Published online on the page : https://ejournal.uinbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Design and Development of an Online Analytical Processing (OLAP) Application for Customer Profiling Analysis of Insurance "X"

Debleng Puja Kesuma Wardanie[1], Slamet Kacung[*2], Chamdan Fauzi[3], Pamudi[4]*

[1,2,,3,4] *Universitas Dr. Soetomo, Surabaya,* Indonesia

Article Information

Article History:

Submitted: November 23, 2024
Revision: November 29, 2024
Accepted: December 13, 2024
Published: December 24, 2024

Keywords

Online Analytical Processing (OLAP),
Customer Profile,
Multidimensional Analysis,
Decision Making,
Insurance

Correspondence

E-mail: slamet@unitomo.ac.id*

A B S T R A C T

The system's slow and inflexible response time is a characteristic of analytical processes based on transactional databases (OLTP), as experienced by PT Asuransi "X." This limitation arises because transactional databases are not designed for OLAP, which can provide various functions to perform synthesis and analysis that improve response time. This study aims to design and develop an Online Analytical Processing (OLAP) application to be used for customer profiling analysis at insurance company "X." In the insurance industry, effective and efficient data analysis is essential to understand customer behavior, perform segmentation, and make more informed decisions in marketing insurance products. The OLAP application developed in this study integrates various customer data dimensions, such as demographics, claim history, and owned products, facilitating multidimensional analysis for its users. The application design process includes system design, data collection, OLAP technology implementation, and application testing. The study results indicate that the application reveals that the majority of customers are male (56%), aged between 30 and 45 years (45%), and employed in the private sector. Additionally, in the city of Surabaya, there is a higher tendency to purchase the Mitra Sakinah life insurance policy. This information enables the company to better understand customer demographic characteristics and tailor its marketing strategies accordingly.

1. Introduction

Insurance companies face complex challenges in managing large and diverse customer data. This data includes demographic information, claim history, and products purchased by customers. Proper data processing is crucial for enhancing customer service and improving the company's operational efficiency. In this regard, data analysis techniques capable of processing information multidimensionally can provide an effective solution, one of which is through the use of Online Analytical Processing (OLAP) applications [1], [2].

OLAP (Online Analytical Processing) is a technology that enables users to analyze data from various perspectives and perform analytical processing rapidly. Through OLAP applications, insurance companies can segment customers, analyze behavior patterns, and evaluate product performance more efficiently. This provides a clearer understanding of customer profiles and supports better data-driven decision-making [3], [4].

However, despite the widespread recognition of OLAP technology, its implementation in the insurance industry remains limited, particularly in the context of customer profiling analysis. Therefore, this study aims to design and develop an OLAP application for customer profiling analysis at insurance company "X."

With this application, the company is expected to more easily evaluate customer performance and identify new opportunities in marketing insurance products[5].

OLAP enables users to view and analyze data from various perspectives or dimensions. As a crucial component of business analytics, OLAP storage methods play a significant role in enhancing the performance and efficiency of OLAP systems. These storage methods aim to support analytical operations such as fast query processing, data aggregation, and interactivity in data-driven decision-making. There are several primary OLAP storage methods, each with its advantages and limitations tailored to the specific needs of users and organizations [6], [7].

1.1. Multidimensional OLAP (MOLAP)

MOLAP (Multidimensional Online Analytical Processing) is one of the most widely used OLAP storage methods. In MOLAP, multidimensional data is stored in cube structures or multidimensional tables, enabling quick access to pre-aggregated data. One of the key advantages of MOLAP is its high-performance capability in answering queries involving large-scale data aggregation.[1][8].

1.2. Relational OLAP (ROLAP)

Unlike MOLAP, ROLAP (Relational Online Analytical Processing) stores data in more traditional relational databases. In ROLAP, data is not stored in a multidimensional format but rather in relational tables. When users execute OLAP queries, ROLAP dynamically aggregates data through SQL and queries on the relational tables[9].

1.3. Hybrid OLAP (HOLAP)

HOLAP (Hybrid Online Analytical Processing) combines the strengths of MOLAP and ROLAP by storing large volumes of data in relational databases (like ROLAP) while calculating and storing data aggregations or summaries in multidimensional structures (like MOLAP). This approach provides a balance between performance and flexibility. HOLAP efficiently handles large datasets while allowing fast access to aggregated data. However, implementing and maintaining a HOLAP system tends to be more complex and requires greater resources[10].

A Decision Support System (DSS) is an information system used to assist decision-making within an organization or company. DSS provides relevant and structured information to support the process of making complex decisions [11]. The benefits of an effective Decision Support System (DSS) include significant budget reductions, such as cost savings in promotions, better understanding of the decision-making environment, more effective decision-making, and improved information value.

A data mart stores data sourced from various databases or other data sources, which can be utilized by healthcare companies to meet their specific needs. In business operations, a data mart is a component of a data warehouse that aids in generating reports and conducting data analysis [12], [13]. This data can originate from various sources, including internal company sources such as operational systems and external sources such as market data. Through multiple data ports, a data mart can provide a theoretical overview of dataset components. The ETL (Extract, Transform, and Load) method is employed to process data within the data mart[14]. This entails selecting and processing data from various sources, transforming it into a format suitable for processing in the data mart, and having the capability to load or present information to users.

A data warehouse is a system designed to integrate, store, and analyze data from various sources to support an organization's data-driven decision-making. A data warehouse is an integrated data collection that is subject-oriented, non-volatile, and time-variant. Its structure differs from regular operational systems as it focuses on analyzing historical and in-depth data, aiding strategic decision-making processes across industries such as finance, healthcare, retail, and insurance[15]. A data warehouse is not a single product[16], but an environment that usually consists of various products as seen in figure 1, which includes :

a. Data Sources: The origins from which data will be extracted.
b. Transmission Tools: Tools used to extract and transfer data into the data warehouse.
c. Data Mart and Data Warehouse: The storage and organizational systems for the integrated data.
d. Desktop Query and Reporting Tools: Analytical tools used for querying, reporting, and supporting decision-making.

