Design and Build Data Warehouse Using Ontology and Rule Base Method in Supporting Sales and Service Information

Muhammad Qusyairi[1], Made Sudarma[2], and Agus Dharma[3]

[1]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[2,3]Department of Electrical and Computer Engineering, Udayana University
*erick.rrtr@gmail.com

Abstract Data warehouse makes company data that is spread out become integrated and concise so that it helps managers or top management in analyzing existing data for strategic and fast decision making. With the application of data warehouses using the rule base ontology method, it makes the class of related data relations easier to analyze so that it helps companies monitor the good results of sales data relating to goods, sales locations, turnover values, and data services related to product service, region , technician, and service value.

***Index of Terms* — DataWarehouse, ETL, Ontology, Rule Based, Olap.**

I. INTRODUCTION

THE development of technology is currently experiencing rapid progress, the need for information in the business world becomes very important in determining the progress of the company. Information is a vital requirement in the design of activities and decision making that affect the development of a company [1]. With accurate, complete and relevant information, the company can compete with its competitors. In an increasingly tight global era with business competition both in the supply of goods and services, the success of a company or organization is highly dependent on information technology that can provide reports for analysis. Basically, a company was founded with the aim of being able to continue to exist in the market and be accepted by the public through the products or services offered. To support the decision making process both in terms of time and the quality of decisions that will be generated, it is necessary to have information technology that can manage data both with large amounts of data that are likely to be integrated to obtain good and accurate information. So to manage the data requires an application of data warehouse.

Data warehouse is a subject-oriented, integrated, non-updateable data collection that has a time dimension, which is used to support the decision-making management process and business intelligence [2]. Data warehouse is able to provide reports and analysis quickly and accurately, because basically The data analyzed is the result of historical data available by the company.

Some research related to data warehouses is research that explains semantic data warehouses with ontology and rule based methods for processing academic data [3]. The semantic data warehouse system that was designed has succeeded in processing college academic data from three OLTP database sources by applying ontology and rule based methods. The data that has been stored in the OLAP semantic data warehouse database is data in a uniform and more meaningful format, so the reports generated from the application of the Kimball nine step method are quality academic data reports to support the decision making process. The semantic search application that was designed was successful in finding the information needed by using natural language keywords in Indonesian. Next research is data warehouse implementation sales information using molap [4]. This research succeeded in developing data

warehouse by using molap to support batik sales, where from the large amount of sales data it was integrated so that the data became centralized for manufacturing final report required by top management is reporting that can provide comprehensive information, so that based on the report the leader can determine the optimal policy in determining the production capacity so as to achieve efficiency in terms of cost and effective in time. The problem occurs because of lack of knowledge about the utilization of such abundant data. Next research is design of virtual data warehouse in supporting the information needs of sales [5]. This research results in a virtual data warehouse design that results in the analysis of sales data in a building material store that has several branches. This study successfully completed a virtual data warehouse that uses the Kimball nine step method and then produces a dimensional star schema model. Next is the research on the application of data warehouse with the nine step Kimball method [6]. This research succeeds in summarizing operational data in a more concise and useful form. Analysis of data warehouse can be done multidimensional. In the student data warehouse, analysis can be done with the parameters majors, semesters, courses, years and grades. Student value data warehouse provides convenience in accessing, searching and comparing student value data as needed. As a result, the data warehouse is expected to provide support to management in the process of evaluation, planning and decision making. Next is data warehouse research that is used to help the decision making process on a television show [7]. This research succeeded in building a data warehouse to support television editorial decisions on Kompas TV, especially in making quick decisions regarding the selection of news to air, including career decision making and providing incentives for employees with good performance. Next research is modeling integration of real time data warehouse [8]. This research succeeded in developing the data warehouse so that it could be real time in data execution, the application of the Nearly Real Time Data Warehouse Modeling was carried out using the Transaction Log-based Change Data Capture method. Change Data Capture is responsible for capturing changes that occur in the Sales database as a result of insert, update, or deletion activities in the retail information system application. This Transaction Log- based CDC method is applied to the Sales database and Sales Stage database. Changes to data captured by the CDC process will then be moved by the Extract, Transform, Load (ETL) process into the NRTDWH database. The process of moving data from the Sales database to the NRTDWH database is done with the help of a scheduler and Integrating the nearly real time data warehouse modeling with service oriented architecture is done by making middleware in the form of web service.

