Analysis Students' Graduation Eligibility Using Data Warehouse

Edward Chandra
Computer Science Department,
BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
edward.chandra@binus.ac.id

Abba Suganda Girsang
Computer Science Department,
BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
agirsang@binus.edu

Ryo Hadinata
Computer Science Department,
BINUS Graduate Program-Master of Computer Science,
Bina Nusantara University
Jakarta, Indonesia 11480
ryo.hadinata@binus.ac.id

Sani Muhamad Isa
Computer Science Department,
BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
sani.m.isa@binus.ac.id

***Abstract*— Nowadays, most universities have a lot of different ways in considering graduation eligibility for their students. The consideration can use the data which is generated by system on-line analytical processing (OLTP). However the OLTP system has a limitation to report intuitively. The aim of this research is to create the system using data warehouse which able to see the progress of performance student and the course failed in each term. This system is expected able to predict student's graduation eligibility. Research method of this research consists some steps to implement data warehouse and then is followed by generating the report and on-line analytical processing (OLAP) analysis. Evaluation is done by benchmarking current system with the proposed system. The result shows the system can provide data for the graduation eligibility.**

Keywords—data warehouses, OLAP, graduation eligibility, university

I. INTRODUCTION

Binus University is one of private university in Indonesia that excels in quality management and has target to have 90% graduates on time. Binus University has encountered a lot of trouble in predicting and analyzing student's graduate time. The problem comes from unexpected failed grades in some core courses and a big number of data regarding students' grade. Therefore, a system for graduate eligibility and failed courses will be generated to ease the analysis.

Data warehouse, "subject-oriented, integrated, time-varying, non-volatile collection of data is considered suitable for implementing to solve this problem. The data warehouse can handle the high quantity and can be for decision making in an organization or company [1]. Data warehouse also supports on-line analytical processing (OLAP), which is quite different from normal on-line transactional processing (OLTP) applications. OLTP applications usually needs detailed and updated data, which usually accessed, through read and write, hundreds of times in a day. Therefore, OLTP applications needs consistency and recoverability, and reflects operational needs. By using data warehouse, it is possible to integrate business intelligence analysis for the current problem. ETL process is designed for migrating data sources to data warehouse. The construction of the system has three steps, which are: problem identification, fact statement and dimension formulation, and schema formulation; ETL process design and data warehouse implementation; and reporting and OLAP analysis.

II. RELATED WORK

A. Data Warehouse

In contrast to OLTP applications, OLAP can help in decision making support [2]. Users of data warehouse evaluates organization performance by demanding hundreds of thousands of transactions to an answer set [3]. They also demanded historically accurate data to support their decision making, thus, normal OLTP application is not efficient for this case.

Data warehouse includes tools to extract, transform and clean, and loading data from multiple operational databases, data marts, and other external sources. This process is called ETL process (Extract, Transform, Load) and usually done periodically. In loading data to data warehouse, pipelined and partitioned parallelism are used [4]. It means that while loading data to data warehouse in batch transactions, application is still possible to query to the extracted databases.

Data warehouse also has refresh techniques that consists of push updated source data to data warehouses. The research problem of constructing correct updates in data warehouse was conducted [5][6][7][8].

978-1-5386-5821-5/18/$31.00 ©2018 IEEE 3-5 September 2018, Bina Nusantara University, Jakarta, Indonesia
2018 International Conference on Information Management and Technology (ICIMTech)

B. Nine Steps of Database Design Methodology for Data Warehouses

There are nine steps of data warehouses design [3] which includes:

Choosing the process means choosing the business process which is going to be used as the main process that will be analyzed by using OLAP. **Choosing the grain** means deciding what is the smallest grain, or deciding factor, which could be time, number of transactions, or invoices. **In identifying and conforming the dimensions** means avoid of creating two dimensions with the same purposes. Dimensions are also connected with facts in order to help us in designing reports to analyze data. **Choosing the facts** states the type of measurement of grain which will be used, and main facts which is broken down from the business processes stated in step 1. **Storing pre-calculations in the fact table** states the fields which will be calculated in the analysis reports. Fields should be a type of aggregation of data, which is broken down to grains, in order to ease user to query and analyze data. **Rounding out the dimension tables** creates descriptions for attributes which are listed in dimensions that has been created. Descriptions can be something easier for non-IT users to understand. **Choosing the duration of the database** means the duration of ETL process from transactional database to data warehouse will be recurred to integrate the data in data warehouse. **Tracking slowly changing dimensions** has 3 basic steps in handling slowly changing dimensions (SCD) [3], which are overwriting the dimension attribute's prior value, creating a new surrogate key to track the changes, or adding a column to keep the previous value. **Deciding the query priorities and the query nodes** focuses more on physical design of data warehouse, including administration, backup, indexing performance, and security.

