Development of Star-Schema Model for Lecturer Performance in Research Activities

M. Miftakul Amin[1], Adi Sutrisman[2], Yevi Dwitayanti[3]

Department of Computer Engineering, Politeknik Negeri Sriwijaya, Palembang, Indonesia[1, 2]
Department of Accounting, Politeknik Negeri Sriwijaya, Palembang, Indonesia[3]

Abstract—**In this study, the researchers developed a multidimensional data model to investigate the activities of lecturers in universities in carrying out research activities as part of the Three Pillars of Higher Education. Information about lecturers' research activities has been managed using spreadsheet (excel) documents. Thus, access and analysis of the information were limited. Data warehouse development was carried out through several stages, namely requirement analysis, data source analysis, multidimensional modeling, ETL process, and reporting. The information generated in this data warehouse (DW) can be used as one of the business intelligence (BI) models in universities. In this study, the star-schema model was used in designing dimension tables and fact tables to facilitate and speed up the query process. The information generated in this study can be used by management in universities to make decisions and strategic planning. The results of this study can also be used as one of the important information in the preparation of institutional accreditation data and study program accreditation.**

Keywords—*Data warehouse (DW); star-schema; multidimensional data; business intelligence (BI)*

I. INTRODUCTION

Lecturers as one of the intellectual assets in higher education have one of the activities that are part of the Three Pillars of Higher Education, namely carrying out research and community service. Every year there is a lot of funding to carry out research and community service activities. This funding source comes from the ministries of education, culture, research, and technology, as well as from local governments, and internal universities.

This research and community service information is used as one of the assessment criteria in the preparation of study program accreditation instruments and university accreditation. Currently, at Politeknik Negeri Sriwijaya, research data and community service are documented at the Research and Community Service Center (Pusat Penelitian dan Pengabdian kepada Masyarakat abbreviated as P3M) Politeknik Negeri Sriwijaya, in the form of spreadsheet (excel) file data, without being further managed into useful information for upper management in the decision-making process.

Referring to the research and community service handbook published by the Ministry of Education, Culture, Research, and Technology in 2020, there are 13 types of research funding schemes, and 10 community service funding schemes [8]. With this large funding opportunity, it is appropriate for universities to carry out adequate data management in managing research information and community service.

Currently, universities have quite large data and are spread across several sub-units within it. These data will continue to grow over time. Upper management needs tools to generate information and to assist the decision-making process [1]. Data Warehouse (DW) technology can be used to extract important information from data scattered across information system management units into centralized integrated storage, to provide management information needs, view data from various perspectives, detailed information, and historical data.

A Data Warehouse is an integrated repository of information, making it possible to query and analyze the data. The basic idea is to carry out the process of extracting, filtering, and integrating relevant data. The development of data warehouses in universities is still rarely carried out, even though universities are very rich in the information contained in them. One of the factors is that business transactions in universities are non-commercial [2]. This is also reinforced by Yu [3] which states that the implementation of a data warehouse is underestimated in a university environment. Because many people think that universities are non-profits organizations. Whereas with the increasing number of study programs, lecturers, employees, and students in the future, a university must consider the integration of a data warehouse-based decision support system to make better decisions.

In this study, the researchers developed a multidimensional data model to investigate various perspectives and points of view related to research activities carried out by lecturers in universities. This research is important to do in order to provide relevant information for top management at the Politeknik Negeri Sriwijaya in decision making and strategic management.

This research is organized as follows. Section 2 describes some of the theoretical concepts that underlie this research, such as the concept of a data warehouse and multidimensional data modeling. Section 3 describes the methodology and stages in model development. Section 4 describes the experimental setup, research results, and discussion. Section 5 contains conclusions.

II. LITERATURE REVIEW

A. Implementation of Data Warehouse in Higher Education

Based on research conducted by Bassil [4] it is mentioned that data warehouse development in universities can be implemented by transforming operational databases into data warehouses that can be used in the decision-making process and perform data analysis, prediction, and forecasting. The

development of this data warehouse can be done through several stages, namely data extraction, data cleansing, data transforming, and data indexing and loading. An operational database is a regular database that is intended to run a business on a database and support daily transactions [5].