II. DATA WAREHOUSE

Data warehouse is a subject-oriented, integrated, non-updateable collection of data that has a time dimension, which is used to support the decision-making management process and business intelligence. Data warehouse can also be considered as a structured copy of transactional / online transaction processing (OLTP) data for analysis, reporting and data mining needs [9]. Therefore, the transaction system never updates data that is in the data warehouse but only adds data to the data warehouse. Data warehouse is also a collection of data generated to support decision making. Data warehouse is also a place for storing current data and historical data from the interests of managers throughout the organization [10]. Data warehouse can also be called a database that stores important current and historical data from information needs for managers in companies [11].

According to Inmon [12] data warehouse is data from various OLTP applications that needs to be centralized, where the data can be used for analysis. Data must be managed in the form of subject oriented, integrated, non-volatile, and time variant. Data must be accessible to the lowest level if a drill-down is carried out, and it can be summarized if a drill is done. Data marts are treated as part of the Data Warehouse. Each data mart is formed from a data warehouse for a department and is optimized for the department's analysis needs. While Ralph Kimbal is more towards Bottom-Up Approach So the approach is from data mart-datamart where in datamart can use the same dimension which is called conformed dimension. The relationship between datamart through this comformed dimension is called Bus Structure. This data warehouse can be a virtual or aggregation of data mart.

A. ETL

Company data in a heterogeneous system has different formats. However, data with different formats must be collected, integrated and then processed to provide information for the company's business processes. Therefore, we need a system or method that is able to overcome the situation. Therefore, we need a system or method that is able to overcome these conditions, namely ETL.ETL is one of the processes in a data warehouse that involves [13]:

- Reading of data from the source.
- Cleaning and adjusting the data format.
- Writing data into storage space for further use.

Data used for ETL is obtained from various sources, such as mainframe applications, CRM, flat file, Excel spreadsheet format, even message queue.

B. Nine Step Kimbal Method

the nine-step methodology proposed by Kimball. The nine steps include[14]:

1. Choosing the process.

At this stage what is done is that the data mart that was first built must be a data mart that can be sent on time and can answer all important business questions.

2. Choosing the grain
   The selection of data sources to decide for certain what is represented or represented by a fact table. For example, if the data source of a fact table for sale property is an individual sale property, then the source of a customer dimension contains details of customers who bought the main property.
3. Identifying Dimensions (Identifying and conforming the dimensionalos) At this stage of identification the dimensions carried out are:
   a. Set of dimensions that are well built, makes it easy to understand and use the data mart. This dimension is important to describe the facts contained in the fact table.
   b. If there are dimensions that appear on two data marts, both data marts must have the same dimensions, or at least one of them is a mathematical subset of the other.
   c. If a dimension is used on two or more data marts, and these dimensions are not synchronized, then the entire data warehouse will fail, because the two data marts cannot be used together.
4. Choosing the facts

