Designing Data Warehouse To Data Visualization For Identify Document Patterns At Customs And Excise Office

Nadya Hastono Putri, Syaifudin, Teddy Siswanto

Abstract: The increase in import volume at the Pasar Baru Post Customs Office caused relatively long service processes. This is due to several processes that must be passed, including physical checking of goods, determining criteria for value of goods, lines, and status by the office. To streamline services, information is needed regarding the process that must be passed.In this study, the design and implementation of a data warehouse were conducted to visualize information in the form of the most frequently used path types and statuses. Users can easily understand the service process that causes buildup. The Case-Based Reasoning (CBR) method is used to determine the similarity of the new case to the old case based on weight

Index Terms: Services, Data Warehouse, Visualization, Information, Case-Based Reasoning (CBR)

———————————————— ◆ ————————————————

1. INTRODUCTION

Indonesia is one of the developing countries, mainly engaged in business, namely by selling online. So that it has a direct impact on the increasing volume of delivery of postal items to and from abroad. A postal operator is a business entity that organizes posts. Shipment is items sent through the postal operator in accordance with statutory regulations in the postal field. The decision of the parties customs and excise to provide effective services and supervision users of international postal mail services, that is by working quickly, efficiently, transparently, and responsive to the needs of users of international postal mail services [1]. In the case of consignments, there is a fee that must be borne by the seller to process the payment or settlement. Currently, shipping is a very important requirement in helping users to import shipments. The increase in import volume at the Pasar Baru Post Customs Office caused relatively long service processes. This is caused by a number of processes that must be passed, including physical checking of goods, determining criteria for the value of goods, routes, and status by the office. The needs required for the decision-making process, users can quickly obtain and facilitate the search for information and can streamline services based on the service time data. In previous research, according to Ngurah and Nata the data warehouse was built with a multidimensional model in the form of a star schema which is equipped with a loading process using the Extract, Transform, and Load (ETL) method [2]. In designing a data warehouse using the Kimball method, then analyzed using the OLAP (On-Line Analytical Processing) process [3]. In another study, according to Darman, Microsoft Power BI displays information on rice plant data using visualization in the form of a digital mapping feature, namely map visualization so that the distribution of rice production is uneven because remote areas are not visualized [4]. Data

————————————————

- *Nadya Hastono Putri is college student in Information System Study Program, Universitas Trisakti,, Indonesia.*
  E-mail: nadya06500160000@std.trisakti.ac.id
- **Syaifudin js currently lecturer in Information System Study Program, Universitas Trisakti, Indonesia*
  Corresponding author E-mail: fudin@trisakti.ac.id
- *Teddy Siswanto js currently lecturer in Information System Study Program, Universitas Trisakti,, Indonesia*
  E-mail: teddysiswanto@trisakti.ac.id

visualization in the Ministry of Home Affairs system uses maps, tables and graphics, but the visualization used is still incomplete [5]. In this study, a data warehouse was designed using service time data that contained data on the value of goods, status, paths and decisions because it made it easier for users to search for information and process data using appropriate tools and continued with data visualization. Then a decision support analysis is carried out to help Customs and Excise to identify patterns in service time data to facilitate the process of further decision making.

2 LITERATURE REVIEW

2.1 Data Warehouse
Data warehouse is a system with an open architecture and is used to build a particular architecture [6]. Data warehouse to collect some raw data that comes from operational systems and other external data. Then to carry out operational data cleaning, integrating, and storing data then it will distribute the results to user [7].

2.2 Extract, Transform, and Load (ETL)
Process Extract Transform and Load (ETL) is a process of part of the business intelligence system and the data comes from different sources of daily transactions.
And data will be extracted and entered into the data warehouse [8].

2.3 Pentaho Data Integration
Pentaho Data Integration (PDI) is an application that functions for the Extract, Transformation and Load (ETL) process in Business Intelligence (BI). Data or information from different sources and formats can be integrated into integrated data for user [9].