Bogdanova [6] developed a model known as CaMeLOT as an Educational Framework for Conceptual Data Modelling. By using Bloom's taxonomy modeling is carried out as a part of software engineering. This study proposes that the model can be used in the preparation of a curriculum that can adapt the use of technology in the implementation of learning in universities.

A previous study conducted by Santoso [1] mentions that the development of data warehouses in universities can be categorized into 2 groups, namely traditional and modern. A modern data warehouse is characterized by the use of big data technology in its implementation. This data can be taken from various information spreads on the internet such as social media, sensors, blogs, videos, and audio as data sources. Meanwhile, the traditional data warehouse sources are only limited to the transaction and operational data that exist in the university environment.

Yulianto [7] has developed a multidimensional data warehouse model for the Integrated Academic Fee (IAF) in universities. This research follows 4 stages in the development of Business Intelligence (BI), namely preparation, integration, analysis, and visualization. The results of the study can be used as part of the admission DSS.

Asroni et.al [16] investigated the implementation of data warehouses in universities to manage alumni data through data tracer studies. The output of this research is to produce a reporting system using the SQL Server Analysis Service (SSAS) tool to view various dimensions such as alumni profile, department, faculties, and salaries. This information is presented in the form of graphs, tables, and diagrams.

B. Data Warehouse Concept

A data warehouse is a collection of integrated databases which is subject-oriented and designed to support decision-making functions [9]. The data flow in the data warehouse comes from the operational level which is transformed into the data warehouse [10]. According to Thakur [13] mentions that data warehouses have data needs that change from time to time. Thus, this will cause dynamic changes in data storage. A data warehouse is a database designed to perform analysis of decision making, where data and information are generated from the ETL (Extract, transform, Load) [14].

Seen from the infrastructure aspect, the data warehouse consists of several technical components that can be grouped into two categories, namely operation infrastructure and physical infrastructure, such as server hardware, operating system, network software, database software, LAN, WAN, vendor resources, persons, procedure and training [15].

C. Multidimensional Data Modeling

A multidimensional data can be implemented into a star-schema model and can use join operations to relate the tables that exist in it during the query process. The star-schema model consists of a fact table in the middle, then surrounded by a dimension table [11]. A good multidimensional data in data warehouse development should have a simple database structure. It aims to speed up the query process which will be carried out in the analytical stage. The fact table contains facts or measures that are used as business parameters, while the dimension table contains descriptions for query processing.

Several advantages are obtained when implementing a star-schema in data warehouse development, including simplifying query, simplifying reporting logic, improving query, performance, and accelerating data aggregation [17].

In the star-schema model, the fact table and dimension table are connected by a key known as a surrogate key which acts as the primary key of the dimension table and becomes a foreign key in the fact table [18]. This relation occurs logically and can be used to perform the JOIN process in the Query command.

III. METHOD

This study involved data on the activities of lecturers in research at the Politeknik Negeri Sriwijaya within a span of 3 years, from 2018 to 2020. The activity data comes from research activities organized by the Ministry of Education, Culture, Research, and Technology, and activities organized by the university internally.

Meanwhile, the software used in building the system in this research is the framework Codeigniter, with MariaDB/MySQL as the database engine, PHP as the scripting engine, and Apache as the webserver.

Fig. 1 is a stage in the development of a multidimensional database. The stages of system development in this study refer to what was conveyed by Zea [12] who built a data warehouse system in several stages as follows:

- Requirement analysis, in this stage some information is formulated which will later be presented in the data warehouse.

- Data source analysis, the data source is taken from the existing database in the university's internal environment.

- Multidimensional modeling, this stage is carried out to formulate a data warehouse design to describe the relationship between fact table data and dimension tables.

- ETL process, this stage is used to carry out the data retrieval process, transformation process, and data storage into the data warehouse.

