Journal of Information Technology and Computer Science (INTECOMS)
Volume 8 Nomor 5, Tahun 2025
e-ISSN : 2614-1574
p-ISSN : 2621-3249

DESAIN DATA WAREHOUSE UNTUK PREDIKSI SEGMENTASI PELANGGAN DI PT X

DESIGN OF A DATA WAREHOUSE FOR CUSTOMER SEGMENTATION PREDICTION AT PT X

Valerie Lawrence[1], Jap Tji Beng[2,5*], Wasino[2], Sri Tiatri[3,5], Rahmiyana Nurkholiza[1,5], Tasya Mulia Salsabila[4,5], Fasia Meta Sefira[1,5]
Information Systems Undergraduate Program, Universitas Tarumanagara, Jakarta, Indonesia[1]
Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia[2]
Faculty of Psychology, Universitas Tarumanagara, Jakarta, Indonesia[3]
Faculty of Computer Science, Universitas Indonesia, Jakarta, Indonesia[4]
Science, Technology and Society Research Centre, Universitas Tarumanagara, Jakarta, Indonesia[5]
t.jap@untar.ac.id[2*]

ABSTRACT

With the rapid growth of the tourism industry, particularly within the hotel and airline sectors, understanding customer behavior and segmentation has become crucial for driving strategic decisions. However, customer segmentation in these sectors often faces challenges due to the complexity of transactional data and varying service demands. This study addresses these challenges by designing a data warehouse using a Star Schema architecture, aimed at integrating hotel and airline booking data from 2023 to 2024. The methodology follows the Nine-Step Kimball approach, coupled with the Extract, Transform, Load (ETL) process, to transform transactional data from OLTP systems into a format suitable for Online Analytical Processing (OLAP). The data warehouse features fact and dimension tables that support in-depth analysis of customer segments, booking trends, and performance across service categories and time periods. Key findings show that the data warehouse significantly improves the ability to segment customers based on booking behavior, seasonality, and service preferences, leading to more effective decision-making. This research is distinguished by its integrated analytical framework that combines hotel and airline data, providing a comprehensive perspective on tourist activities. This integrated system provides a robust foundation for data-driven strategies, allowing businesses in the tourism industry to optimize customer targeting, improve service offerings, and enhance operational efficiency.

Keywords: *Data Warehouse, Star Schema, ETL, Hotel Booking, Flight*

ABSTRAK

Dengan pertumbuhan pesat industri pariwisata, khususnya di sektor perhotelan dan penerbangan, pemahaman terhadap perilaku pelanggan dan segmentasi menjadi sangat penting untuk mengambil keputusan strategis. Namun, segmentasi pelanggan di sektor-sektor ini seringkali menghadapi tantangan akibat kompleksitas data transaksi dan permintaan layanan yang bervariasi. Studi ini mengatasi tantangan tersebut dengan merancang data warehouse menggunakan arsitektur Star Schema, yang bertujuan untuk mengintegrasikan data pemesanan hotel dan penerbangan dari tahun 2023 hingga 2024. Metodologi ini mengikuti pendekatan Nine-Step Kimball, dikombinasikan dengan proses Extract, Transform, Load (ETL), untuk mengubah data transaksi dari sistem OLTP menjadi format yang sesuai untuk Online Analytical Processing (OLAP). Data warehouse ini dilengkapi dengan tabel fakta dan dimensi yang mendukung analisis mendalam tentang segmen pelanggan, tren pemesanan, dan kinerja di berbagai kategori layanan dan periode waktu. Temuan utama menunjukkan bahwa data warehouse secara signifikan meningkatkan kemampuan untuk mensegmentasi pelanggan berdasarkan perilaku pemesanan, musiman, dan preferensi layanan, yang mengarah pada pengambilan keputusan yang lebih efektif. Penelitian ini berbeda karena menggunakan pendekatan analitis terintegrasi yang menggabungkan data hotel dan maskapai penerbangan, sehingga memberikan gambaran komprehensif tentang aktivitas pariwisata. Sistem terintegrasi ini menyediakan landasan yang kokoh untuk strategi berbasis data, memungkinkan bisnis di industri pariwisata untuk mengoptimalkan penargetan pelanggan, meningkatkan penawaran layanan, dan meningkatkan efisiensi operasional.