2.4 Data Visualization
Data visualization is an effective way of presenting data that is converted into information. Data visualization can change information quickly so that it can be universally understood [10]. The purpose of data visualization is to increase understanding and be able to assist in the process of collecting data and be able to conduct an analysis of several forms of data available. And data visualization has the benefit to increase the value of the results of the data that has been

analyzed and provide deeper knowledge to the reader [11].

2.5 Case-Based Reasoning (CBR)
Case-Based Reasoning (CBR) is a new paradigm technique in artificial intelligence, then suggested as a feasible method for solving problems by referring to knowledge and information and experience from the same event as before. CBR techniques have strong potential as a decision support tool to improve the efficiency and quality of safety management in identification and prevention [12]. In the structure of the CBR system in Figure 1 as a black box that includes the reasoning mechanism and external aspects so that the input or case of a specific problem, the expected solution as an output and previous cases are stored as a reference [13].
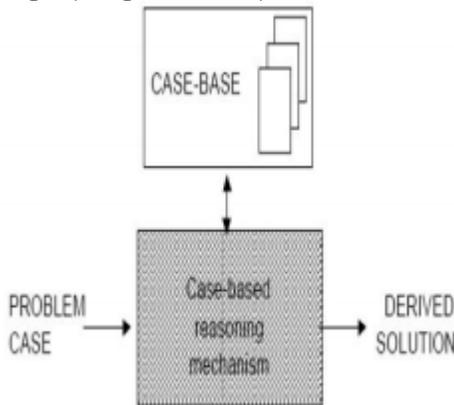
Similary (problem, case):

$$\frac{S1*W1+S2*W2+....+Sn*Wn}{W1+W2+.......+Wn}$$

where:
S = Similary (Similary value) 1 (same) and 0 (different)
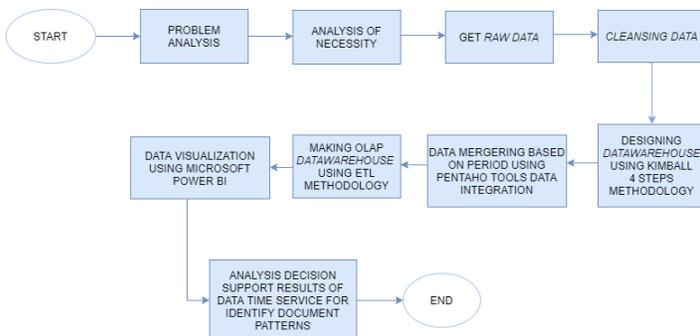W = Weight (Weight obtained)

Fig 1: *Architecture of CBR System*

3 METHODOLOGY
This research uses the methodology in Figure 2 below:

Fig 2: *Research Flow*

The method used in this study began by collecting raw data in the form of Microsoft Excel spreadsheet files. Cleansing data from raw data using Pentaho Data Integration [14]. Used to sort out data that is needed or not in order to cleaning data to be more structured. Data is organized and combined by period

so that it is obtained in the form of Items Receipt Data. Designing data warehouse, table structure design and attributes using Kimball method [15]. The table is made in form of a fact table and several dimension tables. This study forms the dimensional model data using star schema. Fact table in the middle and surrounded by other dimension tables. The tables using include facts of service time, item dimensions, line dimensions, time dimensions, value dimensions of items, decision dimensions and status type dimensions. To make OLAP data warehouse, previously processed data is entered into the database structure using the ETL method [16]. Making OLAP in this study using Pentaho Data Integration. The data visualization process is then carried out aiming to see the results of the data in graphical form, so that it is more easily understood and read using Microsoft Power BI Desktop tools [17]. Data used for visualization are FOB, CIF, Netto, Path, Status, and Decision Data derived from service time data that has been processed in the previous step. The final step is to analyze the results of decision support service time to determine patterns based on the size of the data derived from decision data and status. So, it can be compared from the results of the analysis of previous cases with new cases using Case-Based Reasoning (CBR) method [18]. To search for decisions taken based on the determination of the type of current status that will be chosen by the user. Can provide the results of decisions to support the best process for the needs of the office in making further decisions [19].

