

# Use of Data Warehouse and Data Mining for Academic Data

## A Case Study at a National University



Primasatria Edastama<sup>1</sup>, Amitkumar Dudhat<sup>2</sup>, Giandari Maulani<sup>3</sup>

University of Esa Unggul<sup>1</sup>, Veer Narmad South Gujarat University<sup>2</sup>, University of Raharja<sup>3</sup>  
Jl. Arjuna Utara No.9, Kb. Jeruk, Kota Jakarta Barat<sup>1</sup>, Udhana - Magdalla Rd, Surat, Gujarat<sup>2</sup>,  
Jenderal Sudirman No.40, Cikokol, Kota Tangerang<sup>3</sup>,  
Indonesia<sup>1,3</sup>, India<sup>2</sup>

e-mail: [primasatria@esaunggul.ac.id](mailto:primasatria@esaunggul.ac.id)<sup>1</sup>, [amit000790@gmail.com](mailto:amit000790@gmail.com)<sup>2</sup>, [giandari@raharja.info](mailto:giandari@raharja.info)<sup>3</sup>



**Author Notification**  
October 2021  
**Final Revised**  
October 2021  
**Published**  
October 2021

### To cite this document:

Edastama, P., Dudhat, A., & Maulani, G. (2021). Use of Data Warehouse and Data Mining for Academic Data; A Case Study at a National University. *International Journal of Cyber and IT Service Management (IJCITSM)*, 1(2), 206-215. Retrieved from <https://iast-journal.org/ijcitsm/index.php/IJCITSM/article/view/55>

### DOI:

<https://doi.org/10.34306/ijcitsm.v1i1.55>

### Abstract

*If an organisation has full, rapid, exact, and accurate information, it may perform better in terms of evaluation, planning, and decision-making. The essential information can be derived from operational data maintained in an integrated database by a university that already has an information system. This research looks at how to collect operational data into a data warehouse and then apply data mining techniques to analyse the data. This study yielded a comprehensive data warehouse with a web-based information reporting application. Furthermore, data mining techniques are used to analyse the data warehouse that has been created. The result of the application of data mining is the generation of characteristic patterns of students who take certain specialization programs.*

**Keywords** : Data Warehouse, Data Mining, Academic Data.

### 1. Introduction

Universities are now compelled to use all available resources to gain a competitive edge. Information systems, along with buildings, infrastructure, and human resources, are one of the resources that may be exploited to gain a competitive edge. Information systems can be used to gather, process, and distribute data in order to support both daily operational and strategic decision-making operations. The fact that the existing information system is not yet fully integrated is a common stumbling block, resulting in partial or erroneous information being given.

One technique to retrieve significant information from data dispersed across several information systems is to build a data warehouse [1]. After the data has been integrated, it



may be utilised for information distribution activities that can be seen from multiple perspectives and modified to the degree of detail. Data analysis and data mining techniques are used to make better use of the information in the data warehouse.

A data warehouse was established for the National University in this study (UNAS). UNAS is a Jakarta-based private university that was formed on October 15, 1949. UNAS offers seven faculties with 17 study programmes for undergraduate education (S1), three study programmes for master's education (S2), and three academies with six Diploma Three Study Programs (DIII).

UNAS recognises that relying just on operational data to support decision-making activities is insufficient; a data analysis is required to fully explore the possibilities of current data. As a result, a data warehouse must be built, which can subsequently be evaluated using data mining techniques. The application of data mining techniques is intended to reveal previously concealed knowledge in the data warehouse.

## **2. Data Warehouse Design**

This design phase includes data warehouse architecture design, data source planning, and dimensional data modeling.

### **2.1. Data Warehouse Architectural Design**

The operational data source was the ACADEMIC database, which comprises all UNAS academic data. So that system performance does not interrupt the operational database, data is regularly picked from various data sources and then placed into a secondary database called AKADSL. In data warehouse theory, AKADSL is also known as data staging. Before loading data into the AKADSL database, a validation step is always performed to ensure that the data is legitimate and may be put into the database. Following that, the AKADSL database's data is cleansed and converted. The outcomes of the cleaning and transformation process are subsequently stored in the data warehouse. Figure 1 depicts the UNAS data warehouse's logical architecture design, as well as the method of loading data into the data warehouse.

