

University Accreditation using Data Warehouse

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 J. Phys.: Conf. Ser. 801 012030

(<http://iopscience.iop.org/1742-6596/801/1/012030>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 168.151.158.12

This content was downloaded on 24/03/2017 at 13:45

Please note that [terms and conditions apply](#).

You may also be interested in:

[Study on resources and environmental data integration towards data warehouse construction covering trans-boundary area of China, Russia and Mongolia](#)

J Wang, J Song, M Gao et al.

[A Messaging Infrastructure for WLCG](#)

James Casey, Lionel Cons, Wojciech Lapka et al.

[Measurement Standards: National measurement accreditation service launched](#)

[The design of the automated control system for warehouse equipment under radio-electronic manufacturing](#)

D V Kapulin, I V Chemidov and M A Kazantsev

[Sequential data access with Oracle and Hadoop: a performance comparison](#)

Zbigniew Baranowski, Luca Canali and Eric Grancher

[HOKLAS—Hong Kong Laboratory Accreditation Scheme ILAC 94](#)

[Summary report: Comparison of humidity generators with a dew-point hygrometer](#)

M Heinonen

[Final report of EURAMET 1197: Supplementary bilateral comparison of hydraulic gauge pressure standards up to 50 MPa](#)

Yasin Durgut, Nenad Petrovski and Vanco Kacarski

University Accreditation using Data Warehouse

A S Sinaga¹, A S Girsang²

¹Department in Informatics Engineering, STMIK Pelita Nusantara, Jl. Iskandar Muda No.1, Medan.

²Master in Computer Science, Bina Nusantara University, Jl. Kebon Jeruk Raya No. 27 Jakarta, Email : agirsang@binus.edu

Email : agirsang@binus.edu

Abstract. The accreditation aims assuring the quality the quality of the institution education. The institution needs the comprehensive documents for giving the information accurately before reviewed by assessor. Therefore, academic documents should be stored effectively to ease fulfilling the requirement of accreditation. However, the data are generally derived from various sources, various types, not structured and dispersed. This paper proposes designing a data warehouse to integrate all various data to prepare a good academic document for accreditation in a university. The data warehouse is built using nine steps that was introduced by Kimball. This method is applied to produce a data warehouse based on the accreditation assessment focusing in academic part. The data warehouse shows that it can analyse the data to prepare the accreditation assessment documents.

1. Introduction

The value accreditation of a program study is one of important parameter to classify the quality of university in Indonesia especially private universities. The accreditation is measured based on the evidence assessment related to the specific standards and the expert(s)'s judgement or consideration. In Indonesia, the judgment for accreditation higher education, including university is did by accreditation council national, called BAN-PT. The university should prepare the comprehensive data in a template which is provided by BAN-PT. This data consists academic history, profile lecturers and staff, vision mission university, the facility and so forth. One of the important form is called "borang" which consists 7 forms. In STMIK Pelita Nusantara university, the source data is spread in both systems, manual and information system. Some of them are in document file (.docx) , spread sheet file (.xls) and database system. This fact causes gathering the comprehensive data for "borang" to be a serious problem.

Data warehouse is one of tool to gather all the company data in one place in order to get a better view of a business process and improve the performance of the work/organization. Application of data warehouse are implemented to solve some problems [1-7]. Bassil [1] presented a design model for building data warehouse for an university. The proposed model is based on four stages of data migration: data extraction, data cleansing, data transforming, and data indexing - loading [1]. Sappagh et al proposed a model which is built upon the enhancement of the models in the previous models to support some missing mapping features [2]. Data warehouse and business intelligence have been used in the recent years in many areas for analyzing different phenomena. Rabuzin showed how data



warehouse has been used in educational field for analyzing university exams[3]. Kurniawan and Halim proposed a model that can be applied in data warehouse and data mining techniques to predict student performance (academic) in schools [4]. The data mining and data warehouse model can help the low achiever students, evaluate the course or module suitability, and tailor the interventions to increase student academic performance in schools [4]. Jindal and Taneja started developing a data warehouse identifying and gathering requirements, designing in the dimensional model followed by testing and maintenance [5]. Luan and Jiang [6] present a system architecture of library books based on data warehouse which uses the accumulated data of college library as data resources, simulate simple data warehouse and use OLAP technology. Namusonge [7] found out the effect of information technology on warehousing management. The researcher used descriptive research design taking Jomo Kenyatta University of Agriculture and Technology as a case for this study. This study recommended continuing investment and training in information technology and adoption of better information sharing tools.

2. Related Works

Data warehouse is a collection of large data which is used by the management for making strategic decisions. The data in a data warehouse is gathered from heterogeneous sources and then populated and queried for carrying out the analysis. The data warehouse design must support the queries for which it is being used for [8]. Characteristic of data warehouse can be defined as subject-oriented, integrated, time-varying, and non-volatile collections of data. Nowadays, data warehouse becomes an important strategy to integrate heterogeneous information sources in organizations, and to enable On-Line Analytic Processing (OLAP) [9]. Data warehouse can be built using a number of architectures [10]. It goes beyond the traditional focus on data mining problems to introduce advanced data types such as text, time series, discrete sequences, spatial data, graph data, and social networks [11]. As data warehouse stores huge amount of data, the security of this huge information is crucial for the sustainability and reliability of data warehouse [12]. The quality of correctness and completeness of an information depends on how well the data model is constructed. One way to get a good data model is by utilizing patterns [13].