At the stage of sorting the facts carried out are:

   a. The source of a fact table determines which facts can be used in a data mart.
   b. All facts must be expressed at the level determined by the source.
5. Storing pre-calculations in a fact table (Storing pre-calculations in the fact table). The fact table is the main table in the data warehouse, all the information to be achieved through the data warehouse through the fact table. Many calculation processes are carried out on fact tables, and to make it easier to implement the data warehouse it is necessary to store the results of the pre- calculation.
6. Completing the Dimension Table (Rounding out the dimension tables)
   At this stage what we do is:
   a. Add complete information to the dimension table. b.The explanation must be intuitive and easily understood by users.
7. Choosing the duration of the database
   Based on the use of the database created, at this stage it is determined how long the data is stored.
8. Slowly Tracing Dimension Changes (Tracking slowly changing dimensions). There are three types of slowly changing dimensions, namely:
   a. Dimension attributes that have changed are rewritten
   b. The dimension attribute that has changed gives rise to a new dimension.
   c. Dimensional attributes that have changed give rise to alternatives so that the values of the old and new attributes can be accessed together on the same dimension.
9. Determine Query Priorities and Modes (Deciding the query priorities and the query modes)

C. Ontology and Rule Base Methods

The method of ontology is used as a model for the representation of the knowledge base that can describe the relationships or relations between objects in it so that the objects in the knowledge base become more meaningful [ontology]. This ontology method has several stages (rule based) in the design including (1) Determination of the concept of the domain, (2) Determination of the terminology list, (3) Definition of class and hirarki structure, (4) Relationship diagram class, (5) Definition of the property of the class, (6) Constraint property. Ontology rule based is used to manage relations or relationships between ontology constituent elements such as relations between classes, relations of classes with data types, relations between instances in a class or instances between classes, while rule based queries become a reference in the process of searching for semantic-based information. The type of query used to reference this implementation with the preparation of production rules

III. DESIGN SYSTEM DATA WAREHOUSE

A.General Description of the System

The process in this data warehouse consists of 2 general descriptions for processing where there is service data and sales data to be processed. There are several stages that are passed starting from the data source, ETL, data warehouse, and web-based processing system to do the analysis

1. Service Data

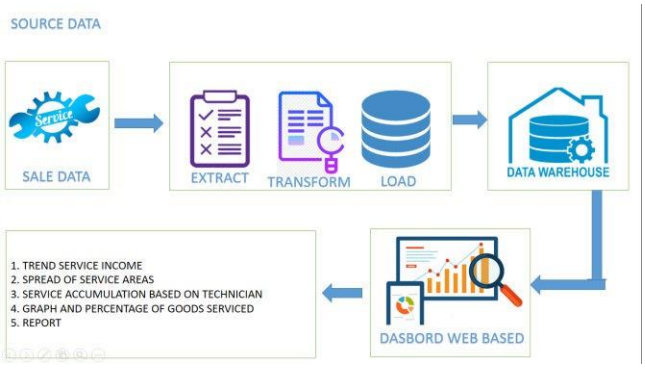

Fig 1. Processing Service Data

The data service processing is passed through the source retrieval step and then the ETL process is performed to clearing data from redundancies and errors, then it will be entered into the data warehouse so that it becomes an integrated data service and is made a web-based dashboard for analysis.

2. Sales Data

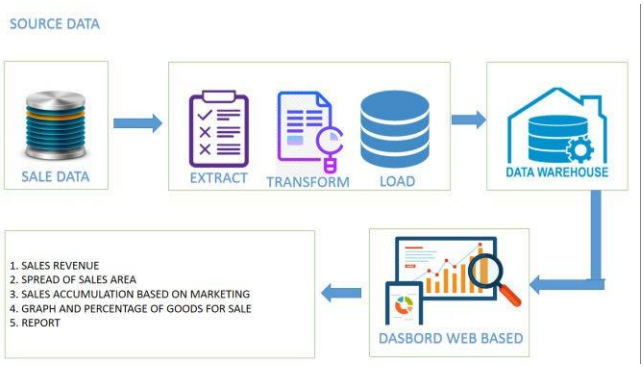

Fig 2. Processing Sales Data

The processing of sales data is passed through the source collection stage and the ETL process is carried out for clearing data from redundancies and errors, then it will be entered into the data warehouse so that it becomes an integrated sales data and made a web-based dashboard for analysis

B. Schema Ontology

In the design of this data warehouse system using the ontology method with rules will produce a mapping and will be transformed into a multidimensional design star schema which will later become a conceptual reference database storage.