C. Star Schema

In data warehouse dimensional modeling, star schema is the simplest form of dimensional model, which organized data into facts and dimensions. Facts are event counted or measured, such as a stocks or login totals. A dimension can contain reference information about the fact, such as date, product, or customer [9]. Stored data in star schema is defined as "denormalized", which means that data has been structured accordingly to reporting purposes. The goal of star schema is to maintain information in fact tables and dimension tables so it can be easier to acquire answers for most business-related queries [10].

III. Proposed Method

By using nine steps design, the data warehouse is proposed to solve the aforementioned problem. They are:

1. Choosing the process. The business process chosen for this case relates two process, student progress and information of course failed. The first process contains number of term left, number of course failed and total credit left. The second process contains the number student who failed in one course.
2. Choosing the grain. The grain for this process is academic year, student, and courses.
3. Identifying and conforming the dimensions. For this case, 4 dimension tables will be used, which two of them are conformed dimensions, which can be seen in Fig. 1.
4. Choosing the facts. This business process is broken down into two big facts, which is student progress fact and failed courses fact, which can be seen in Fig. 1.
5. Storing pre-calculations in the fact table. Pre-calculation fields can be seen in both fact table, which are totalTermLeft, totalCourseFailed and totalCreditLeft for StudentProgressFact, and totalStudentFailed for CourseFailedFact.
6. Rounding out the dimension tables. Different name from current OLTP databases are used to ease the usage of data warehouse which fieldname always represents their fields.
7. Choosing the duration of the database. The duration of the database will be for every 2 weeks after each term ended, because we need the data periodically for every term.
8. Tracking slowly changing dimension. In most of the dimension, we will be using Type 1 SCD handling technique. But in dim_acadadv, CourseCredits will be handled through Type 2 SCD.
9. Deciding query priorities and query nodes. Data warehouse design is completed, which backup tools and design can be seen in Fig 1.

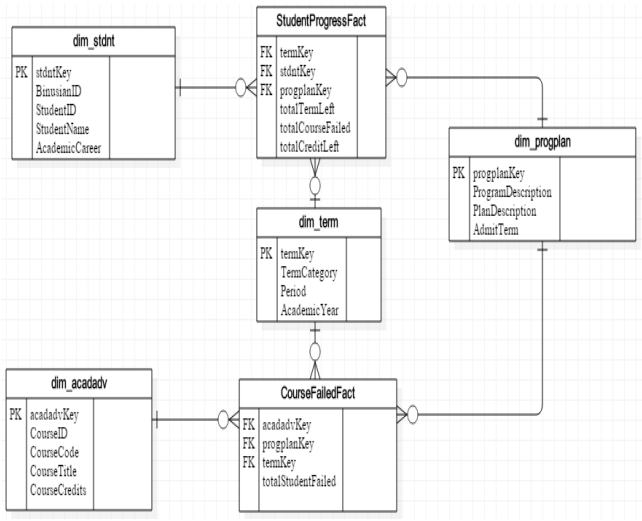


Fig. 1. Star Schema

IV. Result Analysis

A. ETL Process

ETL is a process extraction of data from several sources, cleansing, customization, reformatting, integration, and insertion into a data warehouse. It is an important process since it will keep the consistency data when the data inserted into

dimensional structures. In this paper, the ETL process result will be based on star schema that has been designed and analyzed based on de-normalized table in Fig. 1, which consists of 4 dimensional tables and 2 fact tables.

First fact table, which named StudentProgressFact, contains information about the student's progress that can be tracked down based on the term left, course failed, and credit left so we can know how many students almost dropped out if they don't get attention or advice. The next fact table, which named CourseFailedFact, tells about how many students fail in a certain course, so, university can predict the student progress while taking the current courses based on those data.

B. OLAP Cube

OLAP Cube is a method of storing data in a multidimensional form and used for reporting purposes. Since it give a great performance when do a query. OLAP Cube is another way to represent the star schema. Star schema is known as ROLAP but Cube is known as MOLAP. Fig. 2 represents the star schema that has been explained.

