Data Warehouse Design For Sales Transactions on CV. Sumber Tirta Anugerah

Muhammad Dwiky Syaputra[1], Alwis Nazir[2], Siska Kurnia Gusti[3*], Suwanto Sanjaya[4], Fadhilah Syafria[5]

[1,2,3,4,5]Dept. of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim, Indonesia
[1]118501122278@students.uin-suska.ac.id, [2]alwis.nazir@uin-suska.ac.id, [3]siskakurniagusti@uin-suska.ac.id*,
[4]suwantosanjaya@uin-suska.ac.id, [5]fadhilah.syafria@uin-suska.ac.id

Abstract. Many data warehouses are implemented in companies engaged in retail, CV. Sumber Tirta Anugerah is one of the paint product retail companies that has not implemented it yet. As time goes by, the sales transaction data is getting more and more difficult to process because it is still stored in Microsoft Excel. This is a serious problem in utilizing historical data to assist in making a decision. It is difficult to store sales data because the data is quite large and a lot. Based on the above problems, a data warehouse design is needed for sales transaction data. This data warehouse design uses Kimball's nine-steps method and star schema. To perform the ETL process (extract, transform, and load) using Pentaho software. In this data warehouse design, Tableau software is used to visualize the processed data into a graph and dashboard report. The result of this research is a data warehouse design using nine steps and a star schema which gets a transformation response time of 4048 MS.

Keywords: Data Warehouse, Nine Steps, Pentaho, Star Schema, Tableau

INTRODUCTION

Data warehouse is the historical data processing of an organization or company that can be used to assist in decision making. A data warehouse is also a database designed to perform a special query to analyze in detail and create reports, the resulting output is historical data of the organization or company that is displayed in detail and has been grouped according to categories making it easier for the company or organization to search for data[1],[2],[3].

Data warehouses are widely implemented in companies engaged in retail or sales. CV. Sumber Tirta Anugerah is a small part of companies in the paint retail sector that have not used a data warehouse to store transaction data on their product sales. Currently CV. Sumber Tirta Anugerah has thousands of transaction data in Excel form, but so far it has not been fully utilized. Over time and with the development of CV. Sumber Tirta Anugerah, the greater the data that must be processed. This problem becomes a serious problem in the use of historical data to assist in making a decision[4],[5]. The large and large number of sales transaction data makes it difficult for companies to store historic data and analyze it. To solve all these problems, a data warehouse can be used as a solution to store quite a lot of historical data and can be used to assist in making strategic business decisions[6],[7],[8].

In research conducted by Hasanudin, et al about the design of the data warehouse model for the sale of materials using the star schema. This study uses star schema modeling because it is considered easy to understand and use in the process of making queries. The final result of this research is a data warehouse model that comes from various dimensions and is used to form a managerial report[9],[10],[11].

In another study regarding the use of a nine-step design methodology for business intelligence in data warehouse design by Akbar and Rahmanto. The design usea s nine-step design and star schema modeling to produce a better data warehouse design. The results in this study resulted in a faster data warehouse design which is 1278 MS[12].

Furthermore, in research conducted by Sugiarto et al about the design of a sales data warehouse at PT. SPJ. In this study, the design of the data warehouse model uses Pentaho tools whose data source comes from the

sales information system which then loads it into the database into the data warehouse. The results of this study are data warehouse designs that can assist leaders in making decisions[13].

Based on the background that has been described, research will be carried out to design a data warehouse on CV sales transaction data. Source Tirta Anugerah. In this data, the warehouse designer will use a star schema. The use of star schema in this design aims to minimize data redundancy in each dimension table in the fact table. The design of this data warehouse also uses a nine-step design methodology as its design method, and uses Pentaho Data Integration in carrying out the ETL (extract, transform, load) and Tableau processes to visualize informative data from the designed data warehouse design.