- Reporting, this stage presents data in the form of summary information and other important information that acts as output in system design.
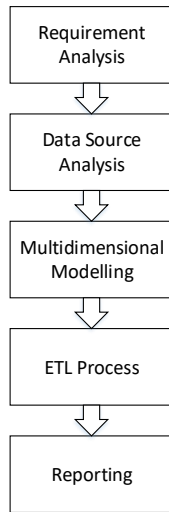
Fig. 1.   Step of Design System.

IV. RESULT AND ANALYSIS

A. Requirement Analysis

This study aims to determine for the availability of data and information as shown in Table I. The information presented in the dashboard as an application on the end-user side can be displayed in the form of tables and graphs. A data table is used to display data in tabular form in which there are facilities for sorting, searching, and paging. Meanwhile, the information in the form of graphs is presented to display a summary of the research data that makes it easier to understand the information.

B. Data Source Analysis

Fig. 2 is an architectural model developed in this study. The data source comes from data sources originating from the academic information system (Sistem Informasi Akademik, abbreviated as SISAK) to retrieve lecturer and department data information. Meanwhile, data on lecturers' research activities come from research information system data (SIMP3M) to obtain data on research schemes, research contracts, and research funding.

Furthermore, at the data warehouse stage, the staging process is carried out to store the extracted data from the data source which has been modified successfully to finally be loaded into the multidimensional database in the data warehouse (DW). Then, through the web browser application, there is a dashboard to process reporting, query, visualization, and analysis of the previously formed multidimensional data.

Data derived from the data source can be taken from tables, or with data in the form of csv files. This csv file type is used to facilitate the ETL process which will later be implemented in the system.

C. Multidimensional Modeling

Researchers logically designed a multidimensional data model using a star schema as can be seen in Fig. 3. This star-schema model is used logically to facilitate the query process later. This star-schema consists of a number of dimension tables and fact tables.

In Fig. 3, there is surrogate_key (SK) in each table, both fact table, and dimension table as a unique record marker in the table and facilitates the data query process. Based on the need for data source analysis that has been defined in the previous stage, there are several dimension tables and fact tables as follows:

1) lecturer_dim: The lecturer_dim table is used to store information on the lecturer who acts as a researcher in research activities.

2) department_dim: The department_dim table is used to store information on departments that are part of research activities and the home base of lecturers or researchers.

3) contract_dim: The contract_dim table is used to store research contract information in every research activity carried out by lecturers.

4) schema_dim: The schema_dim table is used to store research scheme data that can be followed by a lecturer in conducting research activities.

5) date_dim: The date_dim table is used in the system to store date information and hierarchical information such as month name, month number, quarter, and year.

6) funding_fact: The funding_fact table is used to store information on research conducted by lecturers. This table is a fact table that contains the research history.

In each table, both dimension table and fact table, there are effective_date and expiry_date columns to indicate whether a record is active or not used in the system for query and data retrieval processes.
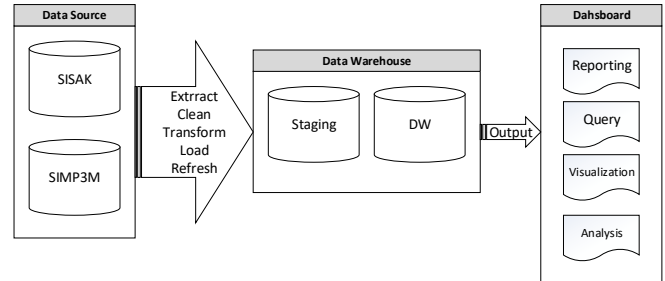

Fig. 2.   Architecture of System.