Kata Kunci: Data Warehouse, Star Schema, ETL, Pemesanan Hotel, Penerbangan

INTRODUCTION

The rapidly growing tourism industry, particularly in the hotel and aviation sectors, requires enhanced data integration to support more accurate and strategic decision-making. Information technology plays a vital role in improving customer data management by enabling integrated and automated systems (Sutanto et al., 2023). These systems allow companies to make smarter decisions based on reliable, real-time data. PT X, a company within this sector, manages vast amounts of hotel and flight booking data. However, this data is fragmented across various operational systems, leading to prolonged reporting times and an inability to generate a unified view of customer travel patterns, which directly hinders effective segmentation and marketing efforts. The pace of technological advancement in society is driving rapid innovation in the industrial sector (Aurelia et al., 2023). One key solution is the use of data warehouses, which consolidate disparate data sources to provide fast, accurate insights (Girsang et al., 2017). Among data warehouse architectures, the Star Schema is recognized for its ability to improve query performance and simplify data relationships compared to the Snowflake Schema, particularly in managing historical data that informs decision-making (Mohammed, 2019). Understanding and predicting consumer behavior is crucial for improving customer satisfaction and creating effective marketing strategies (Mustakim et al., 2024). Using data from databases to build systems that can predict future trends or user needs, as well as automatically adjust and create content that is consistent with user traceability and continuous monitoring results (Florido et al., 2024).

PT X faces a significant challenge: its existing data management system is not fully optimized to provide accessible, actionable data. As a result, the company lacks the ability to efficiently analyze customer behavior and segment the market in meaningful ways. While numerous studies have applied data warehousing for customer segmentation, most focus on a single industry vertical. This creates a significant research gap in developing integrated dimensional models that can the customer journey across multiple, related tourism sectors. To address this issue, the company requires a data integration solution that enables comprehensive analysis of its booking data. A data warehouse, specifically designed with a Star Schema architecture, will allow PT X to better predict business trends, maintain high-quality services, improve competitiveness, and ensure long-term sustainability (Prabawa et al., 2019). Data warehouses offer numerous benefits, such as enhancing Business Intelligence, improving data quality and consistency, saving time, and enabling efficient analysis of historical data (Wiratama & Bagioyuwono, 2023).

The primary objective of this research is to design, implement, and evaluate a data warehouse with a star schema architecture that integrates disparate hotel and airline booking data to enable advanced customer segmentation and trend analysis for PT X. Using the Nine-Step Kimball methodology, this study will apply the ETL (Extract, Transform, Load) process to implement the Star Schema (Syaputra, 2022). This evaluate will focus on key performance metrics, including query execution time, data consistency, and the systems effectiveness in identifying actionable customer segments based on integrated ordering behavior. The research will enable PT X to gain significant benefits, including improved data management efficiency and the ability to analyze booking trends by service category and time period. This approach directly supports the company's goal of optimizing business decision-making, particularly in the area of customer segmentation, which is a crucial component of marketing strategies in the tourism industry (Lawrence et al., 2024).

METHODS

Data Collection

Data collection for this study was carried out through observation and interviews. The primary data used consists of raw data provided by PT X in .csv format, for flight bookings covering the period from 2023 to 2024. Since the focus of this research is on processing and analyzing existing data, the secondary data collection method is highly suitable. Additionally, interviews were conducted through direct question-and-answer sessions to gather supplementary insights. These interviews involved key stakeholders, including the project manager and the development team at PT X, to better understand the company's data requirements. Based on the interview findings, PT X expressed the need for a data visualization tool to analyze and support the prediction of customer segmentation for hotel and flight bookings via the X application. Consequently, the researcher designed a data warehouse based on the Star Schema to enhance the efficiency and effectiveness of data analysis, ultimately leading to the creation of a comprehensive monitoring dashboard for business decision-making

Nine-Step Kimball

The Nine-Step Methodology theory by Ralph Kimball is an approach to data warehouse development aimed at producing an efficient data warehouse system that meets business needs. This method provides a step-by-step guide for building a data warehouse focused on dimensional modelling (Pratama & Raharja, 2023; Suta et al., 2019). Kimball's Nine-Step Methodology, as described in The Data Warehouse Toolkit, includes the following steps in designing a data warehouse (Kimbal & Ross, 2013):

