The Effect of Using a Data Warehouse in Companies

1st Mario Christianto
Computer Science Department,
School of Computer Science
Bina Nusantara University Jakarta,
Indonesia 11480
mario.christianto001@binus.ac.id

2nd Stephanus Nicholas
Computer Science Department,
School of Computer Science
Bina Nusantara University Jakarta,
Indonesia 11480
stephanus.nicholas@binus.ac.id

3rd Syarief Kamal
Computer Science Department,
School of Computer Science
Bina Nusantara University Jakarta,
Indonesia 11480
syarief.kamal@binus.ac.id

4th Alexander A S Gunawan
Computer Science Department,
School of Computer Science
Bina Nusantara University Jakarta,
Jakarta, Indonesia 11480
aagung@binus.edu

***Abstract*—In recent days, especially in the industrial era 4.0, technological developments are increasingly rapid, the need for information is one of the references for companies in making decisions. The need for information or data is increasingly controlled; therefore, a data warehouse is needed to collect data from various sources for reporting and analysis. The data warehouse itself allows a company to access and obtain accurate information in a reasonably short time. Based on this study, we collected data from various respondents to determine how significant the influence of the data warehouse is in analyzing company data. The purpose of this study is to find out that the data warehouse can have a significant influence on decision making and affect the company's performance.**

Keywords—Companies, Data Warehouse, Analysis, OLAP (On-Line Analysis Processing), Effect

I. Introduction

In this 4.0 industrial revolution, the development of information systems in Indonesia is very rapid. This development has also boosted the number of startups in Indonesia. Various startup fields have emerged, such as e-commerce, accommodation, and transportation. Examples of startups in Indonesia include mamikos which is engaged in accommodation, Tokopedia as an E-commerce startup that holds the title of a unicorn. And there is also Gojek which is engaged in transportation. As technology develops rapidly, everything related to digitalization and information is essential for managing data. The use of information systems that are more effective and efficient is vital for a company to continue to compete in this industrial era 4.0.

In a company, Data warehouse can be a tool to facilitate the executive in analyzing patient data as a consideration in decision making [1]. In addition, this decision-making activity is a crucial part of the company in carrying out a company process or activity that can advance company performance. A data warehouse is a system that consolidated information from different and generally fluctuated sources into one far-reaching and effortlessly controlled database [2]. Data warehouse is one of the major IT infrastructures [3]. Data warehouse plays an essential role as an information center and in decision making. A Data Warehouse stores large amounts of data to efficiently analyze historical information and manage it in a proper way [4]. Therefore, through this research, we want to know how much influence the data warehouse has on the performance of E-commerce companies in analyzing data.

II. Literature Review

Data Warehouse is a method used to store large amounts of data [1]. Data Warehouse itself has been known in various industries due to its quite effective impact on company performance [2]. Besides, the data warehouse has an essential role in providing information on the right business strategy for a company. Later, this information will be used to achieve a competitive advantage for the company [3]. This system uses a combination of technologies that help in data management. In a company, decision-making is vital. Making decisions must be based on facts, data, and credible statistics [4, 5]. According to Ponniah, data warehouses have several characteristics, namely (1) Subject Oriented, (2) Integrated, (3) Not changeable, (4) Time Variance, and (5) Granularity [6].

The effectiveness of a data warehouse is usually determined based on how to maintain the data warehouse so that it remains stable, scalable, and efficient. The way to maintain it is to know how the data is designed from data architecture, data modeling, and ETL [3, 7]. Data Warehouse architecture is an overview of the global system from the beginning to the end of how entities or actors are participating in the system [6, 8]. The formation of a data warehouse architecture must ensure that the technical planning must be well understood. The system development in sending information to the data warehouse runs by technical planning. In line with that, data modeling also has a considerable influence in forming a data warehouse that is stable, scalable, and efficient [3]. Data modeling is used to determine and analyze the data requirements needed to support business processes in a company and organization. Data modeling also has a significant influence on the design of a data warehouse [9]. There are two data models, namely Conceptual Data Model (CDM) and Physical Data Model (PDM). The Conceptual Data Model is a database model based on data collection and analysis [10]. At the same time, the physical data model is a database model design that is more specific than the Conceptual Data Model by adding relationships between these data [11]. Both data models have the same function: carrying out the identification process and data analysis [10]. Furthermore, ETL Tool has a function: to extract data from the database, change it, and load it into the company's Data Warehouse [12, 13]. Therefore, an effective data warehouse design can help company management make critical evaluation decisions [14].