The physical architecture of the UNAS data warehouse is seen in Figure 2. Users can access the data warehouse using the application server. Because the AKADSL database and data warehouse are on the same machine, the Extraction, Transformation, and Loading processes may be performed on the data warehouse engine rather than the operating machines. The ACADEMIC, AKADSL, and data warehouse databases all employ the PostgreSQL 8.1 RDBMS.

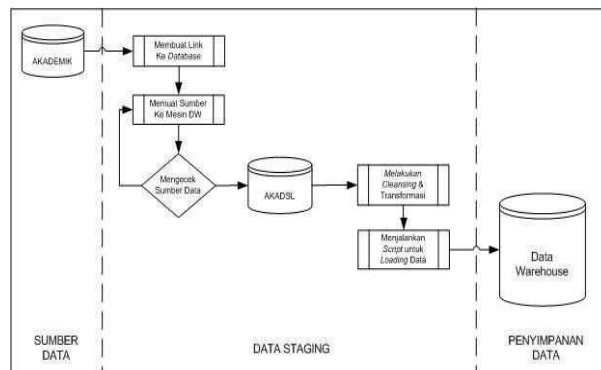


Figure 1. Logical Architecture Data Warehouse UNAS

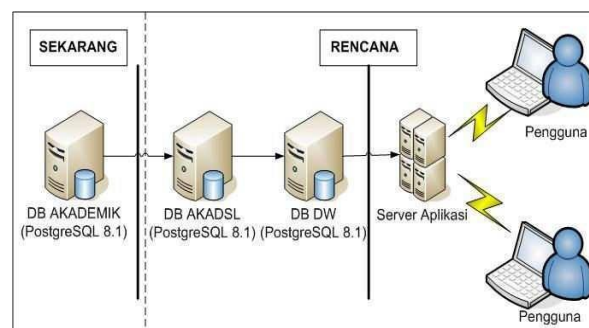


Figure 2. Physical Architecture Data Warehouse UNAS

## 2.2. Data Source Planning

Students, professors, and administrative staff can use a variety of programmes developed by UNAS to assist in the execution of academic activities. Intranet-based apps (available solely on campus through client-server approaches) and web-based applications (accessible over the Internet) were created. The applications are as follows:

- The Academic Information System is the most important academic application (SIA). This programme allows lecturers to fill out grades online.
- Student Academic Services is a student-run programme. This application allows students to fill up their KRS (Card Cards) Study Plan online. You may also examine the grades and class schedules for each subject.
- The New Student Registration Application is a programme that helps new students through the admissions process.
- The Visual Scheduling System (VSS) is a software that keeps track of each faculty member's room schedule during the semester.

## 2.3 Dimensional Data Modeling

The data modelling schema is a star schema with one fact table and many dimension tables. The use of a star schema enables for quicker searching and easier exploration of dimensional data [2]. The fact table that evolved from the data warehouse's design is a table that tracks student academic activity from registration to graduation. The fact tables in the UNAS data warehouse are as follows:

Table camaba fact

Personal information, grades obtained, acceptance status, and registration status are all included in this table for new student admissions. The star schema for new student admissions is depicted in Figure 3.

Tablemhs fact values

This table contains information on student grades, such as the courses taken, the rate of repetition, and the grades received. The star schema for Student Values is shown in Figure 4.

Tabel hasilstudy\_fact

This table contains information on students' academic records, such as the semester achievement index (IPS), cumulative achievement index (GPA), semester credit system (SKS) courses completed, and credits gained during that time period. The student's academic standing is also maintained on file during this period. A star schema for student academic notes is shown in Figure 5.