1. Choose the Business Process: The first step is to identify the core business processes to be supported by the data warehouse. This decision is based on an understanding of the organization's business needs and the available data.
2. Choose the Grain: Determine the level of detail or granularity of the fact table. Choosing the right level of granularity is essential to ensure the flexibility of the data warehouse.
3. Identify and Conform to the Dimensions: Dimensions are descriptive data that provide context to the facts in the fact table.
4. Choose the Facts: Facts are quantitative data to be analyzed, such as sales amounts, items sold, or transaction counts. Facts must align with the previously defined granularity.
5. Build the Star Schema: The star schema organizes the fact table in the center and connects it with dimension tables. This design facilitates efficient data organization and processing. The resulting schema is visualized in figure 1, showing the fact_booking table surrounded by dim_customer, dim_service, dim_time, and dim_location.
6. Preliminary Calculations**:** Before loading data, initial transformations such as calculating derived fields are performed.
7. Design the ETL System: ETL (Extract, Transform, Load) is the process of extracting raw data from source systems, transforming it to meet the data warehouse requirements, and loading it into the target schema. This study utilizes SQL Server Integration Services (SSIS) as the primary ETL tool to automate the data flow from the source system to the data warehouse.
8. Create Aggregates: To optimize performance, aggregate tables are created to summarize data. These tables help improve query response times.
9. Implement and Maintain the System: Finally, the data warehouse is implemented, and ongoing maintenance processes such as updates and ETL system management are established.

Extract Transform Load (ETL)

The ETL (Extract, Transform, Load) process in the context of a data warehouse is a crucial mechanism that connects heterogeneous data source systems to a structured data warehouse environment, making it ready for business analysis. The ETL process plays a vital role in transforming raw data from operational systems into accessible and usable information for business decision-makers (Caserta & Kimball, 2013). For this implementation, the ETL process was executed using Microsoft SQL Server Integration Services (SSIS), which provided a robust platform for designing and managing the complex workflow of extracting data from CSV files, transforming it to fit the star schema, and loading it into a Microsoft SQL Server data warehouse. The core data management process consists of three main stages (Palmer, 2024):

1. Extract: Data is collected from various sources, such as relational databases, APIs, files, or web services. This stage aims to gather raw data from diverse sources.
2. Transform: During this stage, the extracted data is converted into a format that aligns with business requirements. Transformation processes can include data cleansing, data integration, normalization, or aggregation.
3. Load: After transformation, the data is loaded into the data warehouse or other data storage systems, enabling further analysis by users.

Data Warehouse Design

This study used the Star Schema for data warehouse design, chosen for its simplicity and efficiency in supporting fast query performance and clear relationships between data elements. The fact table in the Star Schema contains key metrics such as booking counts and revenue, while the dimension tables include descriptive information such as customer details, service categories, and time periods. Data analysis methods employed in this study include trend analysis, customer segmentation analysis, and historical data reporting, all of which provide valuable insights into booking behaviors, service preferences, and seasonal trends. The core data warehouse and OLAP system provide the basic data for descriptive analysis and predictive analysis for customer segmentation. By implementing this data warehouse solution, PT X will be able to streamline its data management processes, improve decision-making capabilities, and enhance its ability to predict and analyze customer behavior for better marketing strategies and operational planning.

RESULT AND DISCUSSION

Using the Nine-Step Methodology by Kimball, the research begins by selecting the primary business process—hotel and flight service bookings. The fact table, FACT_PREDIKSI, is designed to include summarized transactional data such as customer ID, travel type ID, booking type ID, destination ID, date ID, invoice code, total revenue, and total bookings. This design aligns with the study's hypothesis that a well-structured data warehouse will enhance the ability to analyze and predict booking trends in the tourism industry.

To support this analysis, several dimension tables are created. DIM_CUSTOMER stores customer details, such as name, age, and gender; DIM_TRAVEL_TYPE categorizes various travel purposes or modes; DIM_BOOKING_TYPE outlines different booking methods or categories; DIM_DESTINATION records destination cities for analyzing travel patterns; DIM_TIME stores time-related data, such as dates, months, and years, enabling time-based analysis; and DIM_INVOICE captures payment-related information. These dimensions are designed to optimize data analysis across various categories such as travel type, destination, booking type, and time period.

Each dimension table's attributes are carefully defined to facilitate effective data analysis. For instance, DIM_DESTINATION contains destination city information, and DIM_TRAVEL_TYPE categorizes travel modes or purposes. The FACT_PREDIKSI fact table links all dimension tables by storing reference keys and essential transactional data. Aggregation processes based on these dimensions allow for performance analysis of services over specific time periods or by service categories. This approach is advanced compared to previous studies that focused on a single sector, providing integrated insights into both the hotel and airline industries. Unlike the airline specific data warehouse implementation by Smith et al. (2022), which achieved query performance improvements, our cross sector integration resulted in improvements that enable more comprehensive customer journey analysis.

The final stage of the Nine-Step Kimball Methodology involves testing the system with end users, including business teams or stakeholders, to ensure the data meets the requirements for decision-making and analysis. The end result is a structured data warehouse built using the Star Schema architecture, which efficiently supports the analysis of booking trends and service performance in the tourism industry.