4 RESULT AND DISCUSSION
To create designing data warehouse using Kimball method. The study began with data cleansing using Spoon Pentaho Integration Data. Then the data is combined by period. If you have Determine the grain, then the fact table and dimension table and it can determin too in Figure 3 below:
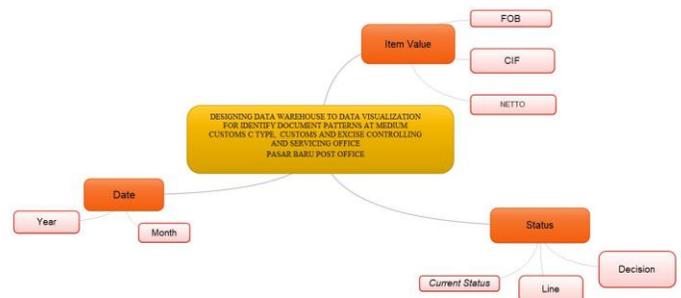
Fig 3: *Mind Map Grain*

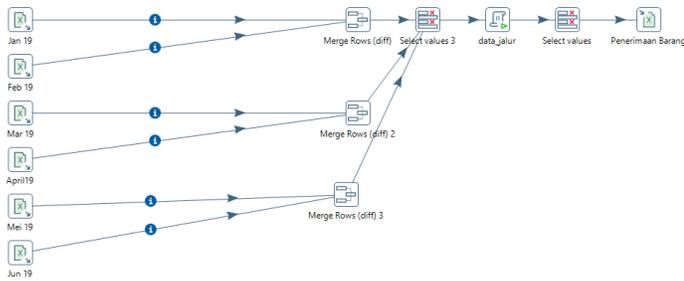Figuer 4 is the first step cleansing data of service time data in January, 2019.

Fig 4: Transformation of cleansing data

To make data mergering of all the data in 2019 as Figure 5.

Fig 5: Transformation of data mergering

After the previous transformation has been successfully, the Items Acceptance get spreadsheet file in Figure 6 below.

Fig 6: Spreadsheet of mergering data of Items Acceptance

To create transformation dimensions table, according to the of the items specified as Figure 7.

Fig 7: Transformation of dimension table

To make a fact table transformation, the data is used from the dimension table to create fact_service_time as Figure 8.
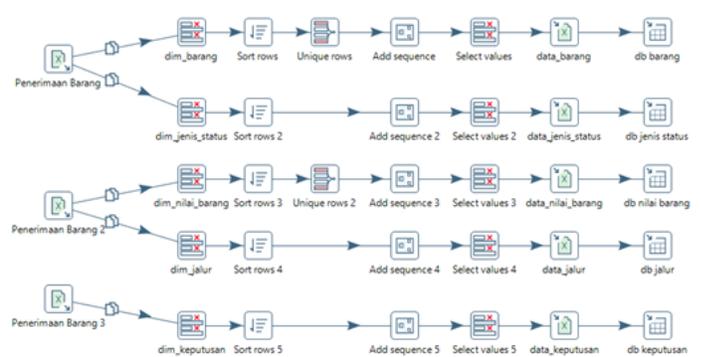

Fig 8: Transformation of fact table

Making OLAP databases contain fact tables and dimension tables that are interconnected. So, to form a star schema and the results of name database is database_itemdata in Figure 9 below.
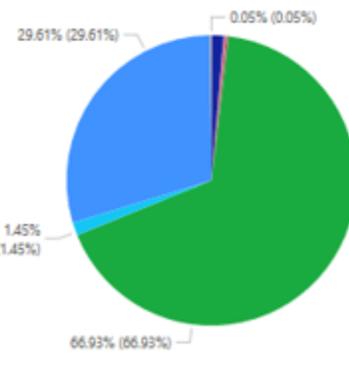