Many models are often used in data warehouses, one of which is the star schema and snowflake schema. According to Conolly, a star schema is a dimensional data model that has a fact table. At the same time, snowflake is a type of star schema where the table does not contain denormalization. The main difference between snowflake schema and star schema is that all tables in the snowflake schema dimensions have been normalized so that the data has an efficient record because there is no data redundancy. In contrast to a star schema, dimension tables are not normalized. This can cause problems in integrity and quality. data [15, 16]. Therefore, designing a good data warehouse will have a significant impact on the company in analyzing data.

There are traditional data warehouses as well as those that are currently developing, namely modern ones. Data warehouses can be said to be a combination of concepts and technologies that facilitate organizations to organize and maintain the history of data that has been obtained. The traditional data warehouse is a traditional method that various agencies have used for a long time and is very efficient in storing data. Data warehouses do have high capacity and performance [17]. However, this traditional data warehouse will be inefficient if the data that has been obtained is in large amounts, such as millions or billions of data, that modern data warehouses have been created and developed until now to make company data warehouses with large amounts of data that can process data quickly and right [18, 19, 20]. Meanwhile, a modern data warehouse has made it possible for a company to store and process large amounts of data. However, because this modern data warehouse is still developing, several problems have no solution. End users in a modern data warehouse need a data scientist who has understood the technology, algorithms, mathematics, and statistics; other than that, if using a traditional data warehouse, it only requires an analyst who does not need to know specific knowledge about technology or data exploration. [19, 21].

Some of the effects or benefits of using a data warehouse in a company are the provision of appropriate information [3], increasing business optimization [22], helping to solve company problems with large amounts of data through the OLTP system, and determining the right strategy through the OLAP system [10, 23], stores the flow and history of data used to make analytical reports of a company [24], can help in analyzing and making the right decisions [25], offers speed in decision making and data analysis [26, 27], as well as for trying to predict the outcome of some data that is accommodated in a data warehouse [28].

III. METHODOLOGY

A. Selected Method

The research method used in this study is a quantitative method. This research was conducted in several stages, namely the first stage, on the data. The data was collected by using one of the fact-finding techniques, namely a questionnaire. Where our group distributed several questionnaires to data warehouse users who work in an E-commerce company. Through the questionnaire data, we can find out more about anything that has a big impact on the data warehouse in an E-commerce company. The second stage, data analysis. From the data we collect through questionnaires, we try to help the data become numeric data through an index method based on the answers given by respondents. Later the results of our research want to show the effective use of a data warehouse in a company by looking at the value of the data warehouse variables based on the interval values that we have determined.

B. Questionnaire

The research method used in this study is to use an online questionnaire. In this survey we asked respondents about the effectiveness of the data warehouse, the factors that influence the effectiveness, namely, ETL, data modeling and data architecture, where these three factors will be used as independent variables and the results of these factors determine how it works. effective data warehouse.

The question such as:

1. Have you ever used a data warehouse in your workplace ?
2. How effective is the use of a data warehouse in your workplace ?
3. In the company, a stable data warehouse is needed. In your opinion, what factors affect the stability of the data warehouse ?
4. Is the performance in processing data better by using a data warehouse ?
5. How significant is the influence of the data warehouse in your company? If the effect is significant, include an explanation.

based on the questions asked above, the purpose of each question is

1. **Have you ever used a data warehouse in your workplace?**
   This question is to find out how many participants have ever used a data warehouse in their workplace.

2. **How effective is the use of a data warehouse in your workplace?**
   This question aims to find out how effective the use of data warehouses in their workplace is.

3. **In the company, a stable data warehouse is needed. In your opinion, what factors affect the stability of the data warehouse?**
   This question is to find out what factors affect the data warehouse based on the opinions of the participants.

4. **Is the performance in processing data better by using a data warehouse?**

This question is to calculate how many respondents think that data warehouses can improve performance in processing data.

5. **How significant is the influence of the data warehouse in your company? If the effect is significant, include an explanation.**
   This question is to determine how much influence the data warehouse has on the company and its impact.