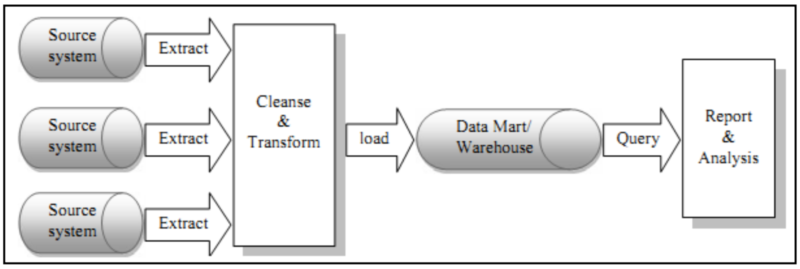

Figure 1. Data Warehouse Environment [17]

Source System: The data source is extracted from predefined transaction databases of the insurance company based on the company's business needs.

Cleanse and Transform: Before source data is loaded into the data warehouse, data cleansing is performed to ensure the data meets expectations.

Data Warehouse: A collection of data composed of several dimensional tables and fact tables ready for processing in the OLAP system.

Report and Analysis: Multidimensional data reports generated from OLAP, featuring drill-up and drill-down capabilities.

2. Method

The research methodology begins with selecting the data required by the organization, which is then stored in a temporary table and subsequently transferred to a staging table before being saved into the data warehouse for reporting and analysis. As illustrated in Figure 2.

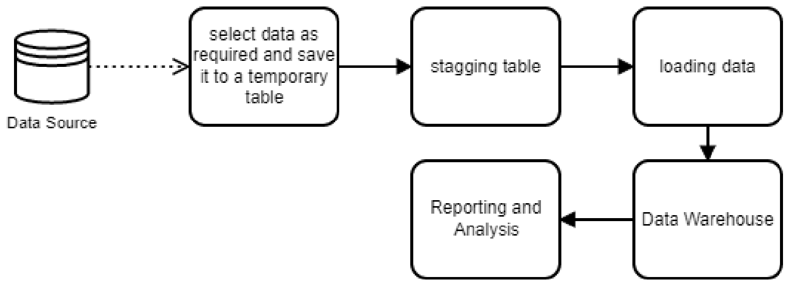

Figure 2. Research Stages

2.1. Data Source and Data Selection

The data source is derived from the transactional database of the insurance company as a foundation for generating insights that can be utilized for decision-making, such as customer profiling. The appropriate use of data sources will result in more relevant and accurate reports and analyses, as shown in Table 1.

Table 1. Insurance Transactional Database

No.	Table Name	Number of Attributes
1	Customer	15
2	Customer_Class	2
3	Product	3
4	Premium	7
5	Policy	7
6	Agent	5
7	Claim	8
8	Claimed	3
9	Employee	12
10	Area	3

2.2. *Staging Table*

The staging table consolidates all data in the warehouse, including dimensional tables and fact tables. This table does not adhere to normalization principles, and while missing values are eliminated, redundancy is present. This design is intentional to accommodate the need for faster processing to produce analyses, which is not achievable with transactional tables.

2.3. *Loading Data*

ECTL (Extract, Cleanse, Transform, Load) is a process for building a data warehouse that enables organizations to collect data from various sources, transform it to meet reporting and analysis needs, and then load the data into the data warehouse. The ECTL process is illustrated in Figure 3.

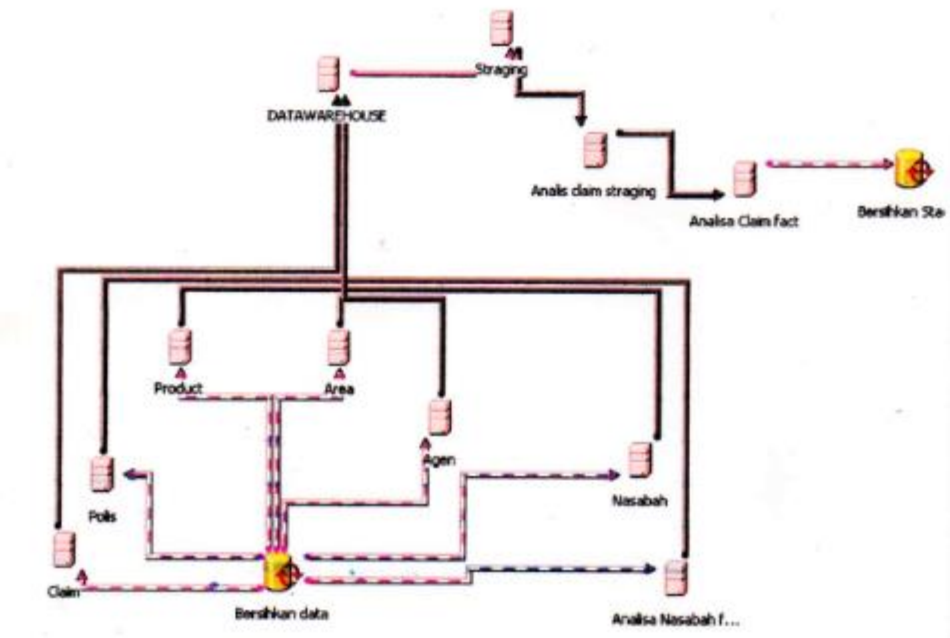

Figure 3. ECTL process.

The process design shown in Figure 3 summarizes the ECTL process in a data warehouse. By implementing these steps, organizations can ensure that the data in the data warehouse is of high quality and ready to be used for better decision-making.

2.3.1. ***Extract :*** *Taking data from various sources in this case taking data that is in accordance with the analysis needs of transactional data. Such as:*

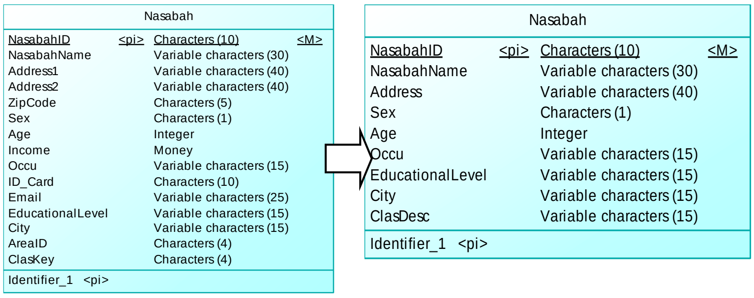

Figure 4. Data Selection to Extract

a) ***Cleanse :*** The cleansing process in a data warehouse is a critical step aimed at ensuring the quality and consistency of data before it is used for analysis and reporting. This process involves identifying and correcting or removing inaccurate, duplicate, or incomplete data [18], [19] or inconsistent before being extracted to the data warehouse. Such as :

SELECT *
FROM StaginOfAssurance
WHERE (ClaimID is Null) or (Tanggal_Claim is Null) or (Besar_Kerugian is Null)
Order by ProduckID, ClaimID