METHODS

The flow of the research below is carried out in several stages which are described in Figure 1 below.
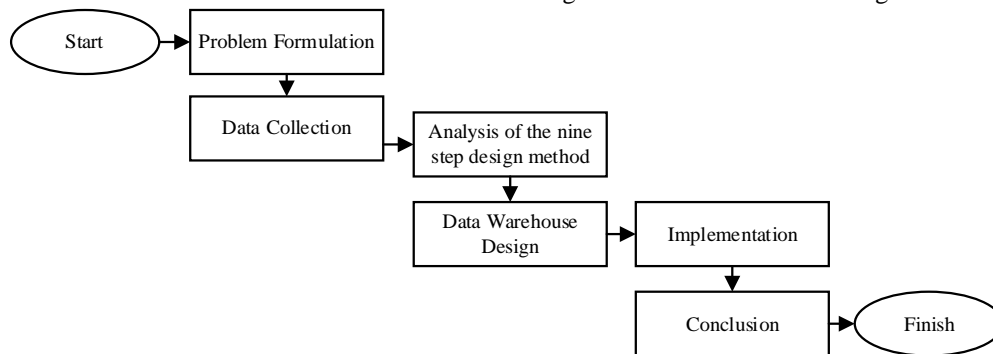

Figure 1. Research Design

A.   Problem Formulation
At this stage, the identification of the problem that becomes the research topic is carried out. The formulation of this problem was carried out to obtain the expected solution from this research. The formulation of the research problem to be studied is the design of a data warehouse on a CV. Sumber Tirta Anugerah sales transaction data.

B.   Data Collection Technique
  1. **Secondary Data**
      The data used is secondary data obtained from sales transaction data CV. Sumber Tirta Anugerah for 3 years starting from January 2019 to December 2021.
  2. **Observation**
      Researchers made observations directly to CV. Sumber Tirta Anugerah to find out how business processes are currently running and what data trends are needed in these business processes.
  3. **Interview**
      Researchers conducted interviews with the owner of the CV. Sumber Tirta Anugerah to get information about the buying and selling process at the company.
  4. **Literatur Review**
      Researchers conduct literature studies from various sources such as books, related journals, and others, to dig up more information related to research.

C.   Research Methodology
The data warehouse dimensional modeling uses the nine-step design methodology by Ralph Kimbal[14]. Here are the 9 steps of dimensional modeling:

  1. **Define Business**
      Process Defining the process means selecting the main subject of the business process.
  2. **Grain Declaration**
      Determines what a fact table will represent. The dimension table grain is defined from the dimension table grain.
  3. **Identifying Dimensions**
      Identify and relate each dimension table that is formed with the fact table that is determined later.

4. **Identifying the Facts**
   In this step determine what information is used in the fact table.
5. **Storing Initial Calculations or Calculations in a Fact Table**
   The calculation process is carried out on the fact table.
6. **Revisit the Dimension Table**
   Add detailed information to the dimension table if the attributes are still difficult to understand.
7. **Selecting the Database Duration Time Range**
   Select the storage time range used.
8. **Tracing Changes from Dimensions Slowly**
   Look for rewritten dimension attribute type changes.
9. **Deciding Query Priority and Query Type**
   This stage is centered on architectural design or physical design.

RESULT AND DISCUSSION

A. Dimensional Modelling Design

The dimensional modeling design in this study follows the 9 steps of the nine-step design method by Ralph Kimball, namely:

1. **Select and Define Business Process**
   Selecting a business process is the earliest stage. The business processes selected in this study can be seen in table 1 below:

Table 1. Select and Define Business Process

Business Proces	Description
Paint Product Sales	Report on product sales data recap that has been collected from all branches into Microsoft Excel with .xls existence

Table 1 shows the sales business process at CV Sumber Tirta Anugerah, sales data is used to assist in data processing so that it can be applied as a reference in analyzing and making decisions.

2. **Grain Declaration**
   In the business process that has been determined, the selected grain can be seen as follows:

Table 2. Grain Declaration

Grain	Description
Paint product sales information	Displays sales information of paint products by branch, sales, product, and time in sales transactions

Table 2 shows the selected grain, namely data on sales of paint products, because the information we want to know is the number of sales of paint products.

3. **Identifying Dimensions**
   Based on the results of grain identification in the previous stage, the dimensional data to be used are as follows:

Table 3. Identifying Dimensions

Dimension Table	Field
dimension_product	SK, code_product
dimension_branch	SK, code_branch
dimension_seller	SK, code_sales
dimension_time	SK, year, month, quarter, day

Table 3 displays the dimension table used in processing data based on each field in the data warehouse.