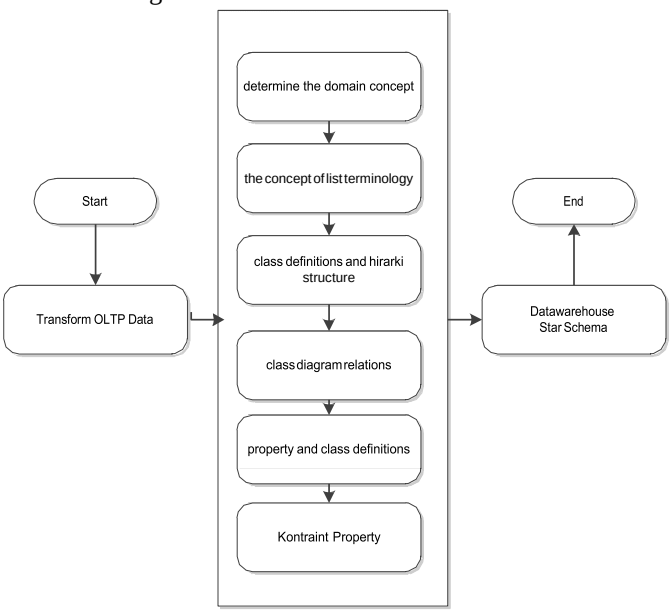

Fig 3. Flowchart Processing ontology

C. Design Multidimensional

In the development of this data warehouse using star schema where there is a facta table surrounded by dimension tables. where the facta table contains numbers and the key history data that is generated is very unique and will be detailed by the dimension table as an explanation of its clarity.

Fig 4. Schema Dimensional Service Data

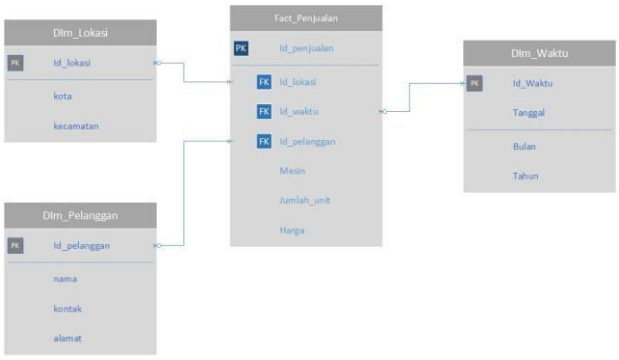

Fig 5. Schema Dimensional Sales Data

D. Processing ETL (Extract, Transform, Load)

In this ETL process, it uses pentaho data integration for processing from the source into a prepared database, in this process data from the source will be cleared and data service integrated with sales data so that it will become a data warehouse that will be analyzed later.

