

Implementation of Data Warehouse and Star Schema for Optimizing Property Business Decision Making

Willsen Wijaya¹, Jansen Wiratama^{2✉}, Santo Fernandi Wijaya³

^{1, 2, 3} Department of Information Systems Study Program, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Indonesia

Article Information

Article History

Received : March 15, 2024

Revised : April 11, 2024

Accepted : April 18, 2024

Keywords:

Data Warehouse; ETL; Pentaho Data Integration; Property Industry; Star Schema.

Kata Kunci:

Data Warehouse; ETL; Pentaho Data Integration; Property Industry; Star Schema.

ABSTRACT

The property industry is a sector that faces intense competition, so innovation is needed in planning business strategies. PT Wijaya Karya Realty faces several problems, such as determining strategic locations for its property products, namely houses or shophouses, and comparing house prices with market prices for houses in certain locations. To face this problem, it is necessary to implement a data warehouse and Extract, Transform, Load (ETL). This research focuses on implementing Pentaho Data Integration (PDI) as a tool for managing and cleaning data that can be used as analysis material to assist in data-based decision-making. Besides that, Mondrian uses the workbench scheme to improve data analysis effectively, and the star scheme, which includes dimensional tables and fact tables, can help optimize the data analysis process. The research results show that implementing Star Schema can provide information on decision-making and business strategy planning in the property industry.

ABSTRAK

Industri Properti merupakan salah satu sektor yang menghadapi persaingan ketat, sehingga dibutuhkan inovasi dalam merencanakan strategi bisnis. PT Wijaya Karya Realty menghadapi beberapa masalah, seperti menentukan lokasi yang strategis produk propertinya yaitu rumah atau ruko, dan membandingkan harga rumah dengan harga pasaran rumah di lokasi tertentu. Untuk menghadapi masalah tersebut, dibutuhkan implementasi data warehouse dan Extract, Transform, Load (ETL). Penelitian ini berfokus pada implementasi Pentaho Data Integration (PDI) sebagai alat dalam mengelola dan membersihkan data yang dapat dijadikan bahan analisis untuk membantu dalam pengambilan keputusan berbasis data. Selain itu, Mondrian melalui skema workbench digunakan untuk meningkatkan analisa data secara efektif dan star schema yang mencakup dimensional table serta fact table yang dapat membantu dalam mengoptimalkan proses analisis data. Hasil penelitian menunjukkan bahwa implementasi Star Schema dapat memberikan informasi dalam pengambilan Keputusan dan perencanaan strategi bisnis pada Industri Properti.

Corresponding Author :

Jansen Wiratama

Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara.

Scientia Garden Jl. Boulevard Gading Serpong, Kab. Tangerang-Banten, 15810.

Email: jansen.wiratama@umn.ac.id

INTRODUCTION

In a rapidly changing business landscape, the property industry demands companies to develop effective strategies to stay ahead of the competition. PT. Wijaya Karya Realty is a subsidiary of PT. Wijaya Karya specializes in constructing a wide range of properties, including private homes, offices, apartments, condotels, shopping centers, educational facilities, and other commercial buildings (Kannapadang & Ta'dung, 2022). PT Wijaya Karya Realty is a prominent player in the Indonesian property sector. It boasts a wide range of experiences and projects that have successfully delivered top-notch housing in various parts of Indonesia. PT Wijaya Karya Realty, being a company engaged in property development, has devised a strategic approach to cater to consumer demands, ultimately aiming to enhance company profitability. Every decision made by PT Wika Realty undergoes a meticulous process. Several factors have been taken into account, including those that impact price determination. The company aims to compare the price of the house it is selling with similar properties in specific areas. Additionally, Wika Realty is developing strategies for identifying the best locations to market houses. To address this issue, PT Wijaya Karya Realty must implement a data warehouse as a crucial solution. Data warehouse implementation will be the focus of the scope of this research to produce analytical results that are useful as information for decision-making.

A data warehouse (DWH) can efficiently store and consolidate diverse data from multiple central systems into a single location (Haryono et al., 2020). A data warehouse serves as a robust system created to gather and store data from multiple sources, enabling more in-depth analysis. Companies have the ability to control the frequency at which data is updated or pulled into the DWH during implementation (Paliwal & Saraswat, 2022). DWH implementation can be customized to meet the specific requirements of a business and accommodate varying rates of data change. Utilizing the power of data analysis, DWH enables companies to make informed decisions (Jayashree & Priya, 2019). DWH functions as a decision support system, enabling decision-makers to draw conclusions based on analyzed data.