	PolisID	AreaID	NasabahID	ProduckID	ClaimID	Tanggal_Claim	Besar_Kerugian
1	P0047	1009	2049	4001	NULL	NULL	NULL
2	P0139	1036	2141	4001	NULL	NULL	NULL
3	P0309	1008	2310	4001	NULL	NULL	NULL
4	P0421	1040	2422	4001	NULL	NULL	NULL
5	P0440	1058	2441	4001	NULL	NULL	NULL
6	P0502	1182	2503	4001	NULL	NULL	NULL
7	P0562	1041	2563	4001	NULL	NULL	NULL
8	P0048	1010	2050	4002	NULL	NULL	NULL
9	P0170	1054	2172	4002	NULL	NULL	NULL
10	P0187	1088	2189	4002	NULL	NULL	NULL
11	P0204	1122	2206	4002	NULL	NULL	NULL
12	P0209	1132	2211	4002	NULL	NULL	NULL
13	P0230	1174	2232	4002	NULL	NULL	NULL

Figure 5. Contents of the StaginOfAssurance Table Before Cleansing

DELETE FROM
StaginOfAssurance
Where ClaimID is Null or Tanggal_Claim is Null or Besar_Kerugian is Null
SELECT * FROM StaginOfAssurance
Order ClaimID

	PolisID	AreaID	NasabahID	ProduckID	ClaimID	Tanggal_Claim	Besar_Kerugian
1	P0001	1001	2001	4001	CL000001	2004-01-01 00:00:00.000	500000.0000
2	P0002	1001	2002	4002	CL000002	2003-03-30 00:00:00.000	750000.0000
3	P0003	1001	2003	4003	CL000003	2003-07-25 00:00:00.000	250000.0000
4	P0004	1002	2004	4004	CL000004	2003-07-20 00:00:00.000	8000000.0000
5	P0005	1002	2005	4005	CL000005	2004-05-05 00:00:00.000	500000.0000
6	P0006	1002	2006	4011	CL000006	2004-10-07 00:00:00.000	10000000.0000
7	P0007	1003	2007	4019	CL000007	2003-01-01 00:00:00.000	1200000.0000
8	P0008	1003	2008	4020	CL000008	2003-01-05 00:00:00.000	5000000.0000
9	P0009	1003	2009	4009	CL000009	2003-01-09 00:00:00.000	2500000.0000
10	P0010	1004	2010	4010	CL000010	2003-01-12 00:00:00.000	1500000.0000
11	P0011	1004	2011	4016	CL000011	2003-01-15 00:00:00.000	2000000.0000
12	P0012	1004	2012	4015	CL000012	2003-01-20 00:00:00.000	1750000.0000
13	P0013	1005	2013	4013	CL000013	2003-01-25 00:00:00.000	1950000.0000

Figure 6. Contents of the StaginOfAssurance Table After Cleansing

b) ***Transform :*** After data cleansing, the next step is the transformation process, which involves converting the data to meet the requirements for customer analysis and claims payment analysis before being extracted into the data warehouse. This includes data aggregation, such as calculating totals, averages, or other relevant metrics.

c) ***Load :*** Loading is the final stage in the ETL process. After data has been extracted, cleansed, and transformed, the last step is loading the extracted data from the staging area into the data warehouse for further analysis [20]. This process consists of:

a. Initial load

The initial loading involves importing the entire dataset from various sources into the data warehouse for the first time. This process is typically performed in bulk.

b. Incremental load

This process involves loading new or updated data since the last loading process. It is performed periodically to keep the data warehouse synchronized with the data sources

2.4. *Data warehouse*

Building a data warehouse is a mechanism for determining how data will be organized, stored, and accessed by users through the creation of a data model. In data warehouse design, the data model is typically divided into two main categories: identifying dimensional tables and fact tables in designing a star schema [21]. At this stage, identification and adjustment of dimensions related to the fact table is carried out [4]. The dimension table used in the insurance customer profiling analysis is :

Table 2. Dimension and Fact Table

No.	Dimension Name	Description
1	Product Dimension	Contains the type of insurance product
2	Claim Dimension	Contains the claim subcategory
3	Policy Dimension	Contains the type of policy used by the customer
4	Area Dimension	Contains the location of the customer's distribution
5	Customer Dimension	Contains customer profile data
6	Time Dimension	Contains analysis based on a certain time, namely date, week, month, quarter, year.
7	Agent Dimension	Contains a list of insurance agents
8	Claim Payment Fact	Contains the amount of loss due to claims and when the claim occurred
9	Customer Analysis Fact	Contains the amount of premium value from customer payments

Star Schema

Star Schema design is used in OLAP to create multidimensional structures, the characteristics of star schema are:

1. The center of the star schema is the fact table
   a. The fact table contains relevant KPIs of information and time objects.
   b. The KPIs are attributes of the fact table
   c. Information and time objects are elements of the key.
2. The center of the star schema is the dimension table
   a. Dimension tables contain data about information or time objects
   b. The fact table and dimension tables are connected through multipart primaries in the fact table
   c. Each dimension table is directly connected to the fact table through a key column.

According to the results of the system analysis, two star schemas are needed, namely the Customer Analysis star schema and the Claim Payment Analysis star schema as seen in Figures 5 and 6 below :