TABLE I.   INFORMATION REQUIREMENT

| No.	Requirement		
	Information	*Format*	*Timeframe*
1	All Lecturer/ Researcher	Table	All Period
2	All Research Funding	Table	All Period
3	All Research Contract	Table	All Period
4	All Research Scheme	Table	All Period
5	All Departement	Table	All Period
6	Top 5 Lecturer/ Researcher	Chart	All Period
7	Top 5 Department Research	Chart	All Period
8	Top 5 Research Scheme	Chart	All Period
9	Research By All Department	Chart	All Period
10	Research By Department and Year	Chart	Custome Filter
11	Trend Research Funding	Chart	Custome Filter
12	Funding Amount	Chart	Custome Filter
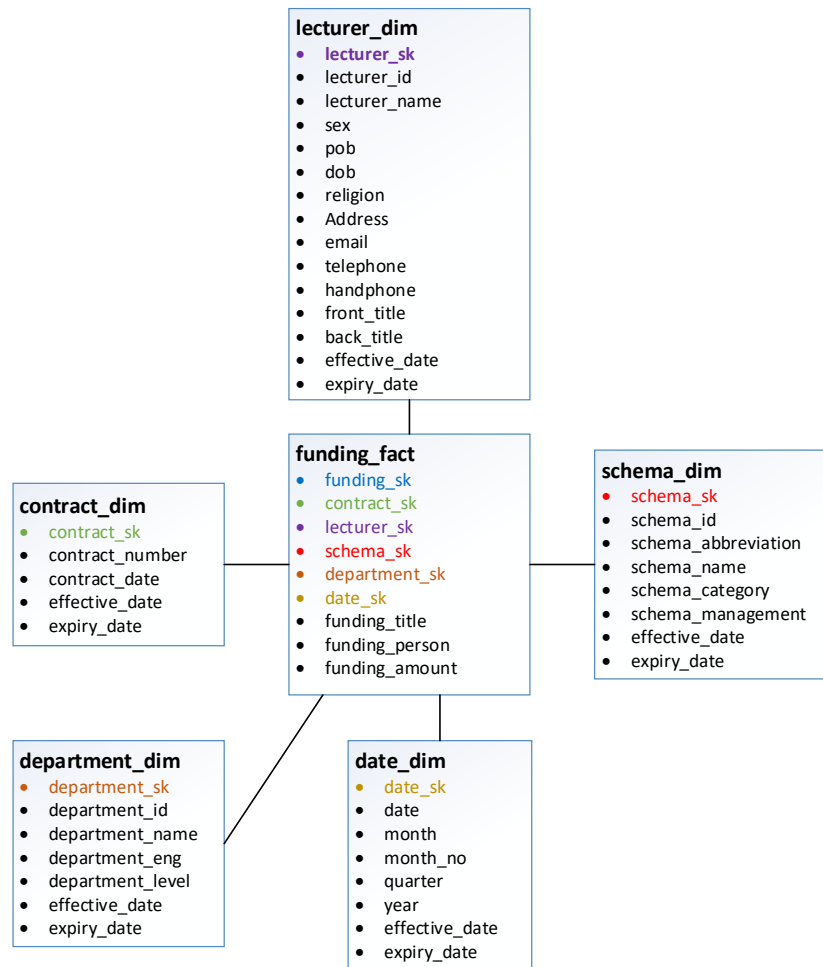
Fig. 3. Star-Schema Model.

D. ETL Process

In the development of this data warehouse, the activity that is quite time-consuming is the ETL (Extraction, Transformation, and Loading). This activity includes the source extraction and performing data population. This activity is to ensure that the data taken from the data source Table I is guaranteed its integrity and validity. Table I is a matrix that describes the ETL (Extract, Transform, Load) process mechanism in the system design. The extract is a step to get data from a data source which is then stored in the data warehouse, while the transform process is a process to prepare data, and load is a process to store data into the data warehouse. In practice, this ETL stage is not a separate stage, but sometimes it is a stage that is an integrated series of processes.

Referring to Table II, there are two mechanisms for the data extraction process from a data source, namely, push and pull. Pull mode is a process to pull data from the data source which is carried out by the data warehouse system. In the design of this system, all data retrieval processes use pull mode. Meanwhile, push mode is a data extraction process carried out by the data source to send data to the data warehouse.

Judging from the data, whether to be sent or withdrawn into the data warehouse, there are two approaches [11], namely, whole part and change data capture (CDC). The whole part is a process to retrieve overall data from data sources when filling out the data warehouse. These data are master data or reference data, such as lecturer data, department data, and schemas which in the timeline review rarely add or update data. Meanwhile, data on contract and funding is processed by the CDC because they are transaction data and there is often be a process of adding and updating data. Pulling data with pull mode is done with the last data change made from the previous pull mode stage.