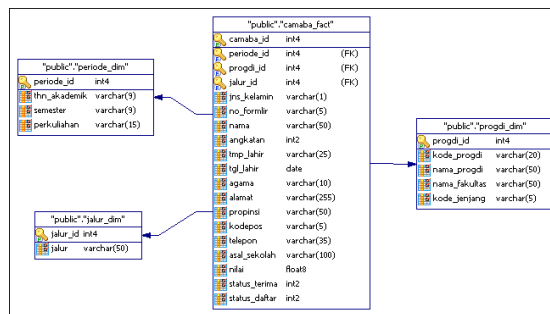


Figure 3. Star Schema New Students

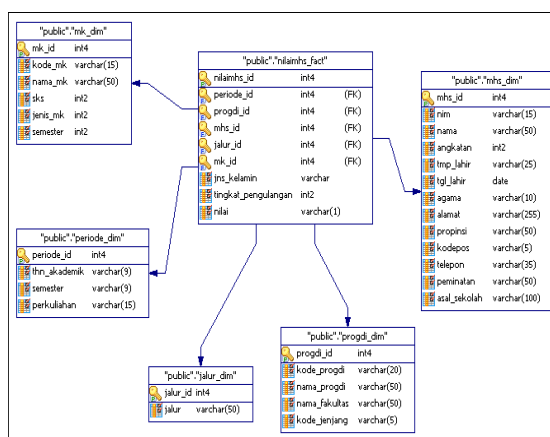


Figure 4. Star Schema Students Value

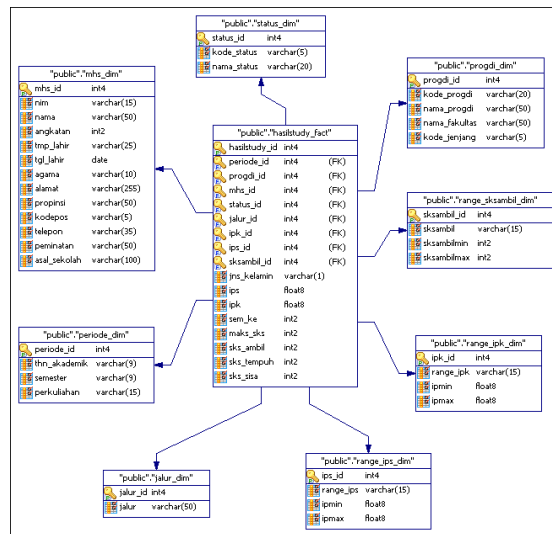


Figure 5. Star Schema Student Academic Notes

A dimension table is a table that contains data that shows an overview from many perspectives. Dimensions allow for the creation of dynamic reports that may be seen from various angles and the level of information changed. The dimension tables in the UNAS data warehouse are as follows:

**Table period dim**

The academic year, semester (even and odd), and lectures are all instances of data with a time component (regular and short).

**Table mhs dim**

Personal data such as name, date of birth, residence, and so on are contained in the student dimension.

**Study dim table**

Information about educational levels, faculties, and departments is included in the department dimension, with each faculty having one or more departments. The educational levels are D3 (Diploma III), S1 (Bachelor), and S2 (Master).

**Table status dim**

Active, Leave, Passed, Unreported Leave, Resigned, and Drop Out are the academic status dimensions.

**Table path dim**

The path dimensions provide you information about how you got into the structure. At UNAS, two new student admissions options are available: the conventional route and the transfer route.

**Table mk dim**

The course dimensions provide information such as the course name, course code, and credits for each course.

**3. Data Warehouse Capabilities**

This data warehouse's ability to give information is used to assess academic activities at UNAS, from registration through graduation, so that university authorities may identify what actions need to be made to improve the quality of academic activities.

### **3.1. Data Warehouse Presentation**

The capacity of a data warehouse to provide information to users is its most important feature. The database platform utilised is PostgreSQL. Users are provided with data via the following tools :

- Mondrian is an OLAP engine that runs on the Java platform [3].
- Jpivot is an OLAP table and chart presentation JSP (Java Server Pages).

A web-based application written in the Java programming language is used to present the data.