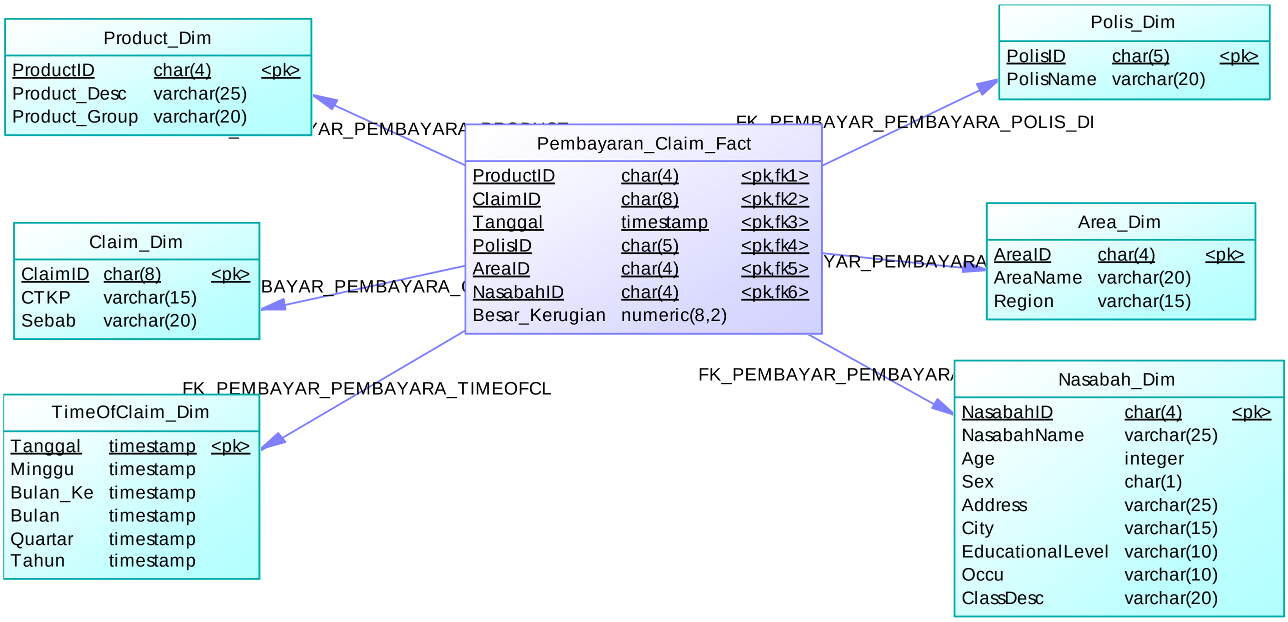

Figure 7. Star Schema Claim Payment

The Star Schema in Figure 7 illustrates that data sourced from the transactional database is interconnected with insurance claim payments, consisting of one fact table and six dimension tables. The Star Schema in Figure 8 demonstrates that data sourced from the transactional database is related to insurance customer profiling analysis, with one fact table and five dimension tables. This analysis is aimed at determining the premium value generated by the organization.

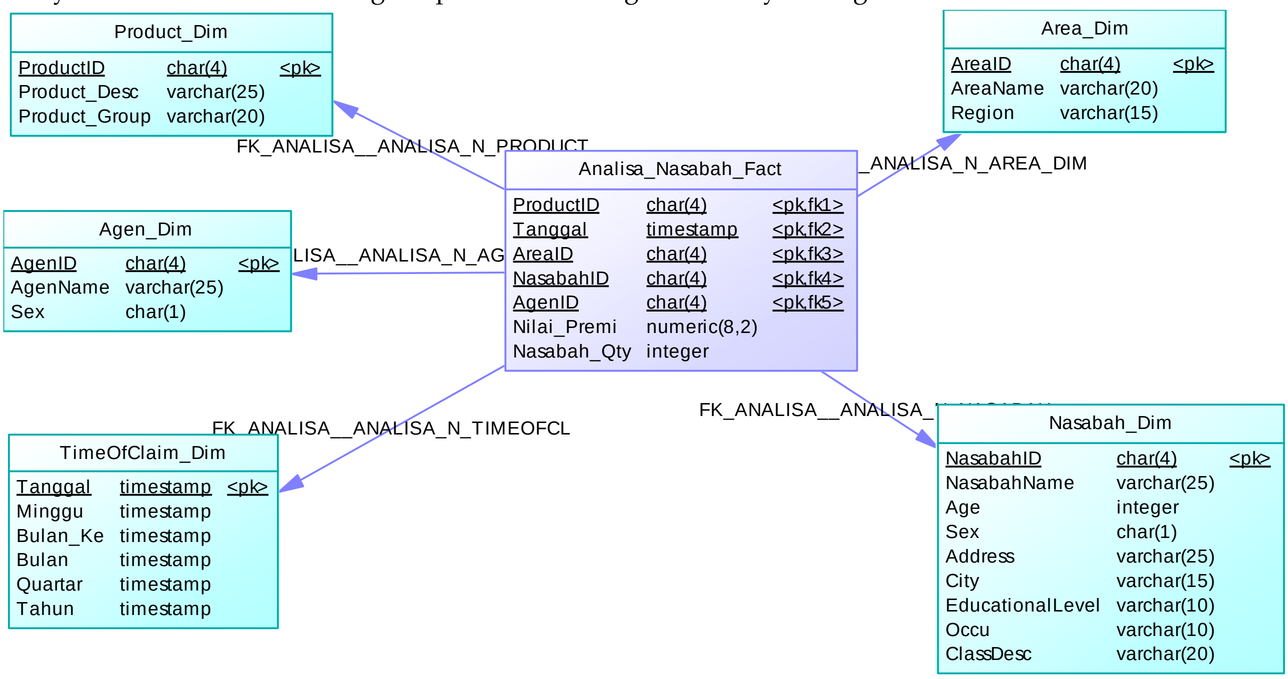

Figure 8. Star Schema Customer Profiling Analysis

The multidimensional design begins with the identification of dimensions in Table 3, which defines how users view the data and perform "drill-down" analysis from the highest level of data into more detailed data in the "drill-up" analysis. This is followed by the creation of the cube design.

Table 3. Multi-Dimensional Design Table

No.	Dimension Name	Level and Hierarchy	Type	Structure
1	Customer	2-City, Customer Name	Standart	Regular
2	Products	1-Product Name	Standart	Regular
3	Areas	2-Area Name, Region	Standart	Regular
4	Claim	1-Cause	Standart	Regular
5	TimeofClaime	3-Year, Quarter, Month	Standart	Regular
6	Education Level	1-Education Level	Standart	Regular
7	Gender	1-Sex	Standart	Regular
8	Customer_Class	1-Class Desc	Standart	Virtual Nasabah
9	Occu	1-Occu	Standart	Virtual Nasabah
10	Product_Class	1-Class Product	Standart	Virtual Product
11	Policy	1-Policy Name	Standart	Regular
12	Agent	1-Agent Name	Standart	Regular

The design of **Customer_Class**, **Occu**, and **Product_Class** with a virtual structure aims to save storage space for multidimensional data. In the customer analysis cube list, calculated members are used to compute values between one cube and another without increasing the storage size of the cubes, as these values are not physically stored.