TABLE II. SOURCE EXTRACTION MATRIX

Data Source	Source Extraction		
	Data Warehouse Table	*Extraction Mode*	*Loading Type*
lecturer	lecturer_dim	Whole, Pull	SCD
department	department_dim	Whole, Pull	SCD
schema	schema_dim	Whole, Pull	SCD
contract	contract_dim	CDC, Pull	unique contract number
funding	funding_fact	CDC, Pull	unique contract number
n/a	date_dim	n/a	pre-population

The process of loading data for the lecturer_dim, department_dim, and schema_dim tables is carried out with SCD (Slowly Changing Dimension) type 1 by replacing the data in certain columns where the data is updated. Meanwhile, the contract_dim and funding_fact tables are carried out with the CDC when the data update process occurs based on the unique contract number. Meanwhile, the date_dim table is not contained in the data source but is generated in the data warehouse system with pre-loading population mode by retrieving and extracting the date parameters during the ETL process.

As an additional feature in the development of this project, part of the data is taken from data sources that come from RESTful web services. This allows two heterogeneous data sources to communicate through a web services interface. The data taken from these web services is used to display summary information in the form of graphic visualizations, making it easier to understand the information presented.

E. Reporting

At this reporting stage, a dashboard is provided in the form of a web application that can present data in the form of tables and graphs. The information presented in the dashboard is generated from the existing SQL commands in the system. The use of star-schema was chosen in this study to facilitate and speed up the query process.

Fig. 4 provides information on the dashboard display that is run on the end-user, side, at the top, there are main menu options consisting of the dashboard, academic, research, human resource, and student. Meanwhile, at the bottom, there is summary information about the number of lecturers, the number of students, the number of departments, the number of research schemes, the number of research funding, and the number of contract funding. In the next section, there is ranking information in the form of top 5 for department, researcher, and funding scheme from lecturer research activities carried out. The following is an example of a query command that is used to produce output in the form of a top 5 research (lecturer) graph as shown in Fig. 4.

```
elect `b`.`lecturer_name` AS
`lecturer_name`,`c`.`department_level` AS
`department_level`,`c`.`department_eng` AS
`department_eng`,count(0) AS `qtyresearch` from
((((((`funding_fact` `a` join `lecturer_dim` `b`) join
`department_dim` `c`) join `schema_dim` `d`) join
`contract_dim` `e`) join `date_dim` `f`) where `a`.`lecturer_sk`
= `b`.`lecturer_sk` and `a`.`department_sk` =
`c`.`department_sk` and `a`.`schema_sk` = `d`.`schema_sk`
and `a`.`contract_sk` = `e`.`contract_sk` and `a`.`date_sk` =
`f`.`date_sk` group by `b`.`lecturer_name` order by count(0)
desc limit 5
```

Meanwhile, in Fig. 5 there is information in the form of graphs that describe the distribution of research activities of lecturers spread across several departments within Politeknik Negeri Sriwijaya. This information is generated in the all-period time range according to the number of data records in the database. Information is presented in a bar chart where the x-axis is the department's data, while the y-axis is the number of studies conducted by lecturers as researchers in the department.