C. Research Question

In this study we have several research questions that proposed the main objective of this research. The question are as follow:

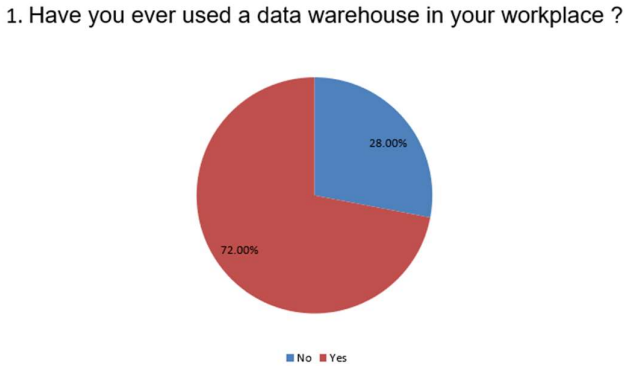
1. Will the level of data effectiveness increase significantly when using a data warehouse?
2. Does implementing Data Warehouse make it easier to access historical data?
3. How to measure the effectiveness of the Data Warehouse in a company?

D. Participant

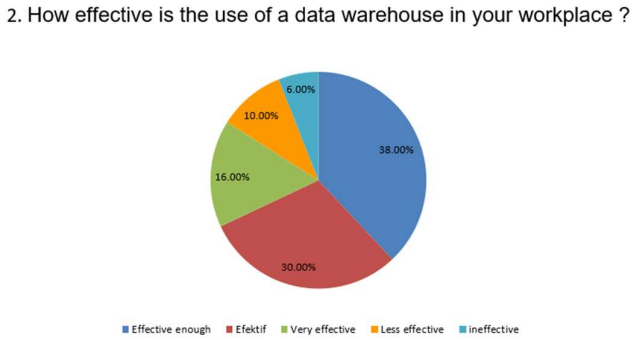
This research is specifically designed to target data warehouse users, where the data we have collected will be used to conclude how much influence the data warehouse has on the company. To make participants reach this questionnaire, we distributed the questionnaire form to a group containing data warehouse users. We made this questionnaire voluntarily; the respondent has no obligation to answer the questionnaire and there is no coercion.

IV. Result & Discussion

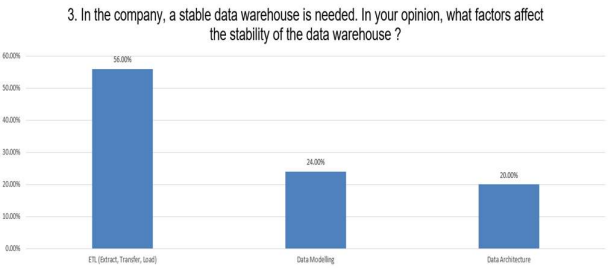
In this section, we will discuss the data that has been obtained from the survey. Out of a total of 50 respondents, the majority were male, and some were female.



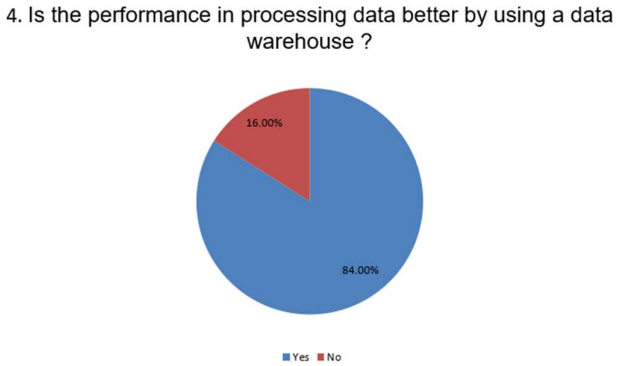
From a total of 50 respondents, 72.00% of the respondents stated that they had used a data warehouse in their company, and partly again, namely, 28.00% of the total respondents indicated that they had never used a data warehouse. From these results, it can be concluded that most companies implement a data warehouse system as a place to store and process data. This may be due to the advantages possessed by the data warehouse itself, namely that it can provide structured information management and well-organized data.



In the second question, 38.00% of respondents answered that the data warehouse is quite effective in their workplace, 16.00% answered that the data warehouse is very effective. On the other hand, some answered that it is not effective as of as 6.00%. From the data obtained, it can be concluded that from 50 respondents, the majority of respondents said that the data warehouse is quite effective in their workplace. These results are influenced by several important things: data warehouses' ability to assist companies in providing data reports and making decisions. This causes most respondents to answer that the data warehouse is quite effective and influential in their workplace.



