



Academic Data Warehouse Modeling in Higher Education Using Nine-Step Design Methodology

Denny Jean Cross Sihombing¹

¹Information System Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: ¹denny.jean@atmajaya.ac.id

Abstract

Data and information are essential in various fields today, as well as in the field of education, especially in universities. Some universities already have information systems that support data and information needs. However, the system has not been integrated, so it cannot provide data and information needs quickly and in an integrated manner. Information systems in universities are still primarily departmental because each was built at a different time and uses another platform. The departmental nature of this information system causes inaccuracies and inconsistencies of data that drive the information produced in reports and data reused in transactions to be invalid. Invalid data, in the end, also impacts decision-making taken by management. This study aims to develop a data warehouse at a university to integrate academic data using a star schema. The method used is the Nine Step Methodology. The result of this research is data warehouse architecture used in the academic field; fact tables and ERDs have been designed at the current stage of designing a Prototype of the Study Program Performance Sheet (LKPS).

Keywords: Data Warehouse, Star Schema, Nine Step Methodology

1. INTRODUCTION

Higher Education as an educational institution is required to provide good services to all stakeholders such as students, alums, parents of students, the industrial world, society, and the government. One of the services needed is academic information services such as student scores, study results, study processes, payment obligations, etc. A reliable and suitable information system is needed to provide these services. On the other hand, data is very crucial for decision-makers because it has a significant influence on the development of the organization. For this reason, data must be precise, accurate, and can be accessed quickly to be processed into useful information. Integrated information certainly has its advantages for management in setting strategies[1].

Data warehouse plays an essential role in helping the alignment between strategies, processes, and technologies that can increase an organization's competitiveness. By applying a data warehouse, a data processing process can be carried out in large quantities. It can effectively analyze performance so that data in the organization can be processed into valuable information for competitive advantage. On the other hand, decision-making in the academic field by management requires integrated, non-departmental, relevant data for an extended period [2]. Data from each operational/transactional information system



must be integrated and grouped based on a specific subject and with relevant time dimensions so that reports can be generated quickly and easily and do not interfere with operational systems that can ultimately be used to help college management to make strategic decisions related to the academy [3], [4].

The data warehouse integrates data from multiple diverse sources of information and transforms it into a multidimensional representation for decision support applications. ETL is a set of processes that must be passed in the formation of a data warehouse[5]. The purpose of ETL is to organize, filter, process, and integrate suitable data from various references to be stored in a data warehouse. ETL process [6] consists of extracting, transforming, and loading. Some data warehouse development research uses data modeling using Star schema. Star Schema is a kind of star-shaped data warehouse schema model with dots emanating from the center where data is organized into fact tables and dimensions[7]. This data warehouse schema model is the most accessible model to analyze and execute data warehouses because table dimensions do not need to be normalized[8]. Research [9] applies the star scheme to data warehouse design at educational institutions in Iranian. Online Analytical Processing (OLAP) is a computer processing technology that enables the fast execution of complex analytical queries. OLAP serves to query data in complex forms and analyze large volumes of data [10] [11].

There are several methodologies for building a data warehouse, one of which is the Kimball method. Kimball & Ross [11] said there are nine steps in building a data warehouse, known as the nine-step design methodology. If the nine-step design methodology steps are carried out systematically, it can build a good data warehouse. Research [1] on developing data warehouses in education shows that the application of data warehouses is beneficial for decision-making effectiveness.

A data warehouse is a technique that retrieves and reduces data periodically from the reference method into a dimension or normalized data store [3]. It usually contains several years of history and is presented for business intelligence or other analytical activities. Updated within a certain period, not every time a transaction occurs in the source system [6]. The components contained in the data warehouse architecture: Subject Oriented, namely warehouse data, is arranged based on the main subjects in the company. The data warehouse analyzes and makes decisions regarding the history of these critical subjects rather than being used to support operational applications. Integrated, namely, data in a data warehouse comes from various separate sources. The data entered into the warehouse will be changed, rearranged, sorted, and summarized to maintain data consistency in the data warehouse to create an integrated data picture of the company. Non-Volatile: Data in the data warehouse is not updated in real-time but is updated regularly from operational data in large amounts of data. A data warehouse's characteristics differ from the operational characteristics of data that can be changed. Data in the warehouse can only load data (retrieve data) and access data. Time Variant means that each data unit in the data warehouse is accurate in a specific time interval. On each record, a form of the timestamp indicates when the time of a record is accurate. For example, timestamping giving a transaction date for each record [4].