Fig 9: Star Schema database_itemdata

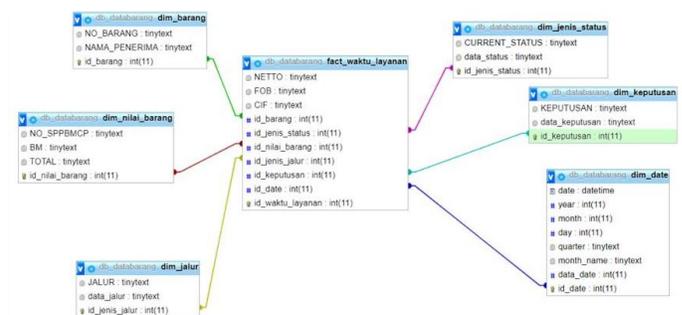The database created in PhpMyAdmin in Figure 10 bellow.

Fig 10: Database om PhpMyAdmin

Power BI Desktop is used to display data visualization, namely (1) the relationship between FOB, CIF and Net with month_name, (2) shows the type of path most often used for physical inspection of items and the type of decision taken from the office, and (3) the relationship between type of status and decision. Visualization using Pie Chart in the form % of relationship between FOB data and CIF data with month_name in Figure 11 below.
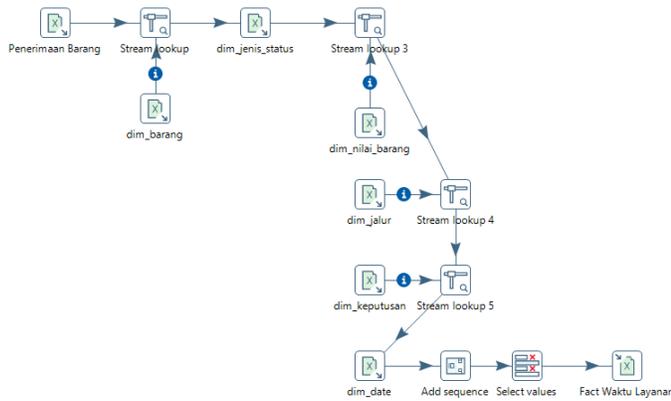
75

Fig 11: *Visualization Dashboard of FOB and CIF Data*

Visualization data using the Pie Chart of Netto and Line is used for physical of items and data visualization of decisions by customs and excise in Figure 12 below.
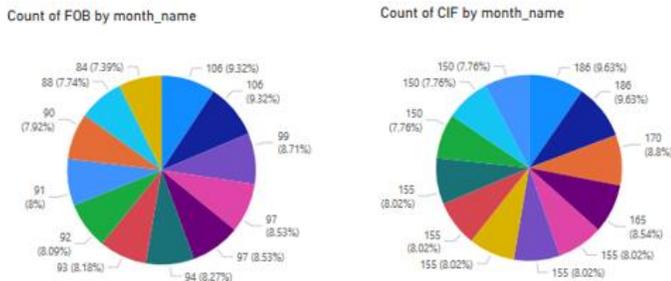

Fig 12: *Visualization Dashboard*
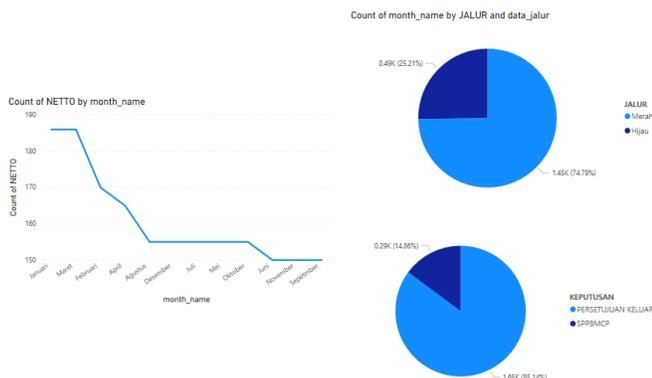
Visualization of current status as Figure 13.