The graph above shows that of the three factors, namely ETL, data modeling, and data architecture, the most critical role for data warehouse stability is ETL. This can be because ETL is a crucial step in converting a data format from a data source to a data warehouse format to be stored centrally in a data warehouse. Therefore, this ETL can be said to be the most influential among others because if data that has been changed has poor accuracy, it can cause the results of data reports to be not good. Making decisions will also be inaccurate and can be detrimental to the company.



From a total of 50 respondents, as many as 84% answered that the performance in processing data was better by using a data warehouse. However, another 16% answered no as the ability of the data warehouse itself is to be able to quickly and easily access and process data because one of the processes of the data warehouse system itself has an ETL process. Where in the ETL process, there is a data cleansing process that functions to remove duplicate data and turn it into quality data so that data processing and access can be carried out properly using a data warehouse.

5. How significant is the influence of the data warehouse in your company ? If the effect is significant, include an explanation.
12 / 50 Respondent

No	Answer	Total Answer	Percentage
1	sangat membantu	1	8.33%
2	N/A	3	25.00%
3	sangat penting dan berperan	1	8.33%
4	sangat penting agar membantu memudahkan pekerjaan setiap karyawan	1	8.33%
5	sangat penting, karena data sangat akurat	1	8.33%
6	sangat penting dan berpengaruh untuk perusahaan	1	8.33%
7	membantu pekerjaan saya dalam menganalisis data	1	8.33%
8	cukup penting untuk management stok dan pengiriman	1	8.33%
9	penting untuk memperbaiki seluruh kendala, mengetahui data	1	8.33%
10	ketika fetching data ceoat, data data dr berbagai sumber pun lengkap hanya dengan ambil key id nya saja meskipun dari berbagai table tidak akan memberatkan proses	1	8.33%
Total		**12**	**100.00%**

Of the total nine respondents who answered our question no 5, the majority answered that the influence of the data warehouse in a company is significant, the impact is of various kinds such as helping performance in analyzing data, managing stock and shipping, accurate and easy data access and simplifying profession.

V. Conclusion

In conclusion, from our research, we can see that a data warehouse has a significant influence and impact on a company. This has shown us the importance of a data warehouse in an enterprise. We found that the data warehouse affects performance in processing and accessing data. However, the data warehouse also stores credible facts and data to make decision-making in a company more accurate. Not only that, based on the results we obtained, we found that the magnitude of the influence of the data warehouse is determined by several factors, including ETL (Extract, Transform, Load), Data modeling, and Data architecture.

In the end, we can conclude that a data warehouse is vital for companies, especially in providing consistent and quality data. Moreover, companies will be faster in accessing data with a data warehouse because data warehouse data is stored centrally so that data can be accessed simultaneously in a short time.

References

[1] I. B. L. M. Suta, I. G. N. A. S. Mahendra and Y. P. Sudarmojo, "Design General Hospital Data Warehouse Base on Nine Step Methodology," *International Journal of Engineering and Emerging Technology, Vol. 4, No. 1,* 2019.

[2] A. Khatoon, S. Singh, S. Kumar and H. Chawla, "Business Analysis Based on Datawarehouse Design and Usage," 2018.

[3] N. Rahman, "An empirical study of data warehouse implementation effectiveness," *International Journal of Management Science and Engineering Management,* pp. 1-2, 8, 2016.

[4] O. Moscoco-Zea, Andres-Sampedro and S. Lujan-Mora, "Datawarehouse design for Educational Data Mining," *2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET),* 2016.

[5] M. Y. Santos, B. Martinho and C. Costa, "Modelling and implementing big data warehouses for decision support," 2017.

[6] M. K. Tino Feri Efendi, "Warehouse Data System Analysis PT. Kanaan Global Indonesia," *Internation Journal of Computer and Information System (IJCIS),* p. 30, 2020.

[7] H. N. Mahadev and A. A. B. Bagwan, "An Effectual Association of Data Warehousing with the Scenario of Trajectory Applications," *International Journal of New Computer Architectures and their Applications (IJNCAA),* 2017.