Within the data warehouse framework, there are different components that play a crucial role, such as the extract, transform, and load (ETL) process. The ETL process extracts data from various sources, both internal and external to the company. (Awiti et al., 2019). This data is then transformed and loaded into the DWH. The data in the data warehouse is seamlessly integrated from a multitude of sources. Data quality is influenced by the source of the data, therefore it is necessary to perform data cleaning (Mathur et al., 2019). Ensuring the quality of the data extracted from transaction records is crucial for accurate analysis and informed decision-making. ETL is crucial for effectively managing data before it gets integrated into the data warehouse. Efficiently optimizing the ETL process is of utmost importance in order to minimize processing time and enhance data availability. Different approaches have been thoroughly assessed and valuable recommendations have been provided, taking into account specific use cases (Hahn, 2019; Setiawan et al., 2023). The data generated from the ETL process is of high quality and plays a crucial role in conducting analysis (Handika et al., 2020). This process strives to enhance the quality and relevance of the information stored in the DWH. Once the ETL process is complete, we can proceed with the star schema modelling activities. Star schema design includes fact tables and dimension tables, with the connection between the fact table and dimensional table established through primary keys in the dimensional table and foreign keys in the fact table. Various dimensional tables of varying granularities are merged with a fact table to create a star schema, as discovered in a recent study (Setiadi & Rahutomo, 2022).

This study utilizes Pentaho Data Integration (PDI) to perform data migration, data cleansing, and file loading into databases, as well as the reverse process, on a large scale (Pratama & Widhiasih, 2020). Utilizing a data warehouse can greatly enhance a company's decision-making process. These decisions are made by analyzing historical and current data, which is managed by the data warehouse (Mathur et al., 2019; Zahra et al., n.d.). This analysis will offer valuable insights into the utilization of data warehouse technology and the ETL process. These insights will help PT Wijaya Karya Realty make data-driven decisions and enhance the efficiency and

effectiveness of their data management and utilization. In addition, it is expected that they can effectively address challenges involving the analysis of house prices in specific markets and identifying optimal marketing locations for the properties.

RESEARCH METHODS

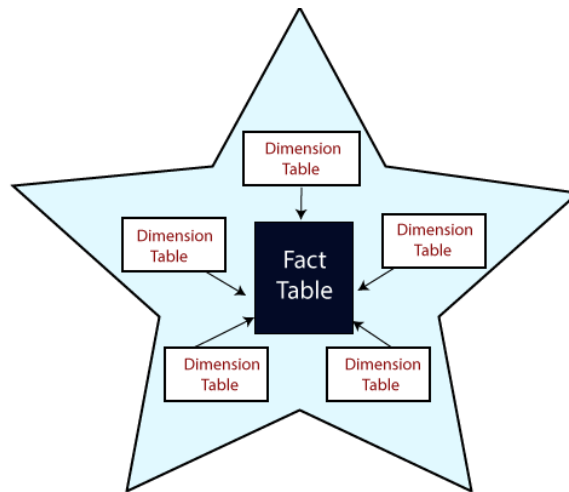


Figure 1. Star Schema on Data Warehouse

A table in a star schema contains facts and is linked to dimensions. A fact table contains two types of columns: one type includes the actual facts, while the other type consists of foreign keys that reference the dimension table. In most cases, the primary key of the fact tables consists of a combination of its foreign keys. A fact table can contain either detailed-level facts or aggregated facts (commonly referred to as summary tables). A fact table typically consists of facts that are aggregated at the same level (Effendy et al., 2019). A dimension is a structure typically made up of one or more hierarchies that organize data. If a dimension lacks hierarchies and levels, it is referred to as a flat dimension or list. The primary keys from the tables of each dimension are included in the composite primary keys of the fact table. Dimensional attributes are essential for defining the value in a precise and structured manner. They typically consist of descriptive textual values. Dimensional tables are typically smaller in size compared to fact tables (Awiti et al., 2020). Fact tables contain information regarding sales, while dimension tables hold data related to geographic regions (markets, cities), clients, products, times, and channels.