F. Performance Evaluation

To see the extent of the performance generated from the developed model, a performance evaluation is carried out by looking at the payloads (bytes) and response time (milli second) of the system. By using network monitoring, a snippet of data is obtained from a web page which is displayed as shown in Table III. It can be seen in Table III that regardless of the number of payloads or data transferred from the server to the client, judging from the response time, the execution time is not much different. As discussed in the ETL Process section, the data source used in this model comes from an internal domain and a RESTful Web Services from a different web domain (cross domain).
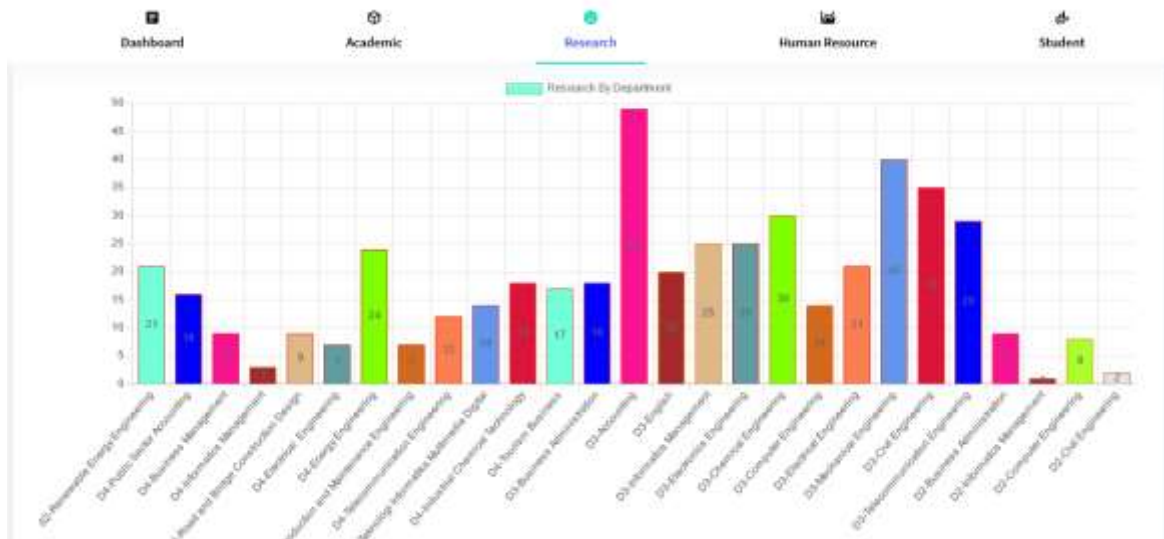

Fig. 4. Dashboard of System.

Fig. 5.   Information of Research by all Departments.

TABLE III.   PERFORMANCE OF APPLICATION MODEL

Services Number	Evaluation Performance		
	Domain	*Payload (Bytes)*	*Time (ms)*
1.	Cross Domain	1670	62
2.	Cross Domain	972	46
3.	Internal Domain	983	61
4.	Internal Domain	814	73
5.	Internal Domain	645	71
6.	Internal Domain	710	61
7.	Internal Domain	743	65
8.	Cross Domain	652	61
9.	Cross Domain	654	60
10.	Cross Domain	651	63
11.	Internal Domain	527	63
12.	Internal Domain	528	64
13.	Internal Domain	526	62
14.	Internal Domain	573	60
15.	Cross Domain	748	54
16.	Cross Domain	719	63
17.	Cross Domain	711	63
18.	Cross Domain	1060	62

By looking at Fig. 6 provides an overview of the distribution of payloads and response times executed by web browsers. By looking at the fast and relatively stable response time for any amount of payloads, the model of this system development can be utilized in developing business intelligence applications by using star schemes and RESTful web services in the application architecture.
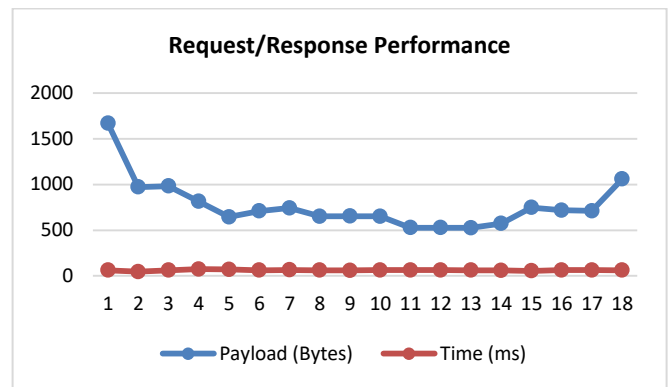

Fig. 6.   Payloads and Response Time of System.