Fig.13: *Visualization Dashboard Current Status*

Case Based Reasoning (CBR) method is required a weight value for the analysis of service time data. The flow used in the process of shipment items business operators send goods through PT. Pos Indonesian. Customs can check the items through the line first. The red lane is the process of carrying out a physical inspection of imported goods. While, the green line is the goods can be submitted to the importer without having to do a physical inspection of goods or an x-ray process. To determine the weight according to individual needs, the weighted value starting from number 1 is the lowest value and number 5 is the highest value. The greater the weight value, the greater the effect of the symptoms on the decision.

The following is the status symptom data for the shipment.

TABLE 1
Symptom Data

Code	A01	A02	Value Weight
G-1	✓	✓	5
G-2	✓	✓	2
G-3	✓	✓	5
G-4		✓	2
G-5		✓	2
G-6	✓		3
G-7	✓	✓	1
G-8	✓		2
G-9	✓	✓	5
G-10	✓	✓	4
G-11	✓		1
G-12	✓	✓	5
G-13	✓		2
G-14	✓		2
G-15	✓	✓	5
G-16	✓	✓	5
G-17	✓	✓	3

The following description for the decisions in Table 2

TABLE 2
Decision Data

Code	Decision
A01	EXIT APPROVAL
A02	LETTER OF DETERMINATION OF PAYMENT OF IMPORT, EXCISE AND OR TAX REVENUE (SPPBMCP)

Following description for the type of current status in Table 3:

TABLE 3
Symptom Data

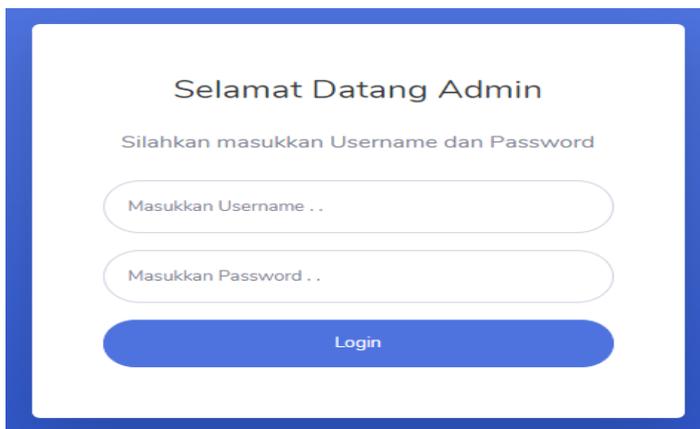| Code	Symptom
G-1	GOODS WILL BE PHYSICAL EXAMINED, WAITING FOR PREPARATION OF GOODS BY P OS
G-2	ITEMS RETURNED TO EXCISE DUTY, SPPBMCP (BILLING) CANCEL
G-3	GOODS AFFECTED WITH LIAR / LIMIT RULES
G-4	PROCESSES P2 / SBP
G-5	DELIVERY OF DELIVERY A
G-6	REPORT OF EXAMINATION HAS BEEN RECORDED

76

The following table results from the relation process in Table 4:

TABLE 4
Reuse Process

| G-7	CANCELLATION OF DOCUME N
G-8	PHYSICAL EXAMINATION ACCORDING TO UNIT
G-9	DETERMINATION OF WAITING SCANNED GOODS (XR AY) AND MANIFES
G-10	DETERMINATION OF N SPPBMCP (PAYMENT AND APPROVALS
G-11	DETERMINATION OF SPPBMCP (PAYER N AND APPROVALS) WAITING FOR SCAN (X-RAY) OR MANIFES
G-12	ITEMS WILL BE PHYSICAL EXAMINED, WAITING FOR THE PREPARATION OF ITEMS BY POST ORGANIZERS
G-13	ITEMS RETURNED TO EXCISE DUTY, SPPBMCP (BILLING) CANCEL
G-14	EXPRESS AGREEMENT WITH EXEMPTION FROM IMPORT DUTY AND TAXES
G-15	SPPBMCP HAS NOT BEEN PAID THROUGH TIME
G-16	SPPBMCP LUNAS
G-17	SPPBMCP HAS BEEN THROUGH A TIME AND IS PAID

Decision Support System Implementation
Admin Page:
The login display functions to display the login page to enter the admin page by filling in the username and password fields in Figure 14 below.