This research aims to establish connections between multiple-dimensional tables and a fact table using foreign keys. The dimensional tables used in this research are the `dim_property`, `dim_harga`, and `dim_demography` tables, which are depicted in Figure 9. `dim_property` has location data and total population which will be processed into relationships in a star schema with `dim_property` and `dim_harga`. Dimensional tables can be linked together as the fact table contains a unique key for each dimensional table. The star schema was designed with the goal of enhancing query performance and making data analysis more user-friendly. This will enable PT Wijaya Karya Realty to have a clear reference and visual representation for comparing house prices based on market prices in specific locations. It will also help determine the optimal location for marketing the house.

RESULTS AND DISCUSSION

In the Extraction, Transformation, and Load (ETL) process, the first step is to connect Pentaho Data Integration (PDI) to the database with a database connection. Next, retrieve or extract tables from the phpMyAdmin database using the input table component. Then, in the transformation process, the data will be tidied using Sort Rows to sort the data in the table based on specific columns. Next, the Select Values component selects the value or data you want to

retrieve for display. In the load process, the data will be moved to the database using the output table and written into Excel using Microsoft Excel Writer. The following is a display of carrying out the ETL process on several data tables used:

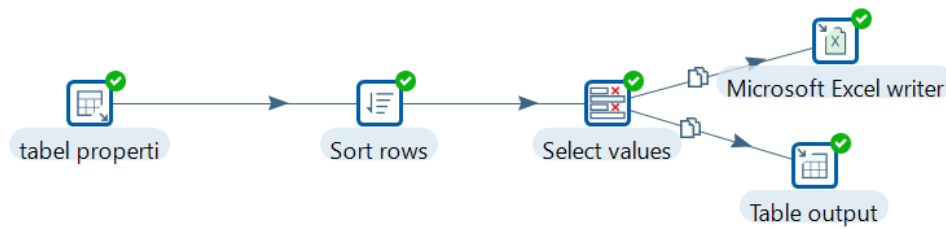


Figure 2. ETL Process on Property Table.

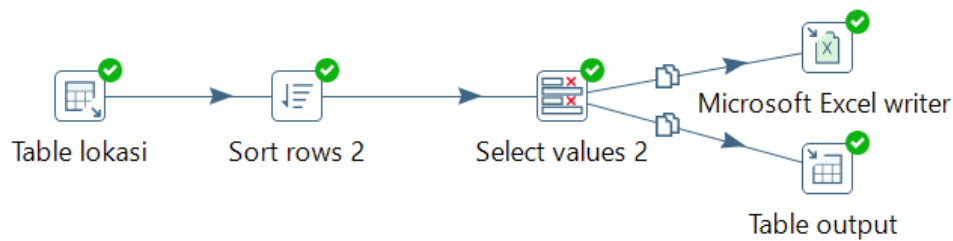


Figure 3. ETL Process on Location Table.

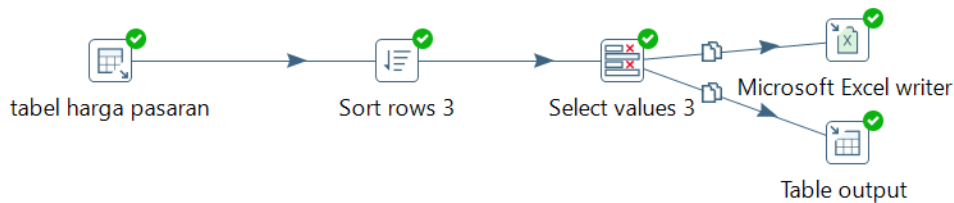


Figure 4. ETL Process on Market Price Table.

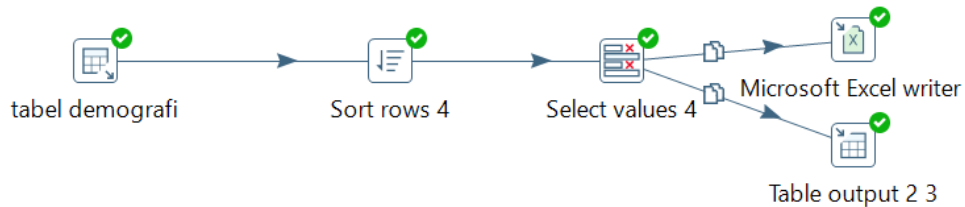


Figure 5. ETL Process on Demographic Table.