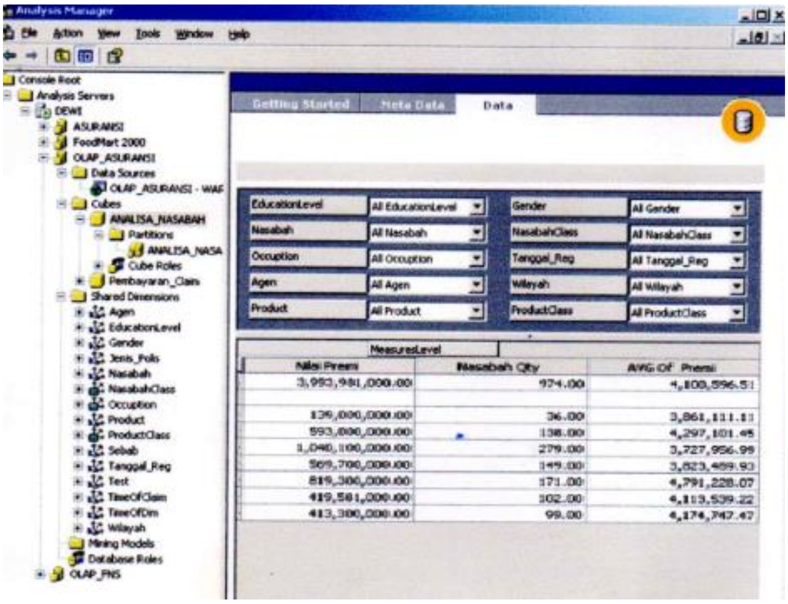

Figure 9. Results of Multidimensional Data Analysis of Customer Profiling

Table 3. Multi-Dimensional Design Table

No.	Cube Name	Dimension	Type	Measure	Calculate Member
1	Customer Analysis	11 Dimensions: Agent, Customer, Customer Class, Education Level, Gender, Area, Registration Date, Occu, Product, Product Class, Time of Dim	Regular	Premium Value and Customer Qty	Avg Premium = Premium Value / Customer Qty
2	Claim Payment Analysis	10 Dimensions: TimeOfClaim, Policy, Education Level, Cause, Gender, Customer, Customer Class, Occu, Product, Product Class	Regular	Loss	-

The aggregation design involves pre-calculated summary data, enabling faster response times for queries and analyses[22], [23], thus, the server does not need to retrieve the entire dataset. The determination of aggregation in SQL Server can be accessed through **Analysis Manager > Cube Editor > Tools > Design Storage**, as shown in Figure 10.

.

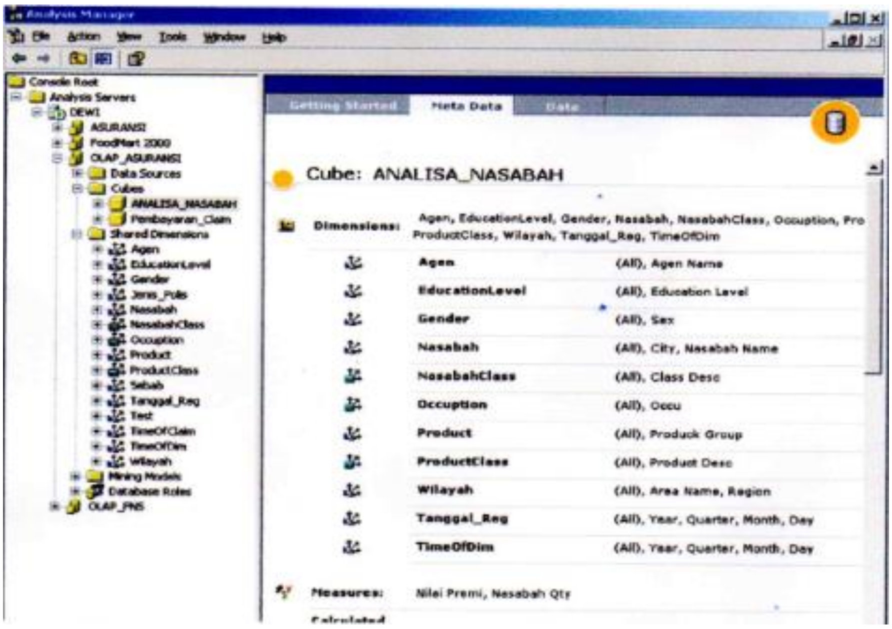

Figure 10. Manager Analysis Design

The aggregation for the customer analysis cube, from the initial data warehouse load, was set at a 90% performance improvement level, showing the optimal balance between performance and size, requiring 8.4 MB of storage, as shown in Figure 11. Meanwhile, the aggregation for the claim payment analysis cube, from the initial data warehouse load, was also set at a 90% performance improvement level, achieving the best balance between performance and size, requiring only 0.4 MB of storage, as depicted in Figure 12.

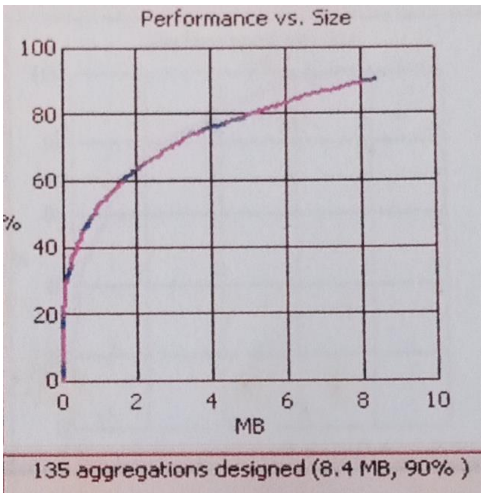

Figure 11 Customer Analysis Cube Aggregation

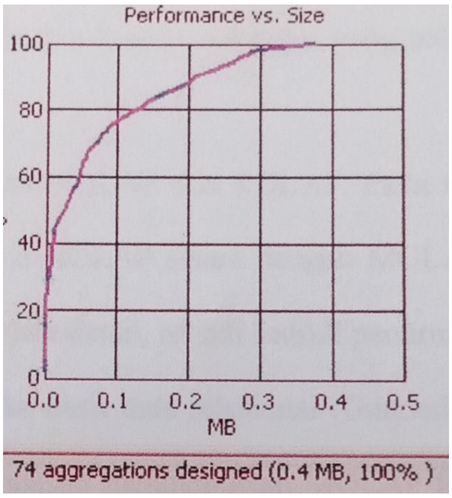

Figure 12 Claim Cube Aggregation

3. Reporting and Analysis

This stage illustrates the reporting and analysis processes of the data warehouse using SQL Server. At the core, the architecture of the data warehouse is represented, where various data sources flow through the ETL (Extract, Transform, Load) process. The reporting section is depicted by various reports and visualizations created using SQL Server Reporting Services. Figure 13 showcases an interactive dashboard and tabular reports that leverage data from cubes, enabling users to perform drill-downs, apply filters, and conduct in-depth analyses.