Fig 14: Display Login

The admin page display functions to display the admin main page which consists of a menu of symptoms, decisions, case basis, and logout in Figure 15 below.
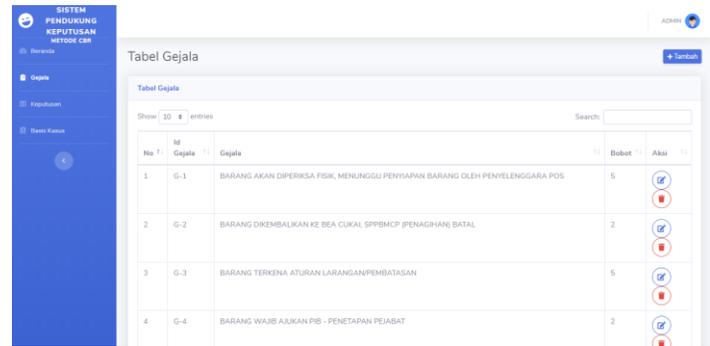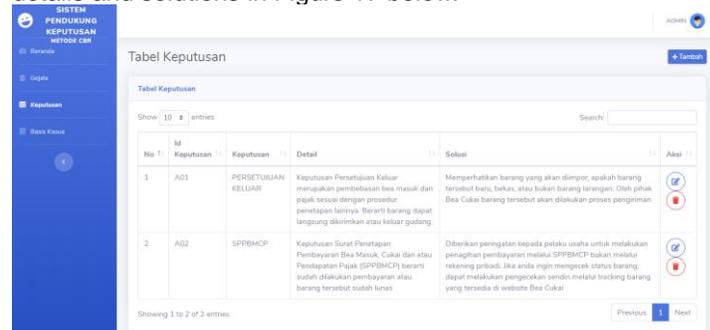

Fig 15: Admin Page

The symptom menu display functions to add, change and delete status symptom data on shipments in Figure 16 below.
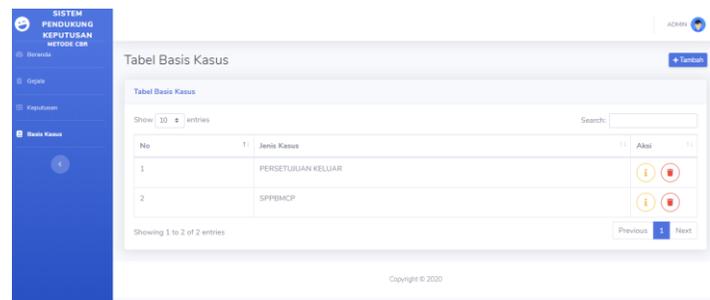

Fig 16: Symptoms Menu Display

Decision menu display functions to add, change and delete decision data. Which contains the name of the decision, details and solutions in Figure 17 below.
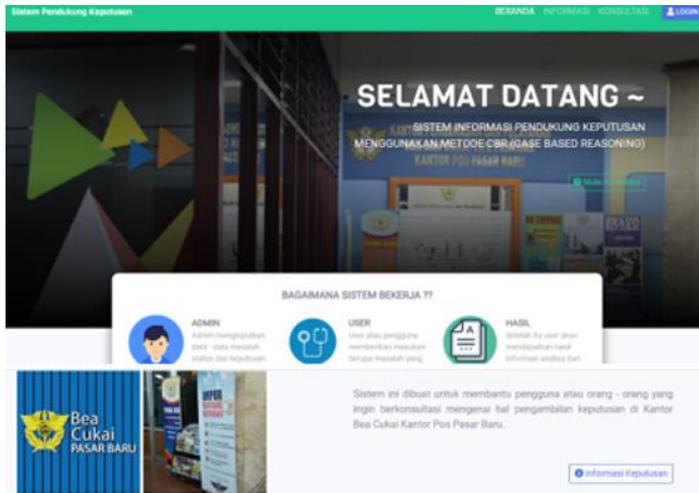

Fig 17: Decision Menu Display

This view serves to manage the case base of symptom data and decision data in Figure 18 below.