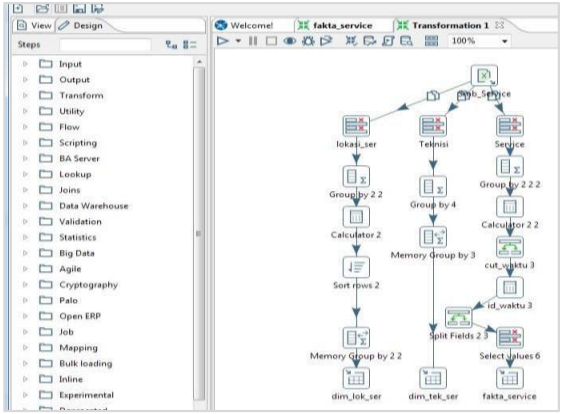

Fig 6. Processing Integration Sales Data

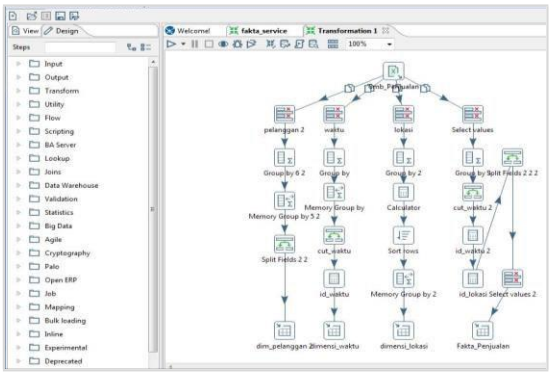

Fig 7. Processing Integration Service Data

IV. RESULT AND DISCUSSION

Data warehouse that has been built using the rule base ontology method is then developed based on web for analysis, while the results of this data warehouse are as follows

A. Interface Data warehouse

Fig 8. Interface Dashboard Data

This page provides an overview of data services and sales, both the amount of income from sales and service

B. Interface Data Service

1. Analysis Service Data

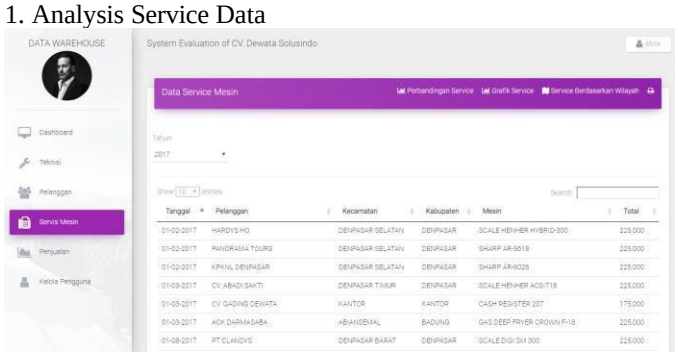

Fig 9. Analysis Service Data

This page displays analysis of service data both from the level of income, technicians and customers involved as well as from the spread of service areas

2. Analysis Chart Service Data

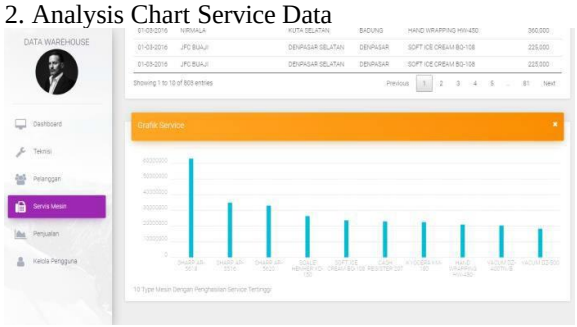

Fig 10. Analysis chart Service Data

This page displays a graphical analysis of the level of service revenue obtained from the product being serviced as well as the value of each period.

3. Analysis Ratio Service Data

Fig 11. Analysis Ratio Service Data

This page shows the comparative analysis of service revenue for each period and the percentage of increase and decrease in the results obtained by the company is followed

4. Analysis Region Service Data

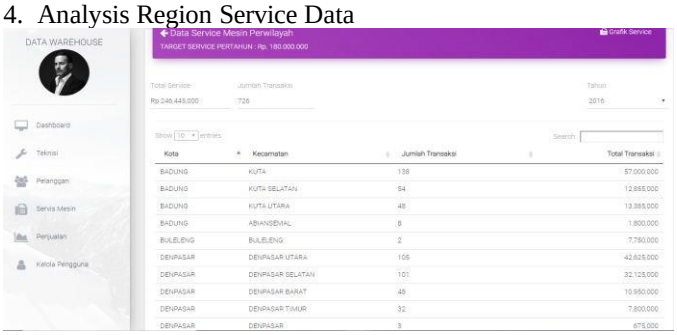

Fig 12. Analysis Region Service Data

This page shows the analysis of any region that provides good results for service to customers so that it will make a description of the area after ranking with the best service

.