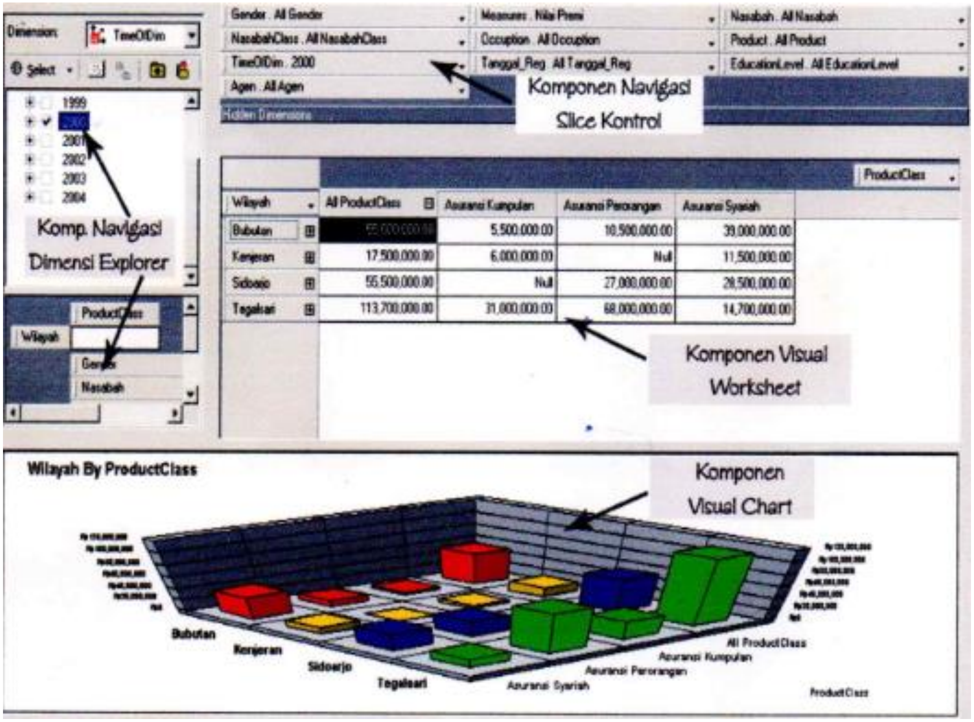

Figure 13. Customer Analysis Interface Design

4. Results and Discussion

This stage illustrates the reporting and analysis processes of the data warehouse using SQL Server. At the core, the architecture of the data warehouse is represented, where various data sources flow through the ETL (Extract, Transform, Load) process. The reporting section is depicted by various reports and visualizations created using SQL Server Reporting Services. Figure 13 showcases an interactive dashboard and tabular reports that leverage data from cubes, enabling users to perform drill-downs, apply filters, and conduct in-depth analyses.

4.1. Result

There are several features developed in this OLAP application:

a. Multidimensional data analysis developed in the form of an interactive dashboard with pivot table graphs.

b. Data visualization can be filtered in a follow-up and drill-down manner to dig up more detailed information based on insurance type and certain demographics.

Figure 11 indicates that the total number of customers is 974, consisting of 545 male customers and 429 female customers. These are further categorized as follows: Individual customers: 388 (217 males and 171 females), Organizations/NGOs: 157 (84 males and 73 females), Companies: 51 (29 males and 22 females), Schools: 127 (65 males and 62 females).

.

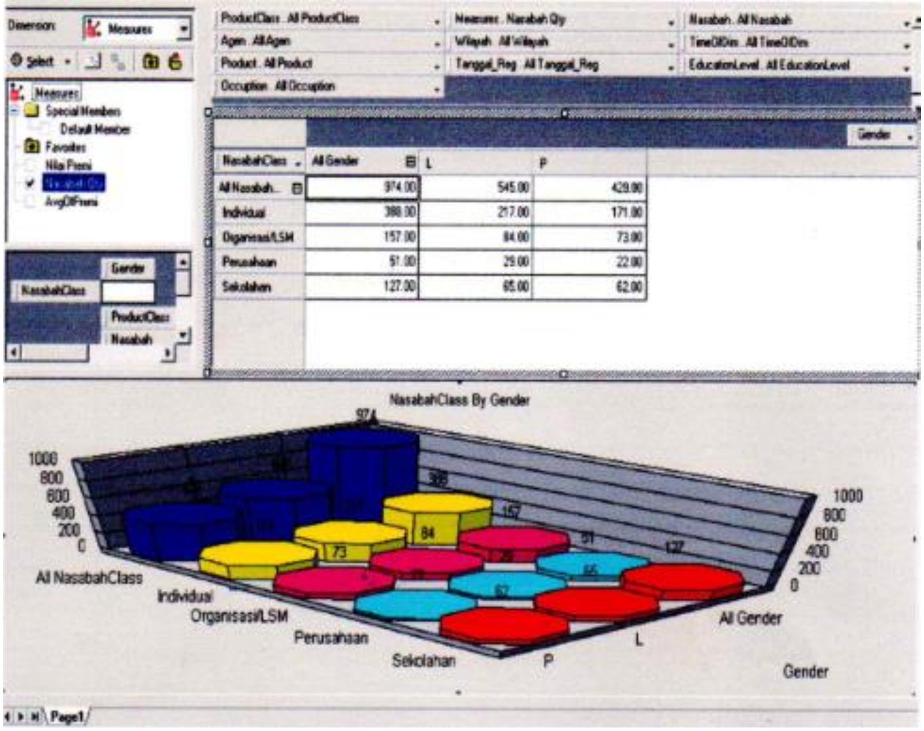

Figure 14. Customer Demographic Analysis

The data in Figure 14 shows that the number of customers in the Bubutan area is 54 customers (16 using group products, 21 individual insurance, and 17 customers using Sharia insurance). While the Semampir area has 45 customers (15 people using group insurance, 20 individual insurance, and 10 Sharia insurance. The number of customers in the Tegalsari area is 73 customers (23 using group products, 38 individual insurance, and 15 customers using Sharia insurance).