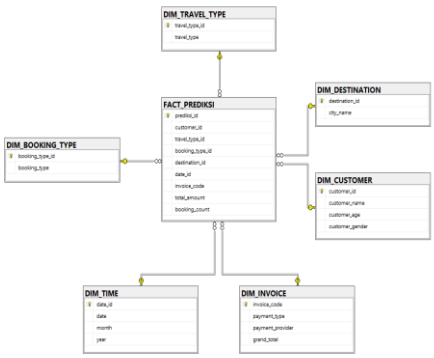

Figure 1. Star Schema
Source: Personal Documentation

CONCLUSION

The results of this study clearly demonstrate that the data warehouse, designed using the Star Schema, successfully addresses the hypothesis that a centralized, well-organized system improves the ability to analyze booking trends and customer behavior. The data warehouse provides accurate, accessible, and reliable data that can be used by PT X for informed decision-making in the hotel and airline sectors. The design and implementation of the system ensure that results can be easily accessed, and that insights gained from the data are reliable for supporting strategic decisions.

However, this study has several limitations, its scope is limited to a single company case study, it uses batch processed historical data rather than real time streaming data, and it applies a descriptive rather than predictive analytical approach.

By enabling in-depth analysis of booking trends, customer segmentation, and service performance, the results contribute to a more effective and efficient decision-making process in the tourism industry, thereby validating the proposed solution and its practical application.

REFERENCES

Aurelia, A., Wasino, W., Chandra, D., & Jap, T. B. (2023). Developing Website-Based Information System Applications to Map PT. XYZ's Properties Using Next. JS Framework with Haversine Method. *International Journal of Application on Sciences, Technology and Engineering*, *1*(1), 59-64.

Caserta, J., & Kimball, R. (2013). *The Data Warehouseetl Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley.

Florido-Benítez, L. (2024). Generative artificial intelligence: a proactive and creative tool to achieve hyper-segmentation and hyper-personalization in the tourism industry. *International Journal of Tourism Cities*.

Girsang, A. S., Satya, D., Isa, S. M., Al Fariz, S., Susilo, B., Ramdani, D., & Lian, M. (2017, November). Decision support system using data warehouse for hotel reservation system. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 369-373). IEEE.

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons.

Lawrence, V., Limbor, E. G., Putri, T. A., Wasino, W., & Beng, J. T. (2024). Perancangan Dashboard Transaksi dan Segmentasi Pelanggan untuk Pemesanan Layanan Travel PT Y. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, *13*(3).

Mustakim, N. A., Aziz, M. A., & Rahman, S. A. (2024). Predicting Consumer Behavior in E-Commerce Using Decision Tree: A Case Study in Malaysia. *Information Management and Business Review*, *16*(3), 201-209.

Mohammed, K. (2019). Data warehouse design and implementation based on star schema vs. snowflake schema. International Journal of Academic Research in Business and Social Sciences, 9.

Palmer, M. (2024). Technical guide: Understanding ETL data pipelines for modern data architectures.

Prabawa, I. N. A., Agung, D., Arimbawa, K., & Janardana, I. G. N. (2019). Analysis and Design Data Warehouse For E-Travel Business Optimization. Int. J. Eng. Emerg. Technol, 4(1).

Pratama, I. P. A. E., & Raharja, I. M. S. (2023). Business Intelligence Based on Kimball Nine-Steps Methodology for Monitoring the Feasibility of Goods in Market. *International Journal of Advances in Data and Information Systems*, *4*(2), 135-144.

Sutanto, R., Wasino, & Beng, J. T. (2023, December). Developing a sales system for Ko Ginhan's special chicken porridge with point of sale system. In *AIP Conference Proceedings* (Vol. 2680, No. 1, p. 020061). AIP Publishing LLC.

Suta, I. B. L. M., Mahendra, I. G. N. A. S., & Sudarmojo, Y. P. (2019). Design General Hospital Data Warehouse Base on Nine Step Methodology. *IJEET (International J. Eng. Emerg. Technol*, *4*(1), 15-19.

Syaputra, M. D. (2022). DATA WAREHOUSE DESIGN FOR SALES TRANSACTION ON CV. SUMBER TIRTA ANUGERAH. *DATA WAREHOUSE DESIGN FOR SALES TRANSACTION ON CV. SUMBER TIRTA ANUGERAH*, *8*(2), 88-96.

Wiratama, J., & Bagioyuwono, M. A. (2023). Improving the Data Management: ETL Implementation on Data Warehouse at Indonesian Vehicle Insurance Industry. *International Journal of Science, Technology & Management*, *4*(5), 1256-1268.