4. **Identifying the Facts**
   Based on the dimensions that have been identified, the fact table that will be created is as follows:

Table 4. Identifying the Fact

Fact Table	Field
fact_sales	sk_product sk_branch sk_seller sk_time purchase_amount total_sale

Table 4 displays a fact table for which data processing will be made based on the results of the dimension table identification.

5. **Storing Initial Calculations or Calculations in a Fact Table**

Table 5. Storing Initial Calculations In Fact Table

Preliminary Calculation Table	Field
price	code_product product name price

Table 5 shows the initial calculation table that is stored for use in fact table calculations to calculate total sales in a fact table.

6. **Revisit the Dimension Table**
   At this stage, review the dimension tables that have been identified to add attributes that are still unclear or difficult to understand.

7. **Selecting the Database Duration Time Range**
   The duration of database that the researcher uses in the sales data warehouse is data for the last 3 years from January 2019 to December 2021.

8. **Tracing Changes from Dimensions Slowly**
   At this stage, observations are made in the dimension table whether there is a change in dimensional attributes which results in attribute data being written over and over again, attribute data changing, and giving rise to alternative attributes.

9. **Deciding Query Priority and Query Type**
   At this stage, it is only centered on the architecture or physical design of the data warehouse where the sales transaction data query will be prioritized to run a query that produces a sales chart output.

B.   Data Warehouse Design

Star Schema or star schema is the schema used in the design of this research data warehouse. Sales transaction data warehouse schematic CV. Sumber Tirta Anugerah is shown in Figure 2:
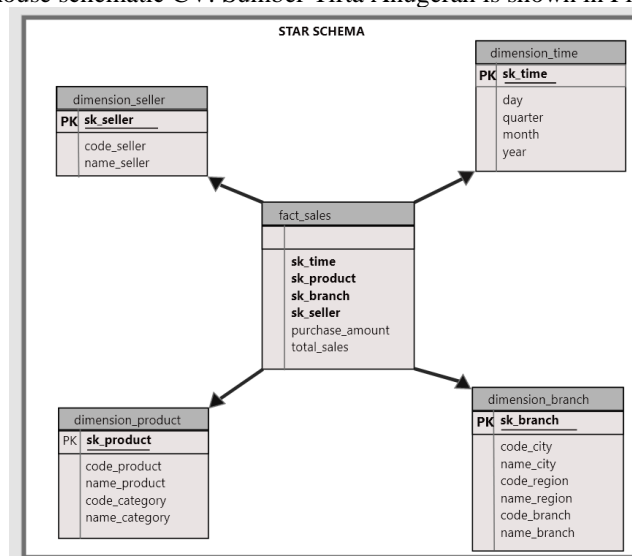

Figure 2. Star Schema Data Warehouse Sales transaction

Figure 2 shows the star schema or star schema in the data warehouse design that has 4 dimension tables, namely sales_dimension, time_dimension, product_dimension, and branch_dimension. And has 1 fact table, namely sales_fact which contains the surrogate key (SK) of each dimension, and has 2 initial calculation tables of fact tables, namely total_belief and total_jual.

C.   ETL (Extract, Transform, Load)

ETL (Extract, Transform, Load) is the process of combining data from several sources to become one new data source. After obtaining the results of dimensional modeling, the modeling will be used as a basis for processing paint product sales data. Furthermore, the sales data is extracted so that it can be transformed into dimension tables and fact tables.

The product and category data process is transformed into the product_dimension table shown in Figure 3:
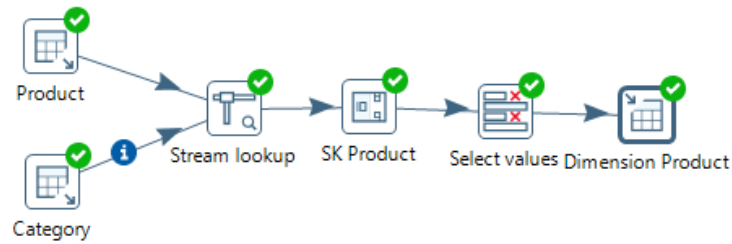

Figure 3. Transformation Process into Product Dimension

Figure 3 shows the process of product data and category data being extracted using Pentaho Software, then the data is combined, and after that it is selected and what data will be used in the product_dimension table. The sales data process is transformed into the sales_dimension table shown in Figure 4:

Figure 4. Transformation Process into Sales Dimension

Figure 4 shows the process of extracting sales data using Pentaho Software, then the data is selected and any data will be taken to be entered into the sales dimension table.