Fig 18: Display Case Base

User Page:
The main page display consists of a home menu containing profiles, decision information, and consultations in Figure 19 below.

Fig 19: Main Page Display

This view is a display of information that contains decision data, which can be seen by the user in Figure 20 below.


Fig 20: Decision Information Display

The consultation page display functions when the user wants to do a consultation by filling out a checklist first. Adjusted for symptoms of status problems on the item in Figure 21 below.

Fig 21: Consultation Page Views

This display is a display of consultation results in the form of decision making in accordance with the symptoms of a status problem and there is a solution.
By using the CBR method, it can display the results of the

analysis in the form of the calculation of Similarity Value which is used to support decisions in further processing in Figure 22 below.
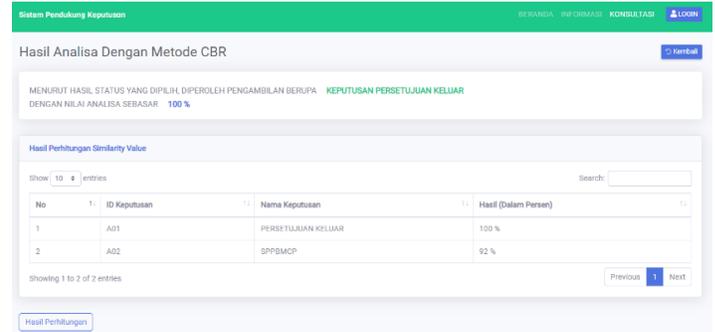
Fig 22: Display of Consultation Results

4   CONCLUSION

This research applied OLAP data warehouse with the appropriate type of dimensional modeling using Star Schema for data warehouse design using the Kimball method and the Extract, Transform, and Load (ETL) method. Dimension tables are: (1) Dimensions of Items; (2) Dimensions of Date; (3) Dimensions of Value of Items; (4) Status Type Dimensions; (5) Line Dimensions; and (6) Decision Dimension. The results of the visualization of the relationship between FOB, CIF and, Netto item value data with month_name to determine the value of goods that meet the standard, increased shipments in January and March. The FOB data was obtained 9.32% (106). CIF and Net data obtained 9.63% (186). While the decline occurred, the FOB data occurred in July obtained 7.39% (84). Then for CIF and Net data there was a decrease in June, November and September obtained 7.76% (150). Based on the results of the analysis of service time data decision support using the Case Based Reasoning (CBR) method, the similarity value between the new cases and the old cases obtained the highest value in the form of 0.44 is the Decree of Stipulation of Payment of Import Duty, Excise and / or Tax Revenue (SPPBMCP). And Decision support system is built to identify problems in the shipment process, which results in decisions taken and their solutions based on the selected status. And provide convenience to the Customs and Excise Supervision and Service Office (KPPBC) Pasar Baru Cukai Post Office in making further decisions.

REFERENCES

[1] J. Akuntansi, M. Bisnis, D. Kartika, T. A. Batam, A. Tipe, and M. Kuning, "Afrianto," Accounting, Econ. and Manaj. Business, vol. 6, no. 1, pp. 107–118, 2018.

[2] G. Ngurah and M. Nata, "Detection of Outlier Transactions Using Olap Visualization in Private Higher Education Data Warehouses," J. Appl. Intell. Syst. UDINUS, vol. 1, no. 2, pp. 77–89, 2016.

[3] A. Armadyah, "Designing and Making Data Warehouses in Yogyakarta STAKIK AKAKOM Library," Tesis. Perpust. MTI, pp. 39–52, 2009.