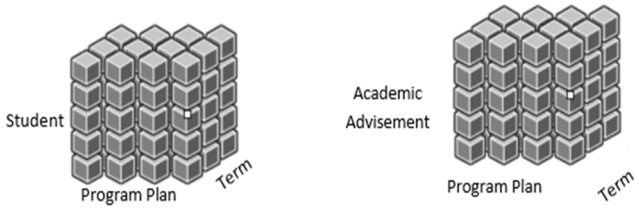
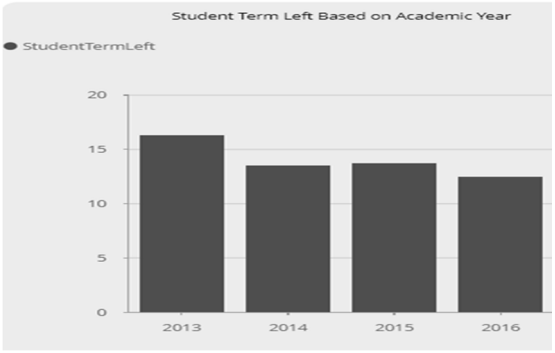


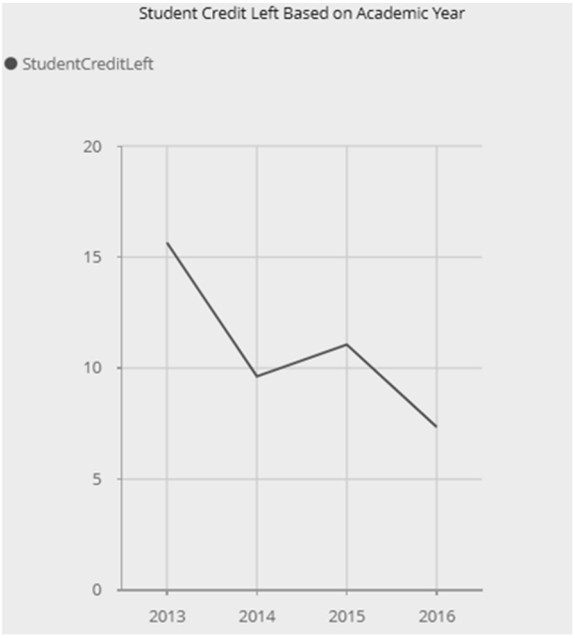
Fig. 2. OLAP Cube

C. Dashboard

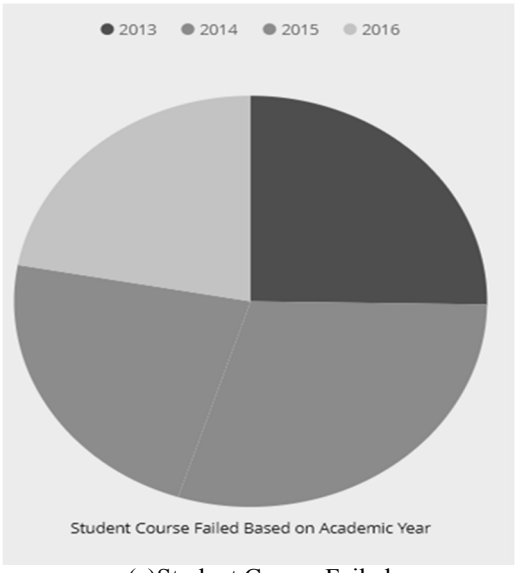
Dashboard contains information for general decision making. It is commonly represented in chart so it can be easier to be understood and evaluated. It contains the main information of certain aspects. For the detail information can be seen on report. Fig. 3 shows dashboard which contains student term left (a), student credit left (b), and student course failed (c) based on academic year.



(a) student term left



(b) Student Course Failed



(c)Student Course Failed

Fig. 3. Dashboard Progress Student

D. BI Reports

BI Report is used to representing the data in more informative ways. It is used for business users for analysis and makes decision to their business. In this paper the report is generated by Pentaho report designer and Pentaho BI server. The first report is created by using Pentaho BI Server. In this process the ROLAP (Star Schema) need to be transformed to MOLAP (Cube) model which can be seen in Fig. 4. Fig. 4 explains information about student term left, student credit left, and student course failed in average based on the academic year and the program.