The process of branch data, regional data, and city data is transformed into the branch_dimension table shown in Figure 5:
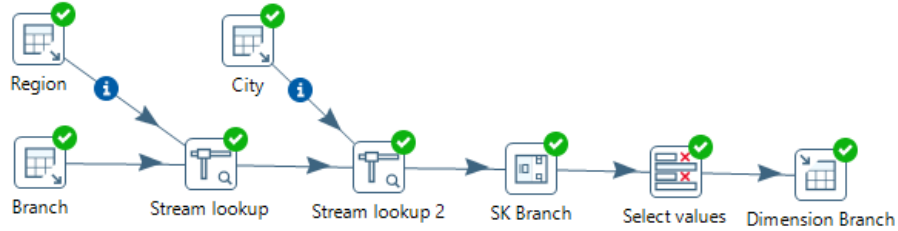

Figure 5. Process of Transformation into Dimension_Branch

Figure 5 shows the process of extracting branch data, regional data, and city data using Pentaho Software, then the data is combined, after that it is selected and what data will be entered into the branch dimension table.

The process of data time of year, month, quarter, and day is transformed into a time_dimensional table shown in Figure 6:
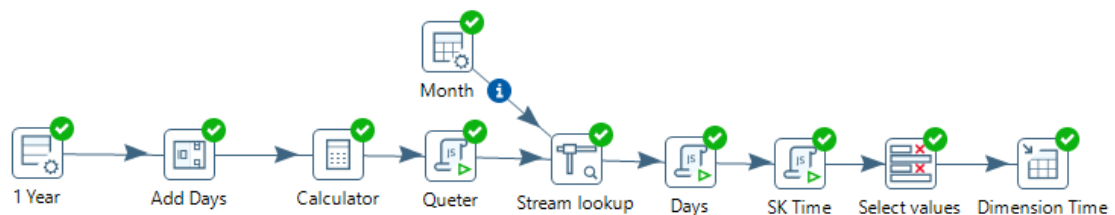

Figure 6. The Process of Transformation into the Time Dimension

Figure 6 shows the process of generating time data for the time_dimension in Pentaho Software, then the time_data is selected and any data will be taken to be entered into the time dimension.

Sales transaction data processing, branch dimensions, sales dimensions, product dimensions, time dimensions are transformed into sales facts shown in Figure 7:
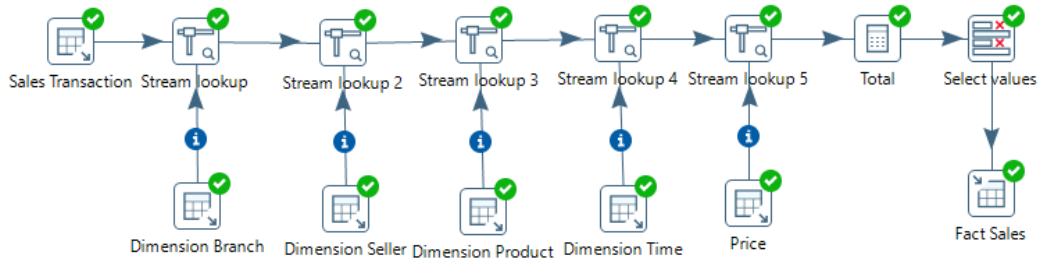
Figure 7. Dimension Transformation Process into Sales_Fact

Figure 7 shows the process of sales transaction data combined (lookup) with each dimension to retrieve the SK or surrogate key from each dimension used in the sales fact_table.