[4] R. Darman, "Visualization Analysis and Mapping of Rice Crop Data in Indonesia Using Microsoft Power BI," J. Ilm. Engineering and Management," J. Ilm. Rekayasa dan Manaj. Sist. Inf., vol. 4, no. 2, pp. 156–162, 2018.

[5] J. T. Geodesi, "Evaluation of Data Visualization System Development (Case Study: Data Management at the Indonesian Ministry of Home Affairs) MUHAMMAD FIKRY

ABYADL, SUMARNO, INDRIANAWATI," vol. 2016, no. 2, pp. 80–89, 2016.

[6] F. Kalna, A. Erraissi, M. Banane, and A. Belangour, "A scalable business intelligence decision-making system in the era of big data," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 12, pp. 2034–2042, 2019.

[7] Firmansyah "The Most Commonly Used Database Type." [Online]. Available:https://www.nesabamedia.com/pengertian-database-dan-fungsinya [20 Jul. 2020].

[8] Wijaya, R., & Pudjoatmodjo, B. (2016). Application of Extraction-Transformation-Loading (ETL) in Data Warehouses (Case Study: Ministry of Agriculture). National Journal of Informatics Engineering Education (JANAPATI), 5(2), 61.

[9] A. Aryati and D. Anwar, "ournal of Industrial and Logistics Management DESIGN OF BUSINESS INTELIGENCY SYSTEM IN JAMU PRODUCT SUPPLY CHAIN BASED ON THE PENTAHO BUSINESS INTELLIGENCE," J. Manaj.Ind. dan Logistik, vol. 1, no. 2, 2018.

[10] A. Sreekant, P. Senthilnathan, G. Gopichand, M. Rajapandy, and N. Kannan, "Necessity of machine learning and data visualization principles in marketing investment management," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 6 Special Issue 4, pp. 527–530, 2019.

[11] Eaton, M. 2017. Seeing library data: A prototype data visualization   application for librarians. Journal of Web Librarianship   11(1):   69–78. DOI:10.1080/19322909.2016.1239236 [30 Jun. 2020].

[12] Naradhipa, A. R., Syaifudin, & Sari, S. (2020). Knowledge management system for broiler and laying cultivation. International Journal of Scientific and Technology Research, 9(3), 1373–1378.

[13] D. P. Utomo and S. D. Nasution, "Expert System to Detect Damage Toner Using the Case-Based Reasoning Method," J. Ris. Komput., vol. 3, no. 5, pp. 3–6, 2016.

[[14] A. Aryati and D. Anwar, "ournal of Industrial and Logistics Management BUSINESS INTELLECTENS DESIGN SYSTEM IN SUPPLY CHAINS PENTAHO-BASED JAMU PRODUCTS BUSINESS INTELLIGENCE," J. Manaj.Ind. dan Logistik, vol. 1, no. 2, 2018.

[15] I. T. Del, P. R. Togatorop, J. Sisingamangaraja, D. Sitoluama, and K. Laguboti, "Dimension Design Model in Data Warehouse for New Student Admissions in," vol. 01, no. 01, 2018.

[16] S. Mathur, V. Bali, S. L. Gupta, and P. Pahwa, "Data warehousetesting and security: A conspectus," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 9 Special Issue, pp. 843–850, 2019.

[17] R. J. Salaki, J. Waworuntu, and I. R. H. T. Tangkawarow, "Extract transformation loading from OLTP to OLAP data using pentaho data integration," IOP Conf. Ser. Mater. Sci. Eng., vol. 128, no. 1, 2016.

[18] H. M. Jumasa, S. Fauziati, and A. E. Permanasari, "Case-Based Application in Determining Equity Based on Oil Palm Land Conformity," Citee 2017, pp. 306–313, 2017.

[19] T. Kautz and J. J. Heckman, The Effect of Import Duties and Taxes in the Framework of Imports (Bm and Pdri) Against Total Revenue" NBER Work. Pap. Ser., vol. 10, no. 1, pp. 1–8, 2013.