		Measures		
ProgramPlan	AcademicYear	StudentTermLeft	StudentCreditLeft	StudentCourseF
Computer Science	2013	15.933	14.467	
	2014	12.837	11.367	
	2015	13.032	10.619	
	2016	11.771	6.623	
	2017			
Industrial Engineering	2013			
	2014			
	2015	18	30	
	2016			
	2017			
Information Systems Management	2013	16.625	16.75	
	2014	14.216	7.961	
	2015	14.328	11.213	
	2016	13.33	8.255	
	2017			

Fig. 4. Student Information Report

Fig. 5 shows "failed course report" in one term which tells how many students failed on certain course.

Course Code	Course Title	Failed
COMP8015	Multimedia Indexing and Retrieval	2
CPEN8003	Network Governance	1
ISYS8005	Ethical Issues in Electronic Information System	2
ISYS8010	Information System Strategic Planning	3
ISYS8017	Technology Management and Valuation	3
M7143	Corporate Information Technology Strategy	1
M7153	Technology Management and Valuation	2
M7166	Thesis	1
M7166	Thesis	37
M7233	Information System Strategic Planning	1
M7240	Pre Thesis	19
RSCH8003	Thesis	12
RSCH8003	Thesis	1
RSCH8003	Thesis	11
RSCH8003	Thesis	26
RSCH8004	Pre Thesis	25
RSCH8004	Pre Thesis	12
RSCH8004	Pre Thesis	3
RSCH8007	Thesis	9
RSCH8007	Thesis	43
RSCH8009	Pre Thesis	12
RSCH8009	Pre Thesis	25
RSCH8011	Thesis	5
T7166	Thesis	24
T7166	Thesis	4
T7193	IT Portfolio Management	1
T7270	Pre Thesis	15
		300

Fig. 5. Failed Courses Report

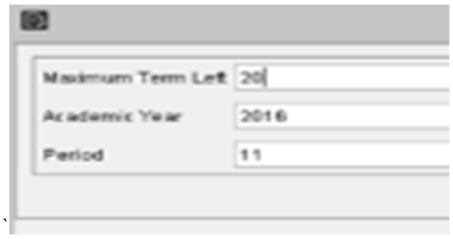


Fig. 6. Parameterized Report

Fig 6 shows the parameterized report which can be processed based on maximum term left, academic year, and period criteria. It will show the student information who met that criteria..

V. CONCLUSIONS

The data warehouse generated can be used for supporting the decision making, through good reporting ability. This research focuses two process, the student progress and the course failed. It is expected helping decision makers to take decision for evaluating the student to optimize target of the graduation student.

The coverage process only involves two process. In future, it can include the other process which influences for the duration of study before graduation for the more complete data warehouse system.

REFERENCES

[1] W. H. Inmon, Building the Data Warehouse. John Wiley, 1992.

[2] E. F. Codd, S. B. Codd, C.T. Salley. "Providing OLAP (On-Line Analytical Processing) to User Analyst: An IT Mandate." Accessed from Arbor Software's website http://www.arborsoft.com/OLAP.html

[3] R. Kimball. The Data Warehouse Toolkit. John Wiley, 1996.

[4] T. Barclay, R. Barnes, J. Gray, P. Sundaresan. "Loading Databases using Dataflow Parallelism," SIGMOD Record, Volume 23, No. 4, 1994.

[5] J. A. Blakeley, N. Coburn, P. Larson. "Updating Derived Relations: Detecting Irrelevant and Autonomously Computable Updates," ACM TODS, Volume 4, No. 3, 1989.

[6] A. Gupta, I. S. Munick. "Maintenance of Materialized Views: Problems, Techniques, and Applications," Data Engineering Bulletin, Volume 18, No. 2, 1995.

[7] Y. Zhuge, H. Garcia-Mollina, J. Hammer, J. Widom. "View Maintenance in a Warehousing Environment," Proc. of SIGMOD Conference, 1995.

[8] N. Roussopoulos, et.al. "The Maryland ADMS Project: Views R Us." Data Engineering Bulletin, Volume 18, No 2, 1995.

[9] M. K. Sandhu, A. Kaur, R. Kaur. "Data Warehouses Schema," in International Journal of Innovative Research in Advanced Engineering (IJIRAE), Issue 4, Volume 2, 2016, pp.47-51.

[10] N. Rausch. "Stars and Models: How to Build and Maintain Star Schemas Using SAS® Data Integration Server in SAS®9," SAS Institute, Paper 096-31, 2012.