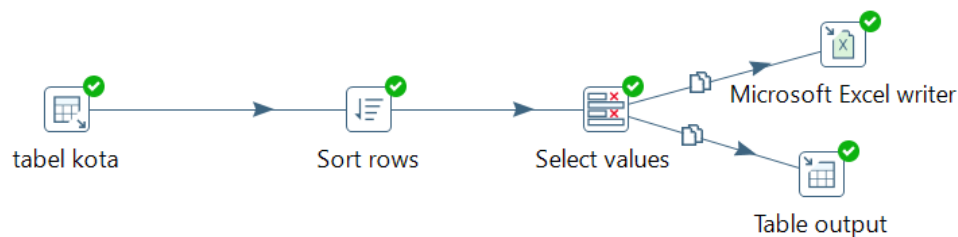


Figure 6. ETL Process on City Table.

Figure 2-6 is a picture of the ETL process on property tables, location tables, market price tables, demographic tables, and city tables where table extraction from the phpMyAdmin database is carried out using the input table component. Next, perform a transformation process, where you tidy up the data using Sort Rows to sort the data in the table based on specific columns. Use the Select Values component to select the value or data values you want to retrieve for display. During the load process, the data will be moved to the database using the output table and written into Excel using Microsoft Excel Writer.



Figure 7. Job's Transformation Data.

Figure 7 shows job transformation as one of the instructions for managing the data transformation workflow. The transformation flow is sequential, according to Figure 2-6. Then, the Job Transformation will be executed only if a specific condition is met.

Left Screenshot: ms_properti table

Showing rows 0 - 49 (50 total, Query took 0.0003 seconds.)

SELECT * FROM "ms_properti"

Extra options

ID_Properti	Nama_rumah	Jenis_rumah	Harga
3	Rumah 3	Tipe A	220000
7	Rumah 7	Tipe A	240000
1	Rumah 1	Tipe A	250000
9	Rumah 9	Tipe A	260000
5	Rumah 5	Tipe A	270000
4	Rumah 4	Tipe B	280000
11	Rumah 11	Tipe A	280000
8	Rumah 8	Tipe B	290000
2	Rumah 2	Tipe B	300000
13	Rumah 13	Tipe A	300000
6	Rumah 6	Tipe B	310000
15	Rumah 15	Tipe A	310000
10	Rumah 10	Tipe B	320000
12	Rumah 12	Tipe B	330000
17	Rumah 17	Tipe A	330000
14	Rumah 14	Tipe B	340000
19	Rumah 19	Tipe A	340000
16	Rumah 16	Tipe B	350000
18	Rumah 18	Tipe B	360000
21	Rumah 21	Tipe A	360000

Right Screenshot: ms_lokasi table

Showing rows 0 - 49 (50 total, Query took 0.0004 seconds.)

SELECT * FROM "ms_lokasi"

Extra options

ID_Lokasi	Nama_Lokasi	Koordinat_GPS
1	Lokasi 1	1.2345,2.3456
2	Lokasi 2	2.3456,3.4567
3	Lokasi 3	3.4567,4.5678
4	Lokasi 4	4.5678,5.6789
5	Lokasi 5	5.6789,6.7890
6	Lokasi 6	6.7890,7.8901
7	Lokasi 7	7.8901,8.9012
8	Lokasi 8	8.9012,9.0123
9	Lokasi 9	9.0123,1.1234
10	Lokasi 10	1.1234,2.2345
11	Lokasi 11	2.2345,3.3456
12	Lokasi 12	3.3456,4.4567
13	Lokasi 13	4.4567,5.5678
14	Lokasi 14	5.5678,6.6789
15	Lokasi 15	6.6789,7.7890
16	Lokasi 16	7.7890,8.8901
17	Lokasi 17	8.8901,9.9012
18	Lokasi 18	9.9012,1.0123
19	Lokasi 19	1.0123,2.1234

Figure 8. Database loading to PhpMyAdmin.

Figure 8 is the display when the data has been successfully loaded into the phpMyAdmin database. After the Transformation is done by cleaning the data, the table from the transformation results will be loaded into the database. The property, price, demographic, city, and market price tables are processed through the transform stage into ms_properti, ms_location, ms_market_price, ms_city, and ms_demography. This table has been cleaned and tidied through the transform stage and entered into the phpMyAdmin database.