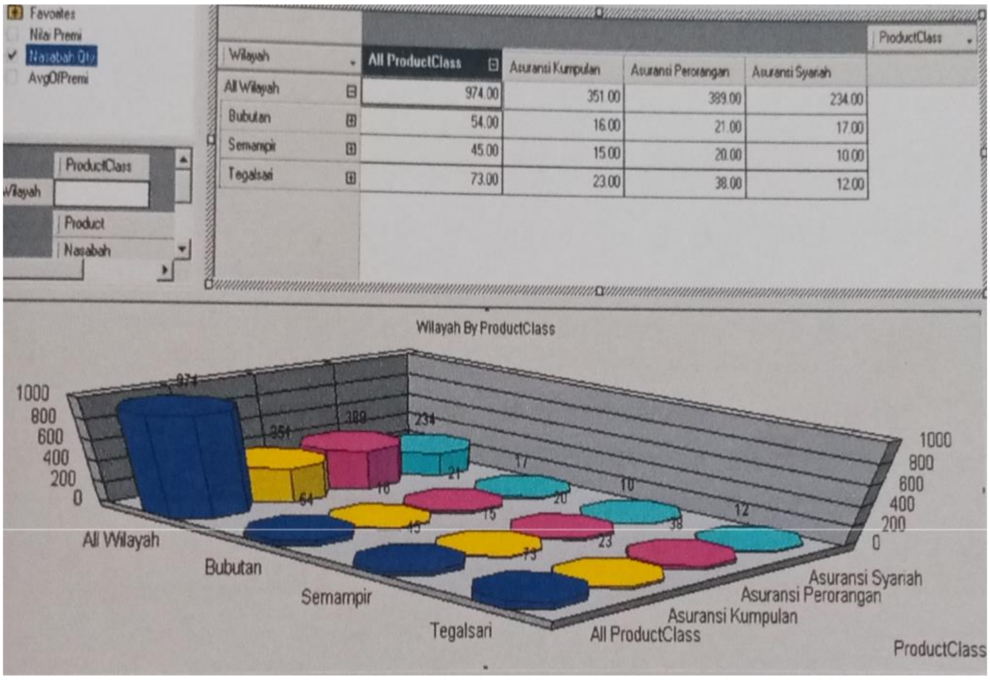

Figure 15. Product Profiling Analysis of Area Customers

The data in Figure 16 shows that there are 974 customers in Surabaya, distributed across several areas. Among them, Gubeng has 61 customers, Krembangan has 58 customers, Simokerto has 56 customers, Tegal Sari has 73 customers, and other areas have additional customers.

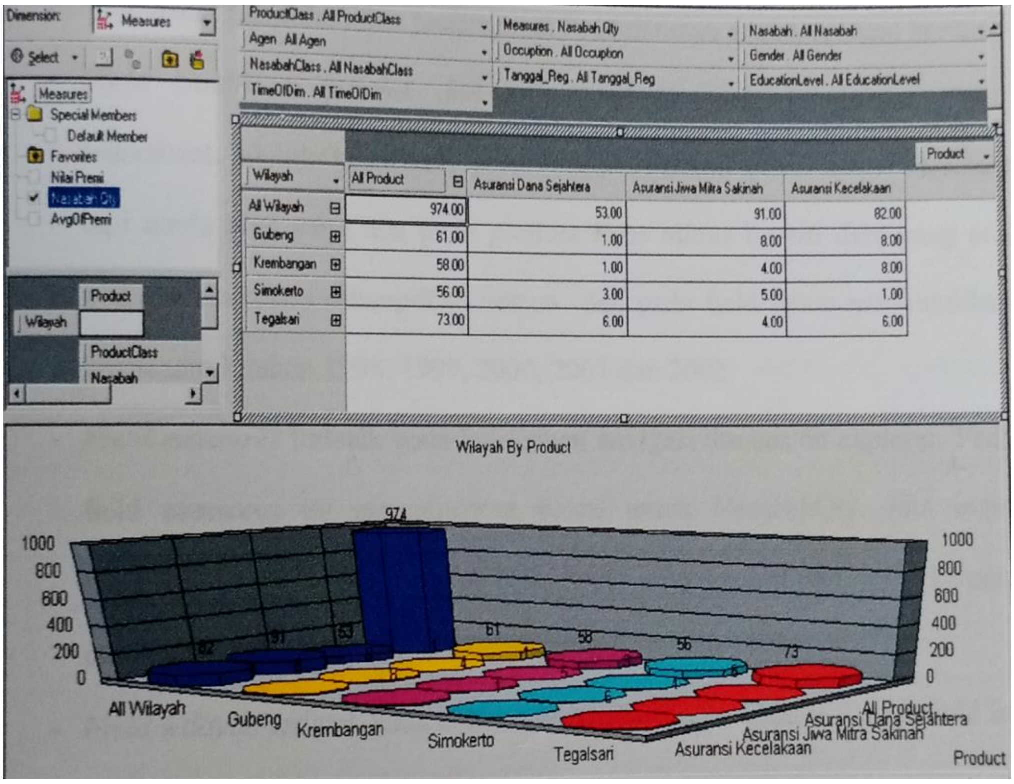

Figure 16. Customer Area Product Profiling Analysis

The data in the claim analysis cube in Figure 17 shows that the largest total claim payment was due to death, amounting to 256.8 million IDR, followed by claims caused by fire at 162.8 million IDR, hit-and-run at 104.8 million IDR, and accidents/collisions at 55.450 million IDR.

All Sebab	Kebakaran	Meninggal Dunia	Tabrak Lari	Tabrakan
579,850,000.00	162,800,000.00	256,800,000.00	104,800,000.00	55,450,000.00
150,700,000.00	67,150,000.00	31,500,000.00	43,250,000.00	8,800,000.00
188,050,000.00	49,700,000.00	103,600,000.00	22,750,000.00	12,000,000.00
236,700,000.00	45,950,000.00	117,300,000.00	38,800,000.00	34,650,000.00

Figure 17. Customer Product Profiling Analysis

4.2. Discussion

The findings of this study demonstrate significant potential in supporting the strategic decision-making processes of insurance companies. Several key points from the design and development of the OLAP application for customer profiling analysis are as follows:

4.2.1. Implementation of Star Schema and OLAP Cube Design

The implementation of star schema and OLAP cube design has proven effective in accommodating multidimensional analysis requirements. Users can easily explore data and derive relevant insights promptly. For instance, the majority of customers are male (56%), aged between 30-45 years (45%), and employed in the private sector. Customers in Surabaya show a higher tendency to purchase *Mitra Sakinah* life insurance policies.

4.2.2. Insurance Product Performance

Individual insurance products contribute the largest share (40%) of premium revenue, while accident insurance exhibits the highest claim frequency, with a claim ratio reaching 70%.