This study aims to develop a data warehouse that integrates academic data using the star scheme (star schema). In this model scheme, the relationships in each table will be used as

a value as a limit (dimension) with a table of magnitude (measure) to describe the desired information needs to answer the information needs of system users. The method used is Nine Step Methodology. Nine Step Methodology is chosen because it has a structured design stage and can be tailored to the user's needs. The benefits of this research, universities have a data warehouse design that can be used as a basis for making academic-related decisions and a data warehouse infrastructure design in the academic field that can be developed for other fields.

2. Research Methods

The stages in this study consist of three main stages: data collection, analysis of information needs, and design of a data warehouse, as shown in Figure 2.

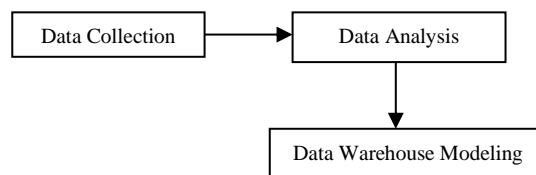


Figure 1. Methods

2.1 Data Collection

The data collection technique is observation and interview. Data Collection involves structural leaders at the Faculty at the University.

2.2 Data Analysis

The data that has been collected will be analyzed. In this stage, conduct a specification analysis of the data warehouse needs in the data warehouse analysis as an Information Package.

2.3 Data Warehouse Modeling

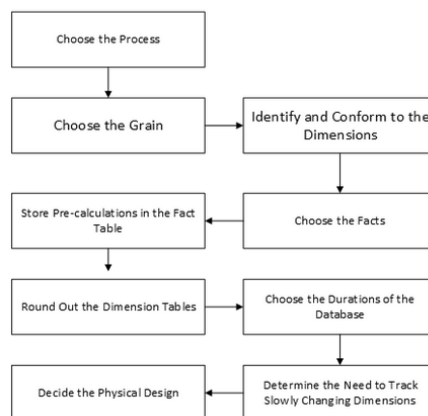


Figure 2. Nine-Step Design

A data warehouse model for academics developed using the nine-step design methodology. Stages of the nine-step design methodology [2], [4], [12]–[15] as show in Figure 2.

3. RESULTS AND DISCUSSION

In this study, the data used was primary data from one of the private universities. This research involved faculty leaders and those involved in the business processes of the college. The development of this system refers to the needs of the system. System development refers to the nine-step design methodology as show in Figure 3.