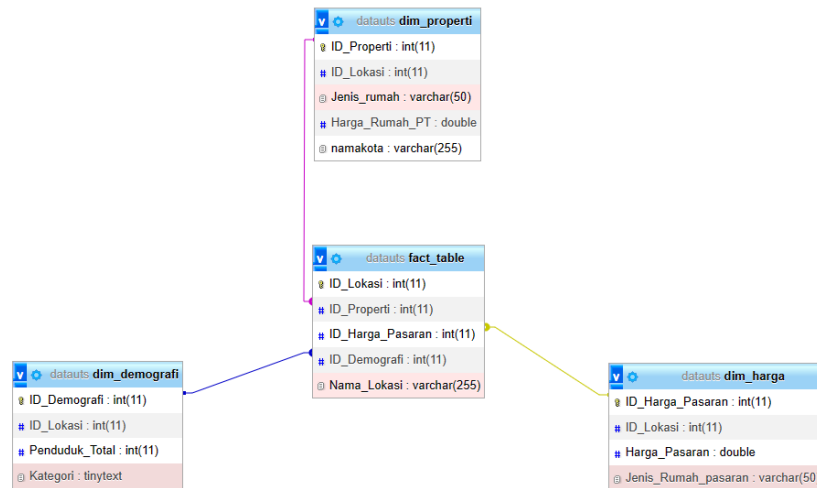


Figure 9. Star Schema with fact table and dimensional table.

Figure 9 is a display of the star schema. In Star Schema, you can see the relationship between several dimensional tables (dim_price, dim_properti, and dim_demography). Dimensional tables can be connected because the fact table has the primary key of each dimensional table. A star schema can be formed to improve query performance, making data analysis easier.

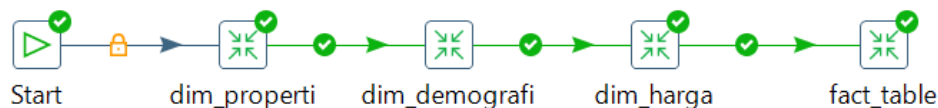


Figure 10. Star Schema on Job.

Figure 10 is a display of Jobs Transformation with star schema on PDI. The job consists of a fact table containing data related to several dimension tables, namely dim_properti, dim_demography, and dim_price. Job Transformation has been successfully carried out with a green checklist indicating the correct process.

#	ID_Lokasi	ID_Properti	ID_Harga_Pasaran	ID_Demografi	Jenis_rumah	namakota	Nama_Lokasi	Penduduk_Total	Kategori	Harga_Rumah_PT	Harga_Pasaran
1	1	1	1	1	Tipe A	Kota 1	Lokasi 1	50000	Tinggi	250000.0	240000.0
2	2	2	2	2	Tipe B	Kota 2	Lokasi 2	60000	Tinggi	300000.0	280000.0
3	3	3	3	3	Tipe A	Kota 3	Lokasi 3	55000	Tinggi	220000.0	230000.0
4	4	4	4	4	Tipe B	Kota 4	Lokasi 4	52000	Tinggi	280000.0	270000.0
5	5	5	5	5	Tipe A	Kota 5	Lokasi 5	65000	Tinggi	270000.0	220000.0
6	6	6	6	6	Tipe B	Kota 6	Lokasi 6	68000	Tinggi	310000.0	260000.0
7	7	7	7	7	Tipe A	Kota 7	Lokasi 7	53000	Tinggi	240000.0	210000.0
8	8	8	8	8	Tipe B	Kota 8	Lokasi 8	56000	Tinggi	290000.0	250000.0
9	9	9	9	9	Tipe A	Kota 9	Lokasi 9	59000	Tinggi	260000.0	200000.0
10	10	10	10	10	Tipe B	Kota 10	Lokasi 10	54000	Tinggi	320000.0	240000.0

Figure 11. Output from Fact Table.

Based on the output from the fact table above, it can be seen that there is a column that displays the total population in specific locations and cities in the location_name and Total_Population columns. This lets you see which locations can be targeted for home marketing. Companies can look at the information in the table and prepare to market in busy locations. Then, there are low, medium and high categories, which make it easier to see the total population in a particular location. Therefore, by managing this data, a table can display strategic locations based on the population of marketing houses. The more strategic the location, the more house sales will increase and provide PT profits. Wijaya Karya Realty. The output results can help PT—Wijaya Karya Realty solve problems in determining strategic locations in marketing houses.