5. Analysis Chart Region Service Data

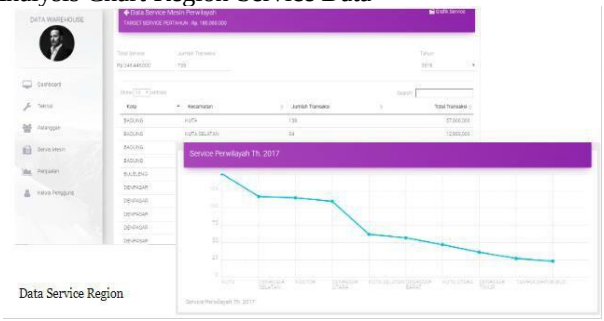

Fig 13. Analysis Chart Region Service Data

This page displays a regional analysis with charts and data services according to ranking from data with period choice.

C. Interface Sales Data

1. Analysis Sales Data

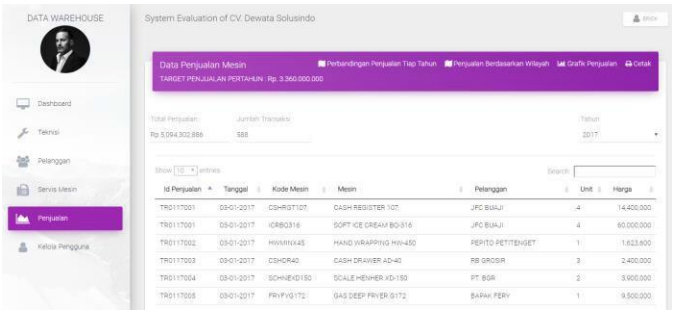

Fig 14. Analysis Sales Data

This page displays analysis of sale data both from the level of income, marketing and customers involved as well as from the spread of sales areas

2. Analysis Chart Sales Data

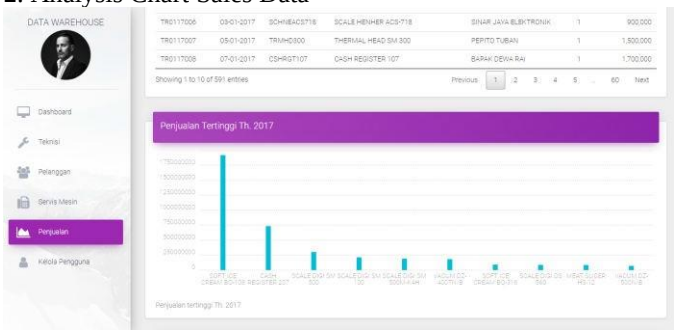

Fig 15. Analysis Chart Sales Data

This page displays a graphical analysis of the level of sales revenue obtained from the product being sale as well as the value of each period.

3. Analysis Ratio Sales Data

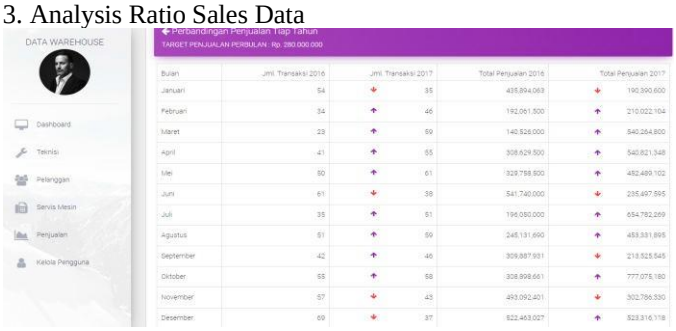

Fig 16. Analysis Ratio Sales Data

This page displays the sales analysis of the existing period and the percentage value comparison so that it can monitor the company's performance going forward

4. Report

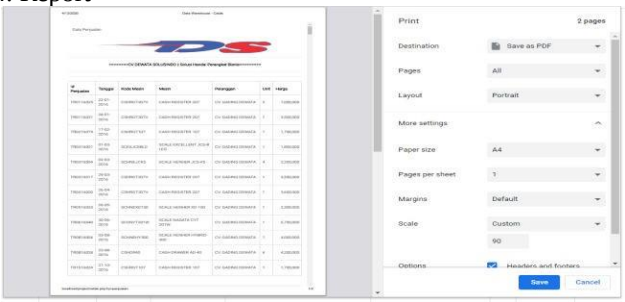

Fig 17. Report Datawarehouse

In addition to displaying an analysis of the data warehouse system, it is also able to print hardcopy reports for the purpose of analysis for managers.