### **3.2. Information Provision**

The UNAS data warehouse was established to provide university administrators with timely, accurate, and trustworthy data so that they may make the best decisions possible when conducting academic activities.

Because it is a web-based application, the provided information may be viewed from any computer with an Internet connection. The existence of a data warehouse can also result in the following capabilities:

- Data mapping and analysis are made easier with drill-down and roll-up features. Roll-up refers to the ability to convey facts in a lower level of detail. The ability to present data in more depth is referred to as drill-down.
- You have the freedom to design your own queries based on your needs.
- Options for customising reports based on information needs.
- Ability to create charts or graphs based on the needs of the report.
- The ability to produce and save reports in Excel and PDF formats.

#### **New Student Admission Information**

Information about new student admissions is used to view:

- Per period, the number of potential students that apply, get accepted, and enroll in each department.
- Which majors saw a drop in the number of potential students who applied, were approved, and then registered? So you can figure out what's causing the problem and how to fix it.
- The composition of a given period's new student population.

#### **Information on the Academic Status of Students**

Academic statuses include active, Leave, Graduated, Unreported Leave, Resigned, and Dropout. Academic status information is necessary for assessing the makeup of a group of students over time based on their academic standing.

Details on Distribution A particular amount of credits are needed for students. The number of credits a student can take is decided by their previous semester's IPS (Semester Achievement Index). As a result, the higher a student's social studies grade, the more credits he or she might earn, and the more credits earned, the more credits earned.

The more tests that have been taken, the shorter the student's study period will be.

The number of students who attend lectures is determined using information on the distribution of credits taken, which is based on the range of credits taken. It's also used to figure out how many credits students took on average during a specific period of time.

#### Student Achievement Index Distribution Information

Social Studies is one predictor of student progress over the course of a semester. The higher the grade for a major, the higher the percentage of students in a department who receive good grades in social studies. The UNAS leadership uses data on student social studies distribution to estimate the number of students based on a predetermined social studies range.

#### Information on the Grade Point Average Distribution of Students (GPA)

The grade point average (GPA) is one measure of a student's academic progress over time. The grade point average (GPA) is calculated based on a student's grades throughout his or her academic career. Data on the distribution of student GPAs is used by university officials to predict the number of students based on a set GPA range. It's also used to figure out what a student's average GPA is.

#### Information about the Graduates

A graduate profile is one of the criteria for an academic report that is a university requirement. The following information regarding graduate profiles is viewed:

- Graduate profile depending on degree of study during a specific time period.
- In a certain time period, the number of graduate students. The higher the percentage of on-time graduates, the higher the department's rating.
- Based on the length of study, the number of graduate students.

#### GPA Distribution Data for Graduates

1. University authorities utilize data on the distribution of graduate GPAs to look at:
2. Graduate student quality.
3. A predetermined GPA range is used to select the number of graduate students.
4. Graduate students' average GPA.

### 3.3. Effect of Change

The report preparation process is simplified with a data warehouse since users may customise reports to their satisfaction, cutting the time it takes to build another programme or perform manual labour from one month to one day. As a result, UNAS may have to reduce the amount of human resources allocated before.

to finish report writing so that I may spend more time to other vital chores and improve the quality of my academic activities Furthermore, this data warehouse keeps prior data on student academic status, making it simple to compile previously complex reports.

Because this report may be obtained over the Internet, university administrators have greater flexibility in meeting their needs, allowing them to get information to help in decision-making even if they are unable to visit campus.

#### **4. Application of Data Mining Techniques**

This study used data mining techniques to analyse academic data retrieved from a data warehouse. Because the purification phase was performed when the data warehouse was built, it is no longer essential. The goal of this study is to determine "What are the characteristics of students who pursue particular specialised programmes?" by examining the condition of the fact table and dimensions created in the previous stage.

Students who are unsure about which speciality degree to pursue might use this information to help them make a decision. Supervisors who are familiar with the pattern of characteristics of successful students from a specialisation programme can give more appropriate recommendations in choosing a specialisation programme based on the interests and qualities of students.