The time loading process transforms sales transaction data, branch dimensions, sales dimensions, product dimensions, and time dimensions into the sales fact_table which is shown in Figure 8:

Figure 8. Display Time Loading Transformation

Figure 8 shows loading time information on the transformation of sales transaction data, branch dimensions, sales dimensions, and product dimensions to the sales fact table in the data warehouse which has taken 4725ms.

D. Implementation

After the data warehouse design on CV. Sumber Tirta Anugerah has been successfully established, then the next process is to analyze the results using the Tableau software with OLAP (Online Analytical Processing). The following are the results of the implementation of the data warehouse design which are visualized in the form of graphs and information in the form of a dashboard using Tableau software.

1. **Graph of Total Product Sales in 2019 to 2021**

   Sales data information visualized into a line chart can be seen in Figure 9 which shows a decrease in total sales figures on CV. Sumber Tirta Anugerah for the last 3 years.
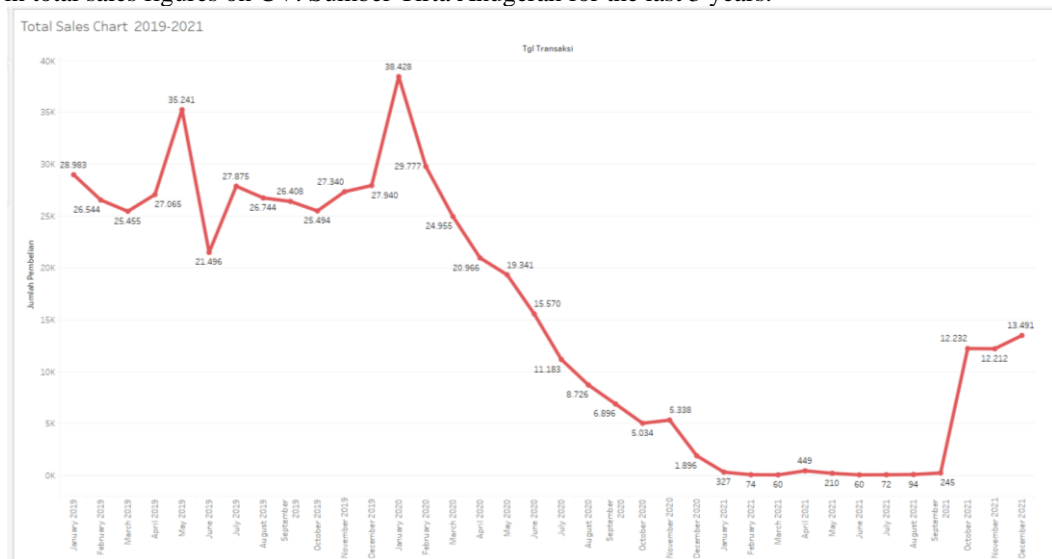

Figure 9. Graph Display of Total Sales Year 2019 to 2021

   Figure 9 shows a graph of total sales from January 2019 to December 2021, from the graph it can be seen that the sales trend for the last 3 years has decreased significantly. A significant decline in the sales trend began in February 2020 until September 2021. The highest total sales were in January 2020 with total sales of 38428 transactions and the lowest sales in March and June 2021 of 60 transactions.