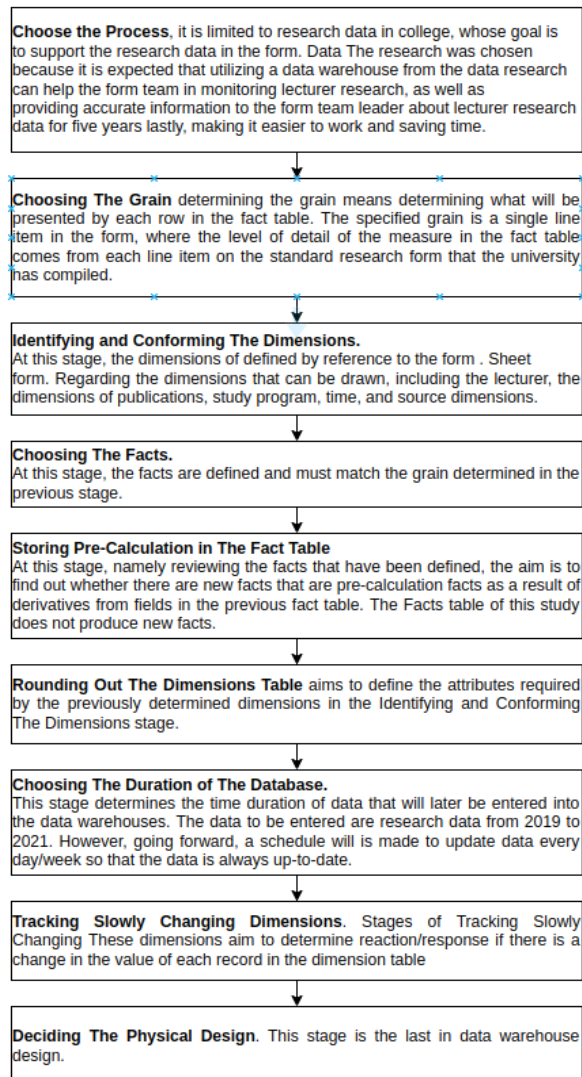


Figure 3. Implementation Nine-Step Design

Choosing a process means determining the main subject. The main subject refers to an organization's business activity that can answer all essential business questions and have specific features. In research, the main subject is the Academic field. Choosing a grain means determining what a table of facts represents or presents. After determining the grain of the fact table, it can then determine the dimension tables that relate to the fact table. The grain on the fact table also determines the grain of the dimension table. Identify and associate dimension tables with fact tables. Dimensions are a collection of points of view essential for describing the facts in the fact table. The grain of a fact table determines the facts that can be used. At this stage, determine the required measures in the fact table. Results of calculations on an attribute need to be considered for storage in the database, reducing the risk of errors in the program every time it performs calculations on these attributes. A description contains structured information about the attributes in the dimension table. The dimension table should be fully denoted and easily understood by the user. The duration of the data entered into the data warehouse is determined at this stage. Dimensions can change slowly and become a problem. There are three basic types of slow dimensional changes: re-author the changed attribute, create a new record on the dimension and create an alternative attribute to hold the new value. The physical design of the data warehouse is carried out. In addition, the determination of the problems that may exist in physical design.

Its principal subject is the college accreditation procedure. The accreditation procedure consists of five implementation stages: the proposal of an accreditation proposal document uploaded /submitted by the university. Acceptance of accreditation proposal documents: at this stage, the completeness of the accreditation documents confirms the requirements. The adequacy assessment process reviews accreditation proposal documents by the Assessor Team. The area assessment approach is a visitation activity by the Assessor Team to universities to verify facts and field conditions against the data/information submitted in the documentation of the accreditation proposal. Determination of accreditation results by National Accreditation Agency for Higher Education (BAN PT) and delivery of accreditation results to universities.

The detailed activities in each implementation stage can be seen in the cross-functional flowchart diagram presented in Figure 3. In this study, the focus of the data warehouse architecture (Figure 4) was used for the Study Program Performance Sheet (LKPS). Data modeling used for LKPS as in Figure 5. The database used in developing the LKPS system is composed of twenty-four tables, each with a relationship with other tables. The twenty-four tables will contain the data needed for accreditation assessment activities, such as data on cooperation with other institutions, lecturers, students, research, articles, recognition, PKM, study programs, student surveys, student achievement, courses, alums, and selection. Available students at the university.