[8] I. Moall, A. Nabli, L. Bouzguenda and M. Hammami, "Data warehouse design approaches from social media: review and comparison," *International Journal of Computing and Business Research (IJCBR),* 2017.

[9] K. A. B. Permana, G. B. Subiksa and M. Sudarma, "Design Data warehouse for Centralized Medical Record," *International Journal of Engineering and Emerging Technology, Vol. 2, No. 2, July,* 2017.

[10] P. S. A. D. Rahadi, P. Widiadnyana and N. Sweden, "Designing Data Warehouse in Finance Company Study at PT ABC," *International Journal of Engineering and Emerging Technology,* pp. 60-61, 2017.

[11] A. Ribeiro, A. Silva and A. R. d. Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *Journal of Software Engineering and Applications,* p. 619, 2015.

[12] S. D. Fajar Zaki Al Faris, "Development of Data Warehouse to Improve Services in IT Services Company," *International Conference on Information Management and Technology (ICIMTech),* p. 483, 2018.

[13] S. K. Makki and M. R. Hasan, "Measuring the Performance of Data Placement Structures for MapReduce-based Data Warehousing Systems," *International Journal of Research in Engineering, Science and Management (IJRSEM),* 2017.

[14] I. N. D. Kotama, A. A. G. O. K. Adnyana and K. O. Saputra, "Design of Data Warehouse for University Library using Kimball and Ross 9 Steps Methodology," *International Journal of Engineering and Emerging Technology, Vol. 4, No. 1,* 2019.

[15] K.I.Mohammed, "Data warehoouse Design and Implementation Based on Star Schema vs Snowflake Schema," *International Journal of Academic Research in Business and Social Sciences,* 2019.

[16] S. W. Felixiang Zhu, "Design and Implementation of Oceangoing Ship Schedulling Data Warehouse," *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FKSD),* pp. 1023-1024, 2018.

[17] V. M. Ngo, N. A. Le-Khac and M.-T. Kechadi, "Designing and Implementing Data Warehouse for Agricultural Big Data," 2019.

[18] M. Y. Santos, C. Costa, J. Galvão, C. Andrade, Ó. P. López and A. C. Marcén, "Enhancing Big Data Warehousing for Efficient, Integrated and Advanced Analytics," 2019.

[19] L. W. Santoso and Y. , "Data Warehouse with Big Data Technology for Higher Education," *Information Systems International Conference (ISIC),* 2017.

[20] F. D. Tria, E. Lefons and F. Tangorra, "Evaluation of Data Warehouse Design Methodologies in the Context of Big Data," 2017.

[21] F. D. Tria, E. Lefons and F. Tangorra, "Cost-benefit analysis of data warehouse design methodologies," 2017.

[22] I. N. A. Prabawa, D. A. K. Arimbawa and I. G. N. Janardana, "Analysis and Design Data Warehouse For E-Travel Business Optimization," *International Journal of Engineering and Emerging Technology,* pp. 25, 30, 2019.

[23] R. A. Setyawan, E. Prasetyo and A. S. Girsang, "Design and Implementation Data Warehouse in Insurance Company," 2019.

[24] R. M. Harris, "Data Warehouse and Decision Support System Effectiveness Demonstrated in Service Recovery During COVID19 Health Pandemic," *International Conference on Open Source Systems and Technologies (ICOSST),* p. 3, 2020.

[25] T. Suprawoto, "Design of Data Warehouse in Library Circulation using Pentaho," *Proceedings of the International Conferences on Information System and Technology (CONRIST 2019),* pp. 59-63, 2019.

[26] M. A. M. I. N. A. S. A. Putu Widiadnyana, "Analysis of Data Warehouse for Transactional Analysis Methods Online Analytical Processing (OLAP) at Company XYZ Silver," *International Journal of Engineering and Emerging Technology,* p. 72, 2017.

[27] G. B. V. T. P. G. G. V. M. Cinzia Muriana, "analysis, Effectiveness of an electronic health record-data warehouse system implementation: return on investment," *International Journal of Medical Engineering and Informatics,* p. 384, 2017.

[28] S. Deepa and R. Maruthi, "Prevention and Detection of Cancer Using Data Warehousing and Data Mining Techniques in Andaman and Nicobar Islands," *International Journal of Research in Engineering, Science and Management,* 2018.