93

2. **Sales Comparison Chart for Each Branch in 2019 to 2021**

Sales data information that is visualized into a bar chart can be seen in Figure 10 which shows the declining sales trend of each branch on the CV. Sumber Tirta Anugerah for the last 3 years.
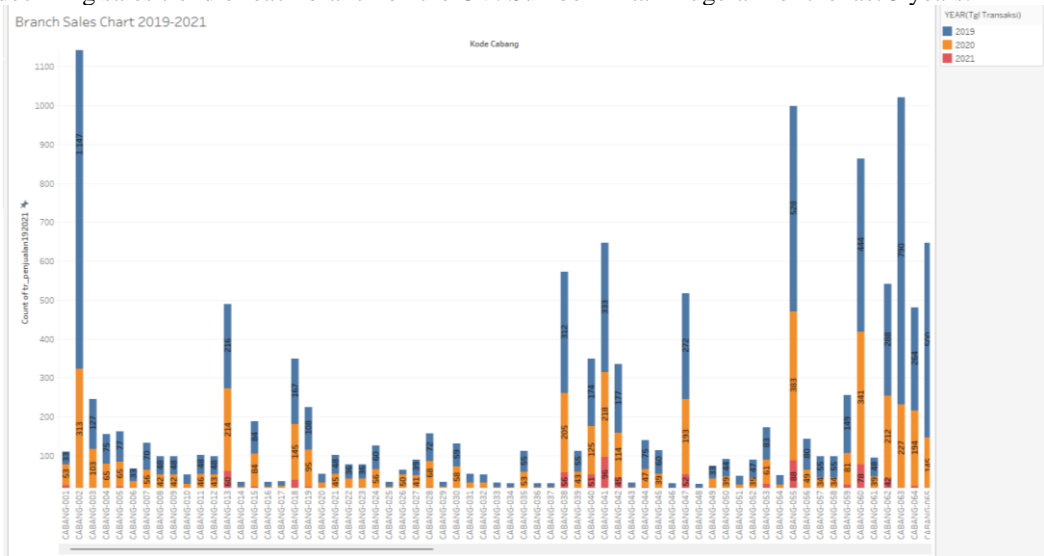

Figure 10. Branch Sales Graph Display 2019 to 2021

Figure 10 shows a graph of branch sales in 2019 to 2021, from the graph it can be seen that branch sales have decreased. The highest branch sales figure was in 2019 and from 2020 to 2021 there was a significant decline. In 2019 branch-002 experienced the highest sales figure of 1147 transactions. In 2020 the highest sales figure was at branch-055 with 383 transactions. And in 2021 the highest sales figure was only 96 transactions at branch-041.

3. **Sales Comparison Chart for Each Sales in 2019 to 2021**

Sales data information that is visualized in the form of a bar chart can be seen in Figure 11 which shows the decreasing trend of sales figures for each sales on the CV. Sumber Tirta Anugerah for the last 3 years.
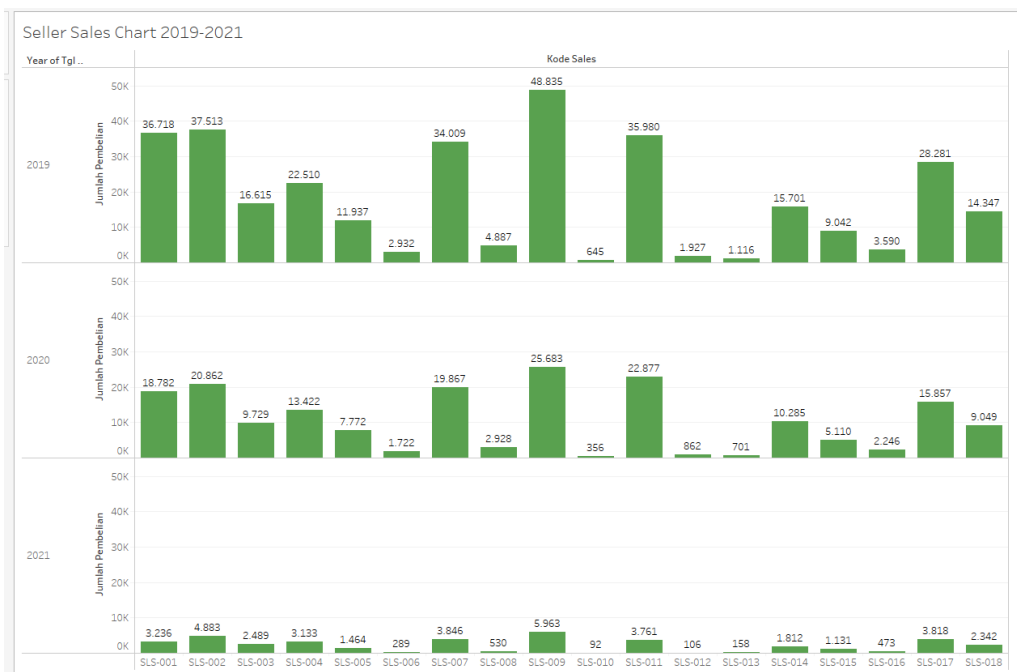

Figure 11. Seller Sales Chart Display Year 2019 to 2021

Figure 11 shows a graph of total sales persales in 2019 to 2021, it can be seen from the graph that the sales figures for each sales per year have decreased. The lowest decline in sales sales will occur in 2021.