V.   CONCLUSION

Through this research, a multidimensional data model has been built that contains information about the research activities of lecturers in the Politeknik Negeri Sriwijaya. This study includes 446 lecturer data records, 34 department data records, 483 record research data records spread over the period of the study 2018, 2019, and 2020. The information generated from the developed model can help higher education management carry out strategic planning. and decision-making. In addition, the resulting information can be used in the preparation of data in accreditation instruments, both for university or institutional accreditation and study program accreditation. The star-schema model was chosen in this study to facilitate the multidimensional database modeling process and to speed up the query process carried out in the reporting and presentation stages of data in a graphical form that can be displayed in a web browser application. Further research from this research is optimizing the data analysis process, and developing data mining applications from data that has been successfully processed.

REFERENCES

[1]   L. W. Santoso and Yulia, "Data Warehouse with Big Data Technology for Higher Education", Procedia Computer Science, vol. 124, pp. 93-99, 2017.

[2] M. Baranovic, M. Madunic, I. Mekterovic, "Data warehouse as a part of the higher education information system in Croatia", in Proceedings of the 25th International Conference on Information Technology Interfaces (ITI), 2003, pp. 121 - 126.

[3] X. Yu, "The Application of Data Warehouse in Teaching Management in Colleges and Universities", J. Phys.: Conf. Ser. 1738 012090, 2021.

[4] Y. Bassil, "A Data Warehouse Design for A Typical University Information System", Journal of Computer Science & Research (JCSCR), vo. 1, no. 6, 2012, pp. 12 – 17.

[5] W. Inmon, "Building the Operational Data Store", 2nd ed. John Wiley & Sons, 1999.

[6] D. Bogdanova and M. Snoeck , "CaMeLOT: An Educational Framework For Conceptual Data Modelling", Information and Software Technology (2019), doi: https://doi.org/10.1016/j.infsof.2019.02.006.

[7] A. A. Yulianto, Y. Kasahara, "Data Warehouse System for Multidimensional Analysis of Tuition Fee Level in Higher Education Institutions in Indonesia", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 6, 2020, pp. 541-550.

[8] DRPM, "Panduan Peneltian dan Pengabdian Kepada Masyarakat Edisi XIII", Kementrian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2020.

[9] Rudy, E. Miranda, E. Suryani, "Implementation of Data Warehouse, Datamining and Dashboard for Higher Education", Journal of Theoretical and Applied Information Technology, vo. 64, no. 3, 214, pp. 710-717.

[10] W. H. Inmon, "Building the Data Warehouse 4th ", Wiley Publishing inc, 2005.

[11] D. Darmawikarta, "Dimensional Data Warehousing with MySQL: A Tutorial", Brainy Software Corp. 2007.

[12] O.M. Zea, J.P. Gualtor, S.L. Mora, "A Holistic View of Data Warehousing in Education", IEEE Access, vo. 6, p. 64659-64673, Okt 2018, DOI:10.1109/ACCESS.2018.2876753.

[13] G. Thakur, A. Gosain, "A Comprehensive Analysis of Materialized Views in a Data Warehouse Environment", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 5, 2011, pp. 76-82.

[14] A. Chakir, H. Medromi, A. Sayouti, "Actioins for Data Warehouse Success", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 8, 2013, pp. 130-133.

[15] A.F. Neamah, "Adoption of Data Warehouse in University Management: Wasit University Case Study", 2021 J. Phys.: Conf. Ser. 1860 012027.

[16] Asroni, T. B. Arimbi, S. Riyadi, "Implementing of Data Warehouse Data Alumni using the Single Dimensional Data Store Method", J. Phys.: Conf. Ser. 1471 012021.

[17] K.A. Shobirin, A.P.S. Ikandar, I.B.A. Swamardika, "Data Warehouse Schemas using Multidimensional Data Model for Retail", International Journal of Engineering and Emerging Technology, vol. 2, no. 1, January-June 2017, pp.84-86.

[18] Y. Peng, "Metadata Based Visualization System for Multidimensional Data", Advances in Intelligent Systems Research, vol. 134, 2017, pp.564-567.