Slicer:

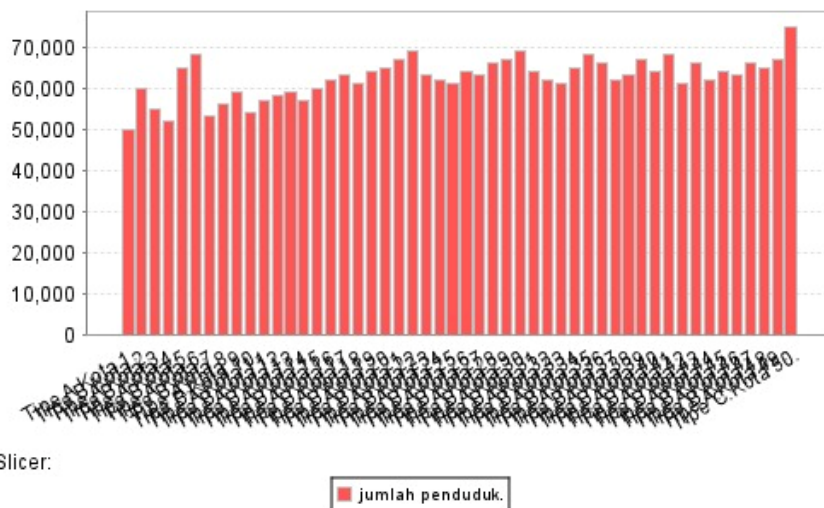


Figure 12. Results from Mondrian using Workbench.

Figure 12 is a view of Mondrian after using the workbench scheme. Mondrian's results can help companies determine strategic locations when marketing their homes. Data visualization in a bar chart helps make analysis more accessible.

Slicer:

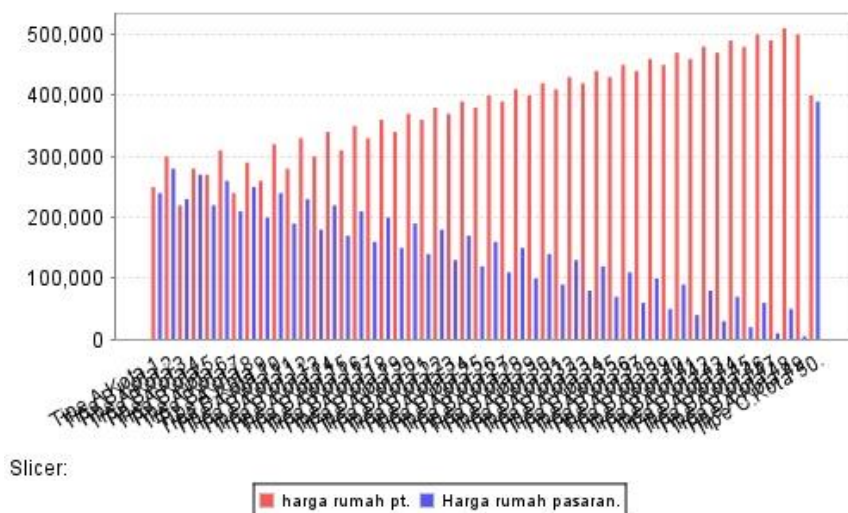


Figure 13. Results from Mondrian using Workbench.

Figure 13 is the second view of Mondrian after using the workbench scheme. The Mondrian results above can help companies compare the prices of houses sold by the company with market house prices in certain cities or locations. Visualization in a bar chart of two colours representing company and market house prices makes it easier to analyse data. The data that has been processed and visualized can be used as a reference for companies in determining house prices in specific locations in the future.

CONCLUSION

The results of the data management process and the use of a data warehouse at the company PT. Wijaya Karya Realty shows that implementing a Data Warehouse with Star Schema can help clean and integrate data in a better format. Data Warehouse makes data more structured and consistent, making it easier to analyze and manage data. Star schema also helps companies analyze data to solve problems. In addition, using Mondrian with a workbench scheme increases the company's ability to analyze better. The implementation of Star Schema helps companies provide solutions to make it easier to choose the most strategic location in marketing houses and compare the prices of the homes the company is selling with the homes on the market in that location.

For further research development, it is recommended that more advanced Datawarehouse tools be used to simplify the data processing process. Apart from that, the number of datasets used is still limited, so there is still potential for further research. Apart from location and price factors, more in-depth research should be done on property developer factors.

ACKNOWLEDGEMENTS

We extend our heartfelt gratitude to Universitas Multimedia Nusantara for their invaluable support, which played a pivotal role in the successful completion of this research endeavor. Their substantial contribution was instrumental in achieving our objectives, and we are deeply grateful for their unwavering assistance.