4. **Comparison Chart of Sales of Each Product in 2019 to 2021**
   Sales data information that is visualized in the form of a bar chart can be seen in Figure 12 which shows a declining trend in sales figures for each product in the CV. Sumber Tirta Anugerah for the last 3 years.
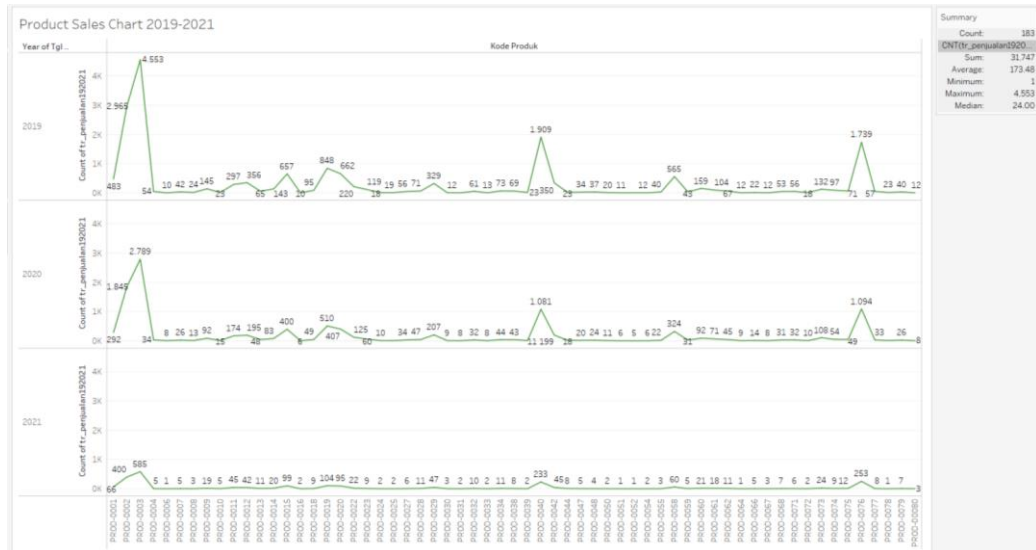

Figure 12. Graphic Display of Product Sales Year 2019 to 2021

Figure 12 shows a graph of sales per product in 2019 to 2021, from the graph it can be seen that the total number of sales for each product has decreased every year. The highest per product sales figure in 2019 and the lowest per product sales figures in 2021.

5. **Information in the form of a Dashboard**
   Figure 13 below shows an informative page or dashboard that contains a combination of several graphs, namely a total sales graph, a branch sales graph, a sales sales graph, and a per product sales graph for the last 3 years in 2019 to 2021.
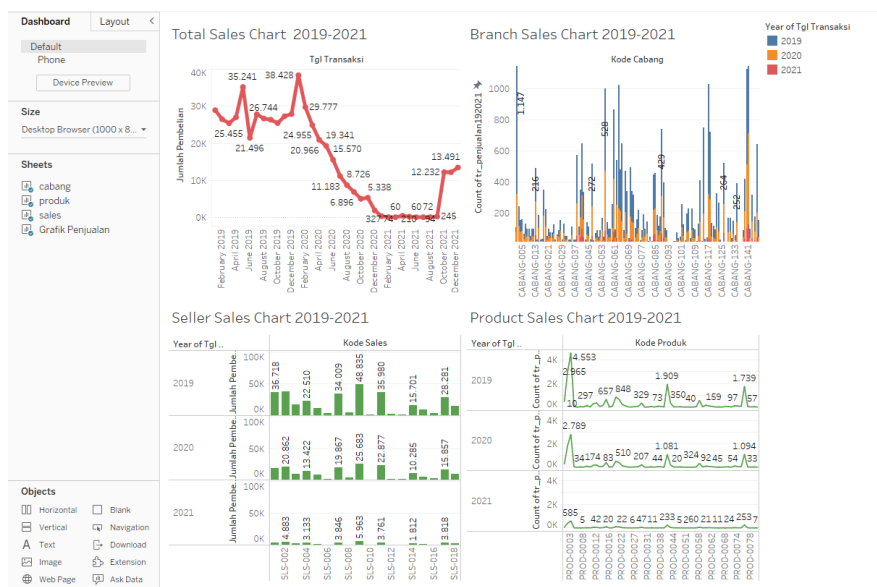