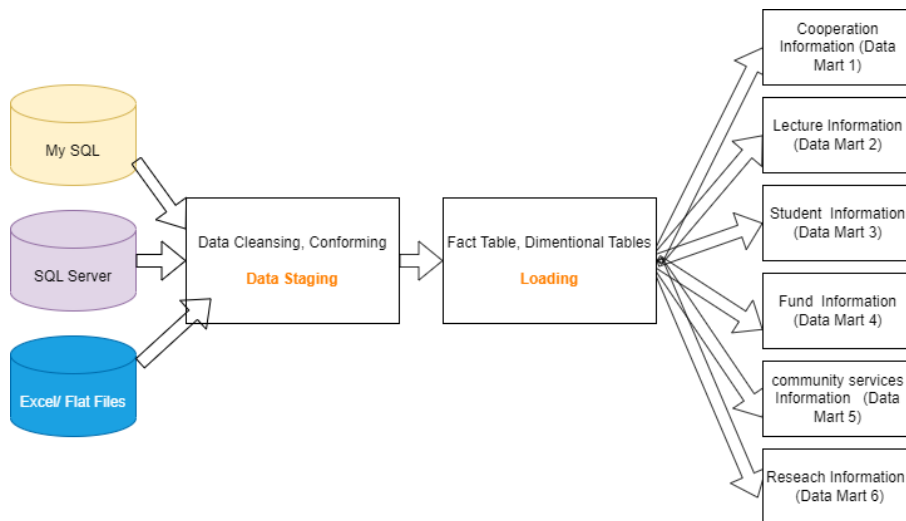


Figure 4. Data Warehouse Architecture

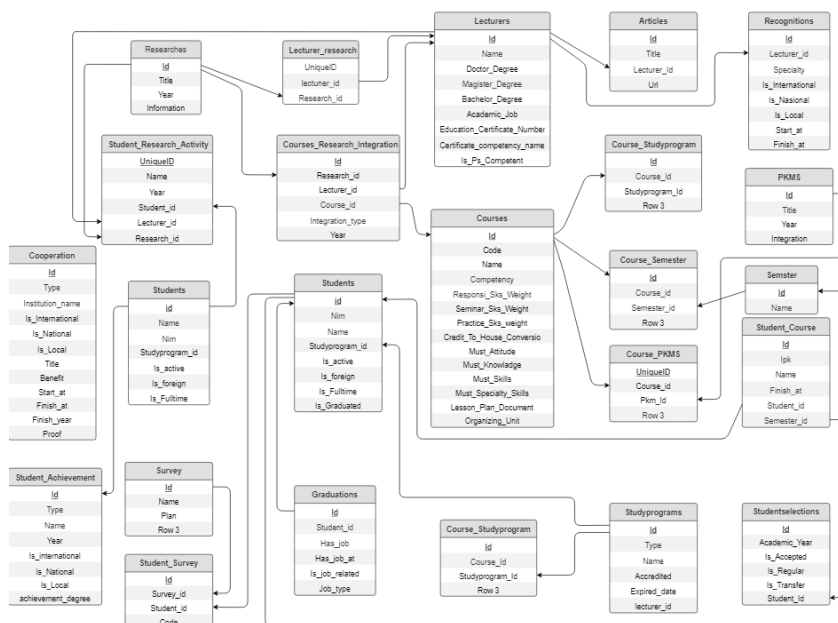


Figure 4. ER-Diagram

Implementation of Study Program Performance Sheets Based on the database design, a prototype of the Study Program Performance Sheet (LKPS) was made, which consisted of the Lecturer Menu, Achievements, Articles, Courses, Cooperation, Graduation, Community Service, Recognition, Research, Students, Study Programs, and Surveys, as shown in Figures 5. The Information

Cooperation menu (*Kerjasama*) displayed consists of cooperation in education, research, and community service. On the lecturer (*Dosen*) menu are the information displayed by the university's permanent lecturer, the final project's primary supervisor, and the equivalence of full teaching time. On the other hand, the information displayed is also data on part-time lecturers who come from practice. The information displayed on the Research (*Penelitian*) Menu is publications from lecturers and information related to citations and research preparations. On the fund's (*Penggunaan Dana*) menu, the information shown is the flow of funds in the university, the source of funds, and the management of funds at the university. The curriculum (*Kurikulum*) menu contains information related to course structure, learning outcomes, and learning plans. On the other hand, this menu will also display courses integrated with research and community service. On the graduation GPA (*IPK Lulusan*) menu, the information shown is related to students' academic expectations, the period of study, and the suitability of the student's field with his current job after graduation.

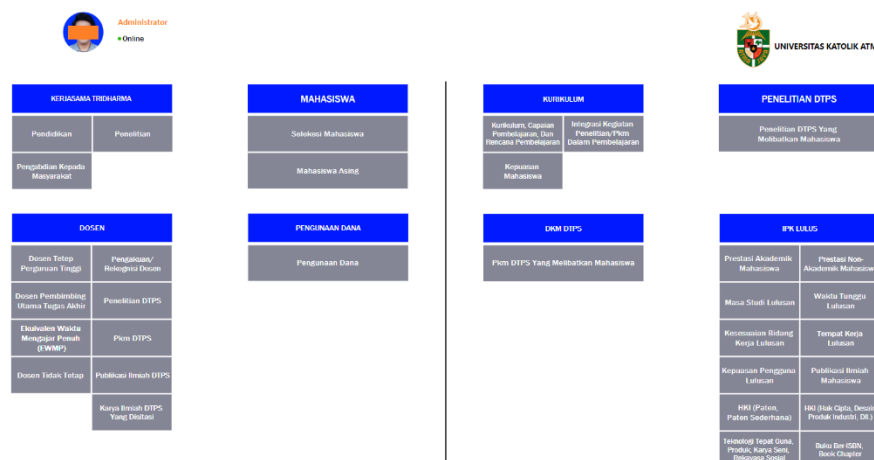


Figure 5. UI Design Main Menu Application

4. CONCLUSION

Data warehouse recreates a critical role in supporting the alignment between strategies, processes, and technologies that can improve an institution's competitiveness. By applying a data warehouse, a data processing process can be carried out in large quantities. It can effectively investigate performance so that data in the institution can be processed into helpful knowledge for competitive benefit. The result of this research is data warehouse architecture used in the academic field; fact tables and ERDs have been designed at the current stage of designing a Prototype of the Study Program Performance Sheet (LKPS). Further research to develop applications and test the success of the application.