REFERENCES

- Awiti, J., Vaisman, A., & Zimányi, E. (2019). From Conceptual to Logical ETL Design Using BPMN and Relational Algebra. In C. Ordonez, I.-Y. Song, G. Anderst-Kotsis, A. M. Tjoa, & I. Khalil (Eds.), *Big Data Analytics and Knowledge Discovery* (pp. 299–309). Springer International Publishing.
- Awiti, J., Vaisman, A., & Zimányi, E. (2020). Design and implementation of ETL processes using BPMN and relational algebra. *Data Knowl. Eng.*, 129, 101837. <https://doi.org/10.1016/j.datak.2020.101837>
- Effendy, M. R., Kusumasari, T. F., & Hasibuan, M. A. (2019). Star Schema Implementation For Monitoring in Data Quality Management Tool (A Case Study at A Government Agency). *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 1–6. <https://doi.org/10.1109/ICIC47613.2019.8985695>
- Hahn, S. M. L. (2019). Analysis of Existing Concepts of Optimization of ETL-Processes. In R. Silhavy (Ed.), *Software Engineering Methods in Intelligent Algorithms* (pp. 62–76). Springer International Publishing.
- Handika, I. P. S., Tama, G. B. A., & Krisnayanti, N. P. M. (2020). Penerapan Teknologi Datawarehouse Nosql Dan Business Intelligence Untuk Analisa Transaksi Penjualan. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, 3(2), 120–127. <https://doi.org/10.31598/jurnalresistor.v3i2.626>
- Haryono, E. M., Fahmi, Tri W, A. S., Gunawan, I., Nizar Hidayanto, A., & Rahardja, U. (2020). Comparison of the E-LT vs ETL Method in Data Warehouse Implementation: A

- Qualitative Study. 2020 *International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 115–120. <https://doi.org/10.1109/ICIMCIS51567.2020.9354284>
- Jayashree, G., & Priya, C. (2019). Design of visibility for order lifecycle using datawarehouse. *International Journal of Engineering and Advanced Technology*, 8(6), 4700–4707. <https://doi.org/10.35940/ijeat.F9171.088619>
- Kannapadang, D., & Ta'dung, Y. L. (2022). Faktor-Faktor Yang Mempengaruhi Perubahan Laba Pada PT. Wijaya Karya Bangunan Gedung Tbk. Yang Terdaftar Di Bursa Efek Indonesia. *Fair Value: Jurnal Ilmiah Akuntansi Dan Keuangan*, 5(5), 2376–2382. <https://doi.org/10.32670/fairvalue.v5i5.2763>
- Mathur, S., Bali, V., Gupta, S. L., & Pahwa, P. (2019). Data warehouse testing and security: A conspectus. *International Journal of Innovative Technology and Exploring Engineering*, 8(9 Special Issue), 843–850. <https://doi.org/10.35940/ijitee.I1136.0789S19>
- Paliwal, M., & Saraswat, P. (2022). Approaches of Data Warehousing and Their Applications: a Review. *International Journal of Innovative Research in Computer Science & Technology*, 1, 117–121. <https://doi.org/10.55524/ijircst.2022.10.1.21>
- Pratama, I. P. A. E., & Widhiasih, N. P. N. D. (2020). Perancangan Data Warehouse Untuk Prediksi Penjualan Pada Orba Express Menggunakan Pentaho. *JUSS (Jurnal Sains Dan Sistem Informasi)*, 3(2), 43–48. <https://doi.org/10.22437/juss.v3i2.8147>
- Setiadi, T. F., & Rahutomo, R. (2022). A Data Warehouse Schema for Monitoring Regional COVID-19 Case Registration. 2022 *International Conference on Information Management and Technology (ICIMTech)*, 144–148. <https://doi.org/10.1109/ICIMTech55957.2022.9915222>
- Setiawan, J., Gousander, V., & Prasetiawan, I. (2023). Unmasking the Sentiments of Labuan Bajo: An Instagram-based Analysis for Tourism Insights through VADER Sentiment Analysis. *G-Tech: Jurnal Teknologi Terapan*, 7(3), 967–976. <https://doi.org/10.33379/gtech.v7i3.2615>
- Zahra, F., Wardhani, D., & Wiratama, J. (n.d.). *Improving the Quality of Service : ETL Implementation on Data Warehouse at Pharmacy Industry*. 18(1), 1–14.