Figure 13. Information Display in Dashboard Form

CONCLUSION

This sales transaction data warehouse has been successfully designed using the nine step design method and using a star scheme which results in a transformation response time of 4048 MS. The design of this data warehouse can be visualized using Tableau software with OLAP technology so that the design of this data warehouse can be used to analyze sales data which can be seen from the fact table which contains various dimensions (time, product, branch, and seller). The results of the data warehouse design are useful for analyzing sales data, seeing sales trends, so that they can help companies find data and analyze sales data to determine sales strategies for the following year.

REFERENCES

[1]   I. P. A. E. Pratama and I. G. A. Pradipta, "Data Warehouse Design and Implementation for Sales Prediction at Mekarsari Stores," *J. Teknol. Inform. dan Terap.*, vol. 2, no. 1, pp. 2–5, 2018.

[2]   R. Ferianto, A. Nugroho, and T. Andriyanto, "Data Warehouse Sales Data Management Case Study UD HF Bersaudara," *Semin. Nas. Inov. Teknol.*, vol. 5, no. 1, pp. 148–153, 2021.

[3]   D. Subuh and W. Yasman, "Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho," *Semin. Nas. Inov. dan Apl. Teknol. di Ind.*, pp. 29–36, 2019.

[4]   I. P. S. Handika and P. P. Santika, "Implementation of Data Warehouse and Business Intelligence for Sales Data at PT. ABC," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 5, no. 2, pp. 76–85, 2020.

[5]   I. Putu, W. Prasetia, I. Nyoman, and H. Kurniawan, "ETL (Extract, Transform, Load) Implementation in Sales Data Warehouse Using Pentaho Tools," *TIERS Inf. Technol. J.*, vol. 2, no. 1, pp. 39–47, 2021.

[6]   Maniah and Supono, "Data Warehouse Model Design (Case Study: Post Shop PT. Pos Indonesia)," *Competitive*, vol. 10, no. 1, pp. 9–16, 2015.

[7]   I. P. A. E. Pratama and N. P. N. D. Widhiasih, "Data Warehouse Design For Sales Prediction At Orba Express Using Pentaho," *JUSS (Jurnal Sains dan Sist. Informasi)*, vol. 3, no. 2, pp. 43–48, 2020, doi: 10.22437/juss.v3i2.8147.

[8]   I. R. Marbun and R. Somya, "Data Warehouse Design for Sales Transaction Data Using Snowflake Schema Case Study : Online Market Dataset," *Univ. Kristen Satya Wacana*, vol. 5, no. 1, pp. 87–91, 2021.

[9]   M. Hasanudin *et al.*, "Material Sales Data Warehouse Model Design using Star Schematic," *Semin. Nas. Inov. Teknol.*, vol. 2, no. 2019, pp. 34–41, 2019.

[10]   R. Riksazany and M. Ayub, "Exploration of Sales Data Warehouse with Tableau," *J. Strateg.*, vol. 1, no. 2, p. 574, 2019.

[11]   E. Triandini, M. S. Rijal, and M. P. Ambara, "Implementation of Star Schema in Development of Data Warehouse Sales of Tour Products," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 12, no. 1, p. 23, 2021, doi: 10.22303/csrid.12.1.2020.23-33.

[12]   M. Akbar and Y. Rahmanto, "Sales Data Warehouse Design Using Nine Step Methodology for Business Intelligence at PT Bangun Mitra Makmur," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, pp. 137–146, 2020, doi: 10.33365/jatika.v1i2.331.

[13]   D. Sugiarto, H. L. H. Warnars, and Winarno, "Sales Data Warehouse Design (Case Study of PT. Subafood Pangan Jaya)," *Semin. Nas. Ris. dan Teknol. (SEMNAS RISTEK)*, pp. 271–276, 2020.

[14]   R. Kimball and M. Ross, *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. 2010.