process in the design	
Attention to different users of the system in design and implementation	
Attention to the location and management of server	
Provide design instructions	
Analysis of existing information system and collection of user needs	
Use constraints and suffixes to represent MD modeling features to UML	Technology Used
Implementation of the WAND tool	
Providing various anti-standardization methods such as increasing data redundancy, increasing derived columns, etc.	
Use the action rules to transform the data model into a logical model	
Mapping rules to convert UML class diagram to multidimensional model	
Definition of an object-oriented MD data model for data description	
Use of decision support system (DSS) in technologies, applications and methods of collecting, integrating and analyzing information or data.	
Attention to server security	
Using a specialized version of the data warehouse	
Using an object-oriented approach based on integrated modeling language	Data Type
Use of various object-oriented model structures such as nodes, detail levels, arcs, stereotypes	
Using MD modeling of MD databases and OLAP support	
Using packages based on UML	
Attention to open issues in warehouse modeling and design	
Using UML to group units	
Translation of UML model to XML logical model	
Consider the actual document(s) when designing the data warehouse	
Provide an architecture diagram for each location and/or a generic high-level diagram	
Existence of data exchange between different locations	

5. Conclusion

In order to address the objective of this phase, the Delphi method was initiated following the selection of panel members using the three-step approach outlined by Schmidt et al. (2001). Questionnaires were distributed and collected offline via Google Form through various social networks. The distribution of questionnaires occurred in three rounds on different dates over a period of approximately three months. The data collected was then analyzed using SPSS26 software to address the research questions in the initial phase.

During this stage, panel members were chosen from experts in technology companies. They were tasked with identifying indicators related to the issues surrounding data warehouse design and implementation in technology companies. Initially, 30 questionnaires were distributed in the first round, with 28 questionnaires ultimately being returned. Based on their insights into the pathology of data warehouse design and implementation, participants affirmed 18 indicators and proposed 9 additional indicators.

Subsequently, all member contributions were consolidated and categorized. Similar terms were standardized for the factors, resulting in 27 ideas that could be examined and summarized. These were then presented to all experts to gather their opinions on the importance of the factors and indicators. Factors with moderate and lower importance were then eliminated based on Kendall's coordination coefficient. One index, specifically the construction method based on previous methods, was removed from the 27 indicators due to its weak and moderate importance coefficient. Ultimately, 26 indicators were analyzed based on Kendall's correlation coefficient.

The 26 indicators were then presented to the members to determine the dimensions. The members unanimously approved and classified the indicators into three dimensions: business conditions, technology used, and data type.

6. References

Catherine Heath (2017-11-01). "What Is A 'Tech Company,' Anyway?". Tech Nation. Archived From The Original On 11 July 2019. Retrieved 2018-07-09.

Dedić, Nedim; Stanier, Clare (2016). Hammoudi, Slimane; Maciaszek, Leszek; Missikoff, Michele M. Missikoff; Camp, Olivier; Cordeiro, José (eds.). An Evaluation of the Challenges of Multilingualism in Data Warehouse Development. International Conference on Enterprise Information Systems, 25–28 April 2016, Rome, Italy (PDF). Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016). Vol. 1. SciTePress. pp. 196–206. doi:10.5220/0005858401960206. ISBN 978-989-758-187-8. Archived (PDF) from the original on 2018-05-22.

Elmasri, R. & Navathe, S.B, (2000) Fundamentals of Database Systems, Addison Weasely PubCo. ISBN 0201542633

Gardner, S. (1998) Communications of the ACM, vol. 41, Nº 9, September, pp 52-60.

Inmon, W.H., Hackathorn, and R.D (1994) Using the data warehouse. Wiley-QED Publishing, Somerset, NJ, USA.

José A. Rodero, José A.Toval, Mario G. Piattini (1999) The audit of the Data Warehouse Framework, Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99) Heidelberg, Germany, 14. - 15. 6.

KELLY, S. (1997) “Data Warehousing in Action”. John Wiley & Sons.

Mary K. Pratt (2023) Definition data warehouse, https://www.techtarget.com/searchdatamanagement/definition/data-warehouse

Rajni Jindal1 and Shweta Taneja (2012) Comparative Study of Data Warehouse Design Approaches: A Survey, international Journal of Database Management Systems (IJDMS) Vol.4, No.1, February 2012 DOI: 10.5121/ijdms.2012.4104 33

S. Rizzi, A. Abelló, J. Lechtenbörger, J. Trujillo (2006) “Research in data warehouse modelling and design: dead or alive?” DOLAP, ACM, , pp. 3–10.

Trujillo, M. Palomar, J. Gómez, I.-Y. Song(2001),” Designing data warehouses with OO conceptual models”, IEEE Comput. 34 (12) (2001) 66– 75.