V. Conclusion

From the explanation above, it can be concluded:

1. Data warehouse design and development begins with an analysis process, which defines the company's business information needs, determines data sources and data warehouse schemes created using Pentaho Data Integration (PDI) and then is developed web-based to provide reporting and results that are easy to understand for managers.
2. Data warehouse is able to minimize the cost, time and effort spent to get fast, accurate results to users.
3. Data warehouse that builds is able to analyze data services and sales so that it can see pertanse and graph of the results obtained so as to be able to monitor the company's performance so far.
4. Web-based data warehouse dashboard application makes analyzing data complete with graphs and reports making it easy to make decision-making processes.

References

[1] Vardiansyah, Dani. Filsafat Ilmu Komunikasi: Suatu Pengantar, Indeks, Jakarta 2008.

[2] Inmon, W. H. Building Data Warehouse, 3th Edition. Canada: John Wiley & Sons, Inc.2005.

[3] Ambara, Pradnyana, Made Sudarma. Dan I Nyoman Satya Kumara. "Desain Sistem Semantic Data Warehouse dengan Metode Ontology dan Rule Based untuk Mengolah Data Akademik Universitas XYZ di Bali", Teknologi Elektro, Vol. 15, No. 1, Januari-Juni 2016.

[4] Basakara, I Made Adi. Luh Gede Putri Suardani. Dan Made Sudarma. "Data Warehouse ImplemantationTo Support Batik Sales Information Using MOLAP". International Journal of Engineering and Emerging Technology, Vol. 3, No. 1, January—June 2018.

[5] Sumichan, Andrew. I Made Gede Yudiana. Muhammad Ridwan Satrio. Made Sudarma. "Designing a Virtual Data Warehouse in Supporting Sales Information Needs". International Journal of Engineering and Emerging Technology, Vol. 4, No. 1, January—June 2019.

[6] Ganda, Wijaya. "Perancangan Data Warehouse Nilai Mahasiswa Dengan Kimball Nine-Step Methodology". 2017.

[7] Eugenius, "Analisis Dan Perancangan Data Warehouse Untuk Mendukung Keputusan Redaksi Televisi Menggunakan Metode Nine-Step Kimball (Studi Kasus Pada Redaksi Kompas Tv Jakarta)", 2018.

[8] Sulastra, I Made Dwi Jendra. Made Sudarma. I Nyoman Satya Kumara. "Pemodelan Integrasi Nearly Real Time Data Warehouse dengan Service Oriented Architecture untuk Menunjang Sistem Informasi Retail". Teknologi Elektro, Vol.14, No. 2, Juli-Desember 2015.

[9] Kadir, Abdul. "Data Warehouse, Data Mart, OLAP, dan Data Mining". 2008.

[10] Turban, S, Delen, Business Intelligence Dashboard. Yogyakarta. Indonesia, 2011.

[11] Laudon, K.C., Laudon, J.P. Management Information Systems (11th Edition). New Jersey : Pearson Prentice Hall, 2010.

[12] Inmon, W.H, Building the Data Warehouse, Third Edition. New York: John Wiley & Sons, Inc, 2002.

[13] Songini, M.L. "ETL", Computerworld, [Online], vol. 38, no. 5, pp. 23. Reach1to1 Technologies, (2015).

[14] Kimball, R., & Caserta, J. (2004). The Data warehouse ETL Toolkit. Indianapolis: Wiley Publishing, In