4.2.3. Customer Risk Analysis

Customers aged ≥ 50 years exhibit a higher frequency of claims due to death under *Mitra Syariah* life insurance products, indicating a risk that requires careful management by the company.

4.2.4. Ease of Use

The intuitive user interface enables business analysts with minimal technical backgrounds to utilize the analytical features effectively. This aspect is critical for increasing technology adoption within insurance companies.

4.2.5. Benefits of OLAP-Based Customer Profiling:

a. The OLAP application provides faster and more detailed insights compared to conventional analysis methods.
b. Through drill-down features, companies can identify high-risk customer segments, such as elderly groups requiring specialized strategies to offer life insurance products or *Dana Sejahtera* loyalty programs.
c. Develop products more aligned with market needs.
d. Effectively target agents for more efficient marketing efforts.

4.2.6. Potential for Development:

a. AI and Machine Learning Integration: With the integration of artificial intelligence, the application can offer predictive customer behavior insights and actionable recommendations.
b. Enhanced Security: Given the sensitivity of customer data, system development must prioritize data security, including implementing encryption across all levels.

5. Conclusion

Based on the research conducted, it can be concluded that the OLAP application designed for customer profiling analysis in the insurance company offers significant potential in supporting more data-driven strategic decision-making. The implementation design of the star schema and OLAP cube has proven to be effective in accommodating multidimensional analysis, providing faster and deeper insights compared to conventional analysis methods. Customer profiling based on demographic characteristics and claims behavior enables the insurance company to better understand their market segments, identify risks that need to be managed, and adjust marketing strategies and product development to meet customer needs.

Furthermore, this application also demonstrates ease of use, enabling business analysts with minimal technical background to take advantage of its analytical features, which in turn will increase technology adoption within the company. With the drill-down feature, the company can more easily identify high-risk customers, develop products that are more aligned with market needs, and target marketing agents more effectively.

From the perspective of development potential, this research highlights significant opportunities to integrate artificial intelligence (AI) and machine learning to provide more accurate predictions of customer behavior and actionable recommendations.

References

[1] R. T. Aldisa, F. Teknologi, and U. Nasional, "Penerapan Online Analytical Processing (OLAP) dalam Pengelolaan Data Karyawan," vol. 9, no. 1, pp. 55–61, 2022.

[2] K. Smith and J. K. Cohen, "'OLAP Technology and Data Warehousing,'" *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 5, pp. 2021–2029.

[3] B. K. Easterita *et al.*, "Pengembangan Data Werehouse Dan Online Analytical Processing (Olap) Untuk Analisis Data Artikel Pada Jurnal Teknologi Informasi Dan Ilmu Komputer (Jtiik) Data Development Of Warehouse And Online Analytical Processing (Olap) For Articledata Analysis ," vol. 1, no. 1, pp. 12–22, 2020.

[4] T. Angelya *et al.*, "Perancangan Data Warehouse Online Analytical Processing (Olap) Data Hasil Kerja PT . ABC," pp. 656–664, 2023.

[5] R. Insani, "Perancangan Data Warehouse untuk Profiling Data Pelanggan pada Bidang Telekomunikasi," vol. 3, no. 2, pp. 19–30, 2021.

[6] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Wiley, 2013.

[7] R. Ben Messaoud, O. Boussaid, S. Rabas, L. Eric, and B. Cedex, "A New OLAP Aggregation Based on the AHC Technique," pp. 65–72.

[8] J. Sistem and A. Cucus, "Fakultas Ilmu Komputer."

[9] J. P. R. A. S. Arora, "ROLAP: A Flexible OLAP Method for Enterprise Business Intelligence," *Int. J. Bus. Intell.*, vol. 28, no. 3, pp. 101–109, 2014.

[10] J. P. Smith, "HOLAP: Hybrid Solution for Data Warehouse Analysis," *Bus. Intell. Rev.*, vol. 12, no. 4, pp. 39–45, 2016.

[11] N. H. Alza and H. Suryamen, "Decision Support System for Selecting Marching Band Field Commanders Using Profile Matching," vol. 03, no. 01, pp. 69–81, 2023.

[12] M. Guntara, "Analysis And Design Of Data Warehouse And Data Mart Budget," pp. 6–11.

[13] I. Zaelani, "Implementasi data Mart Terhadap sistem penjualan pada perusahaan bidang distributor di pt . Eigen implementation of data mart on sales system in distributor companies in PT . EIGEN TRIMATHEMA," vol. 1, no. November, 2021.

[14] W. S. Fana, R. Soviab, and R. Permana, "Data Warehouse Design with ETL Method (Extract , Transform , And Load) for Company Information Centre," vol. 5, no. 2, pp. 132–137, 2021.

[15] W. H. Inmon, *Building the Data Warehouse*, 4th ed. John Wiley & Sons, 2005.

[16] Harry Singh, *Data Warehousing: Concepts, Technologies, Implementations, and Management*. Prentice Hall PTR, 1998, 1999.

[17] L. P. S. Slamet Kacung, "Rancangan dan Pembuatan Data Warehouse untuk Kebutuhan Sistem Pendukung Keputusan (Studi Kasus: Departemen Marketing dan Komunikasi Universitas Dr Soetomo Surabaya)," *Palimpsest (Iowa. City).*, vol. 5, no. 1, pp. 1–10, 2014.

[18] A. Prasad and P. Dutta, "Data Cleansing Framework for Data Warehouses," in *in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2018, pp. 761–767.

[19] F. Ridzuan and W. M. N. Wan Zainon, "A review on data cleansing methods for big data," *Procedia Comput. Sci.*, vol. 161, pp. 731–738, 2019.

[20] H. W. Hah and K. Siau, "A Framework for Data Transformation in ETL Processes," in *in 2018 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2018, pp. 515–520.

[21] W. Suharso *et al.*, “IMPLEMENTASI STAR SCHEMA PADA STUDI KASUS PERPUSTAKAAN,” vol. 4, no. 1, pp. 1–11, 2021.

[22] Neni Purwati, *DATA WAREHOUSE*. Bandar Lampung: Darmajaya (DJ) Press, 2018.

[23] Data Warehousing Guide, “SQL for Aggregation in Data Warehouses,” *Oracle9i*, 2024. [Online]. Available:https://docs.oracle.com/cd/B10500_01/server.920/a96520/aggreg.htm#10981. [Accessed: 23-Nov-2024].