##### **4.1. Selection of Data Mining Techniques**

Data mining techniques were used to solve the aforementioned queries utilising classification techniques and association rules [4]. Using the two approaches outlined above, the pattern groups of students who pursue various specialised programmes will be seen. The supporting evidence should be looked into further.

The software used to carry out the data mining process in this research is :

1. CBA. Data mining tool developed by the School of Computing, National University of Singapore [5]. The algorithm used is Classification Based Association.
2. Microsoft SQL Server 2005. The algorithm used is Association Rules.

##### **4.2. Data Mining Modeling**

The data mining modelling in this study adapts to the information needs and characteristics of the data in the data warehouse. The specialised curriculum for this study was chosen from the UNAS Corporate Management study programme. Because this study aims to reveal a pattern of characteristics of successful students from a speciality programme, the dataset used in this study is data from students in the Corporate Management study programme who graduated with a GPA more than or equal to 2.75. Criteria for selection

The specialisation programme has the highest and most comprehensive amount of speciality programmes when compared to other academic programmes.

Because there was insufficient data, the MO specialist programme was not included in the analysis (only 33 data). There were a total of 701 data points utilised in this enquiry.

The characteristics of the data utilised in data mining. The data reveals a link between the course values and the selected speciality, resulting in a distinctive pattern of students who follow a certain specialisation programme. As a consequence, the specialisation programme was chosen as the output, while the course values were chosen as the input.



### 4.3. Test Result Analysis

The minimal amount of assistance employed in this study is 10%. Figure 9 depicts the results of using CBA to apply the classification approach, whereas Figure 2 depicts the results of using Microsoft SQL Server to apply the association rules technique. Students with an Introduction to Accounting course value of C are more likely to pick the MK specialisation programme, whereas students with an Introduction to Accounting course value of C are more likely to choose the MP specialisation programme. Given that the Introduction to Accounting course is one of the disciplines that forms the foundation of the MK speciality programme, this demonstrates a meaningful link.

Another fact is that students whose Cost Accounting course has a characteristic value of A are more likely to take the MK specialisation programme. Because the Cost Accounting course is one of the courses that forms the basis of the MK specialisation programme, there is a meaningful link between the course value and the selected specialisation programme.

There was no information on the HRM speciality programme gathered from the two tests conducted. This is because the HRM specialised program's mix of items does not fulfil the required minimum support.

### 5. Conclusion

The following are some of the inferences that may be taken from UNAS's data warehouse architecture and use of data mining techniques to academic data:

1. The presence of a data warehouse at UNAS simplifies the reporting procedure. The option to modify reports according to the user's preferences reduces reporting time from one month if done manually or one week if done by establishing a new programme. UNAS will undoubtedly save money in terms of human resources and report preparation costs as a result of this circumstance.
2. Further data analysis was successfully carried out using data mining techniques using the information stored in the UNAS data warehouse to determine the distinctive patterns of students who pursue various specialisation programmes in the UNAS Corporate Management study programme. As a consequence, a pattern of students who enrolled in the MP and MK specialised programmes was discovered. The HRM specialisation programme provided no results because the combination of items produced in the HRM specialisation programme did not match the stipulated minimum support. MO is not included in this study due to the tiny number of data it has (just 33).

### References

- [1]. Paulraj Ponniah. 2001. Data Warehousing Fundamentals. Wiley-Interscience Publication.
- [2]. Jeffrey A. Hoffer, Mary B. Prescott, dan Fred R. McFadden. 2005. Modern Database Management. Seventh Edition. Prentice Hall.
- [3]. Kathleen Polo, Juan Gonzales dan Edwin Rojas. 2006. Quick Tutorial about Mondrian Datamart. International Potato Center (CIP). Research Informatics Unit (RIU).

- [4]. Witten, Ian H dan Eibe Frank, 2005. Data Mining : Practical Machine Learning Tools and Technique. Second Edition. Morgan Kaufmann.
- [5]. <http://www.comp.nus.edu.sg>. Retrieved May 07, 2007. Data Mining II. Classification Based Association (CBA).