REFERENCES

- [1] L. W. Santoso and Yulia, "Data Warehouse with Big Data Technology for Higher Education," in *Procedia Computer Science*, 2017, vol. 124, pp. 93–99. doi: 10.1016/j.procs.2017.12.134.
- [2] S. Bouaziz, A. Nabli, and F. Gargouri, "Design a data warehouse schema from document-oriented database," *Procedia Comput Sci*, vol. 159, pp. 221–230, 2019, doi: 10.1016/j.procs.2019.09.177.
- [3] M. Mirzaei, N. Zaerpour, and R. de Koster, "The impact of integrated cluster-based storage allocation on parts-to-picker warehouse performance," *Transp Res E Logist Transp Rev*, vol. 146, Feb. 2021, doi: 10.1016/j.tre.2020.102207.
- [4] M. AlMeghari, S. Taha, H. Elmahdy, and X. Shen, "A proposed authentication and group-key distribution model for data warehouse signature, DWS framework," *Egyptian Informatics Journal*, no. xxxx, 2020, doi: 10.1016/j.eij.2020.09.002.
- [5] A. Filiana, A. G. Prabawati, M. N. A. Rini, G. Virginia, and B. Susanto, "Perancangan Data Warehouse Perguruan Tinggi untuk Kinerja Penelitian dan Pengabdian kepada Masyarakat," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, pp. 174–183, 2020, doi: 10.28932/jutisi.v6i2.2557.
- [6] M. Souibgui, F. Atigui, S. Zammali, S. Cherfi, and S. ben Yahia, "Data quality in ETL process: A preliminary study," *Procedia Comput Sci*, vol. 159, pp. 676–687, 2019, doi: 10.1016/j.procs.2019.09.223.
- [7] V. Khatibi, A. Keramati, and F. Shirazi, "Deployment of a business intelligence model to evaluate Iranian national higher education," *Social Sciences & Humanities Open*, vol. 2, no. 1, p. 100056, 2020, doi: 10.1016/j.ssaho.2020.100056.
- [8] A. Cuzzocrea, "SpPolap: Computing privacy-preserving OLAP data cubes effectively and efficiently algorithms, complexity analysis and experimental evaluation," *Procedia Comput Sci*, vol. 176, pp. 3831–3842, 2020, doi: 10.1016/j.procs.2020.09.337.
- [9] D. Nurmallasari, D. H. Qudsi, M. S. Zulvi, and W. Nengsih, "Pemodelan Data dengan Skema Galaksi pada Data Lulusan," pp. 123–129, 2020.
- [10] O. A. Omitaomu et al., "A new methodological framework for hazard detection models in health information technology systems," *J Biomed Inform*, vol. 124, Dec. 2021, doi: 10.1016/j.jbi.2021.103937.
- [11] R. Kimball, M. Ross, and A. A. Anisimov, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (2nd Edition)*, vol. 32, no. 3. 2003. doi: 10.1145/945721.945741.
- [12] L. W. Santoso and Yulia, "Data Warehouse with Big Data Technology for Higher Education," *Procedia Comput Sci*, vol. 124, pp. 93–99, 2017, doi: 10.1016/j.procs.2017.12.134.
- [13] R. Martins, M. T. Pereira, L. P. Ferreira, J. C. Sá, and F. J. G. Silva, "Warehouse operations logistics improvement in a cork stopper factory,"

- in *Procedia Manufacturing*, 2020, vol. 51, pp. 1723–1729. doi: 10.1016/j.promfg.2020.10.240.
- [14] S. W. Y. Cheng, K. L. Choy, and H. Y. Lam, “A workflow decision support system for achieving customer satisfaction in warehouses serving machinery industry,” in *IFAC-PapersOnLine*, May 2015, vol. 28, no. 3, pp. 1714–1719. doi: 10.1016/j.ifacol.2015.06.333.
- [15] R. K. Singh, N. Chaudhary, and N. Saxena, “Selection of warehouse location for a global supply chain: A case study,” *IIMB Management Review*, vol. 30, no. 4, pp. 343–356, Dec. 2018, doi: 10.1016/j.iimb.2018.08.009.