The integration of data sources is achieved by using ETL (Extract, Transform, and Load) processes. The ETL process is to identify the main characteristics, notation, and activities of company. A mapping study provides a systematic and objective procedure for identifying the nature and extent of the available research by means of research questions [14]. The ETL extraction step is responsible for extracting data from the source systems. The transformation step tends to make some cleaning and conforming on the incoming data to gain accurate data which is correct, complete, consistent, and unambiguous. This process includes data cleaning, transformation, and integration. It defines the granularity of fact tables, the dimension tables, data warehouse schema (star or snowflake), derived facts, slowly changing dimensions, and fact less/ fact tables. All transformation rules and the resulting schemas are described in the metadata repository. Loading data to the target multidimensional structure is the final ETL step. In this step, extracted and transformed data is written into the dimensional structures actually accessed by the end users and application systems. Loading step includes both loading dimension tables and loading fact tables.

The source data consists event streams that indicate changes in the business process state i.e. progression of the process. The idea is to automate ETL process using scripting technologies [15]. As Business Intelligence evolves from off-line strategic decision making to on-line operational decision making, the design of the back-end Extract-Transform-Load (ETL) processes is becoming even more complex. Many challenges arise in this new context like their optimization and modelling. Wilkinson et al investigated the disconnection between the IT-level view of the enterprise presented by ETL and the business view of the enterprise required by managers and analysts [16].

3. Proposed Method

3.1 Preparing Data Source

There are seven forms (standard1, standard2,...,Standard7) of accreditation council national for university (BAN-PT). However, this research focuses only form standard3 and standard4. It is because most of the rest form standards (1,2,5,6,7) are from the data text. For example standard1 is vision mission, standard2 is the leadership. The standard3 is about students and alumni, while standard4 is about human resource. Table 1 shows the standard3 and standard4 of accreditation council (BAN-PT). Not all of items are treated for data warehouse in this study. This research focuses on items a, c, d, e, f, l and m on standard3, and items d, h, k on standard4. To prepare for data staging, three steps of ETL, extraction-transformation-load, are conducted [18] as described on Section 2.

Table 1 Form Standard 3 and 4 of BAN-PT

| Standard Form | Value |
|----------------------------------|---|
| Standard3 Students and Alumni | a) The ratio candidate students compared to the capacity of the new students. b) The ratio candidate new students compared to the new graduated student. c) The ratio new students transfer compared to regular students. d) Average GPA of students in the last five years. e) Percentage of timely graduation. f) The percentage of students drop out or resign. g) The percentage of students who have access to student services. h) The quality of service to students. i) Efforts tracking and data recording the alumni. j) Use of tracking results for improvement. k) Opinion of industry about quality of the alumni. l) The average waiting time alumni get jobs. m) The percentage of the similarity job with the field of study for alumni. n) Participation alumni participation of academic and non-academic study programs. |
| Standard4 Human Resource | a) Written catalogue for the selection system. b) Monitoring system. c) Evidence of monitoring and evaluation of lecturers. d) The number of faculty member lecturers who have at least master degree. e) The number of faculty member lecturers who have been associate professor and professor, f) The number of faculty member lecturers who have the certified educators. g) The ratio of the number of students compared to lecturers who are suitable areas of expertise. h) The average of the lectures load (credits) per semester. i) The similarity the lecture's expertise with the course being taught. j) The attendance of the lecturer. k) The percentage of part-time lecturers compared to the entire lecturers. l) The dissimilarity the part time lecture's expertise with course being taught. m) The attendance part-time lecturers in teaching. n) The activities of the expert. o) Improvement by sending the lecturers study which is similar with field program study. p) The activities of faculty member lectures whose expertise appropriate study program. q) The performance in obtaining grants / awards to institutions. r) Reputation and networking lecturers in academia. s) Librarian with qualifications. t) Laboratory assistant, technician, operator and programmer. u) Administration staff. |

3.2 Design of Data Warehouse

The design data warehouse is implemented by using Nine-Step Methodology Kimball, consisting of 9 design stages [19]: Choosing the process, Choosing the Grains, Identifying and Conforming the Dimension, Choosing the Fact, Storing Pre-Calculation in the Fact Table, Rounding

Table, The duration of the Election, Slowly Changing Dimension, The determination of the priority of the query model. There is some star schema designed for this case.

- Registration for new students, can be shown on Figure 1. This star schema can solve the accreditation standard 3 items (a) and (c)
- Studying, can be shown on Figure 2. This star schema can solve the accreditation standard 3 items (d), (e), (f), (l) and (m)
- Lecturer, can be shown on Figure 3. This star schema can solve the accreditation standard 4 items (d), (h), and (k)

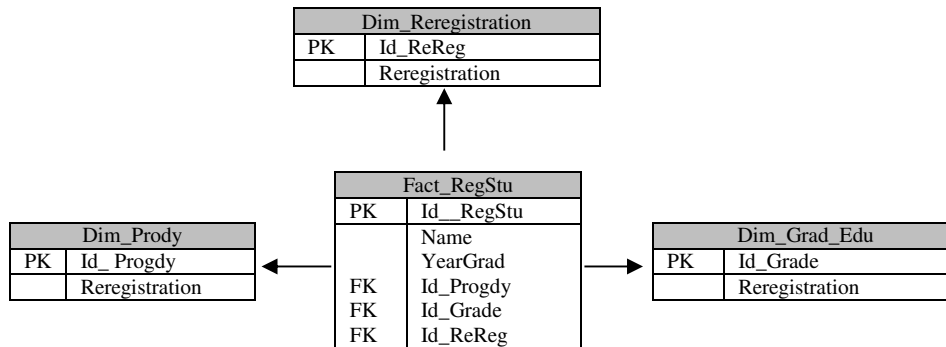


Figure 1. Star schema registration for new students

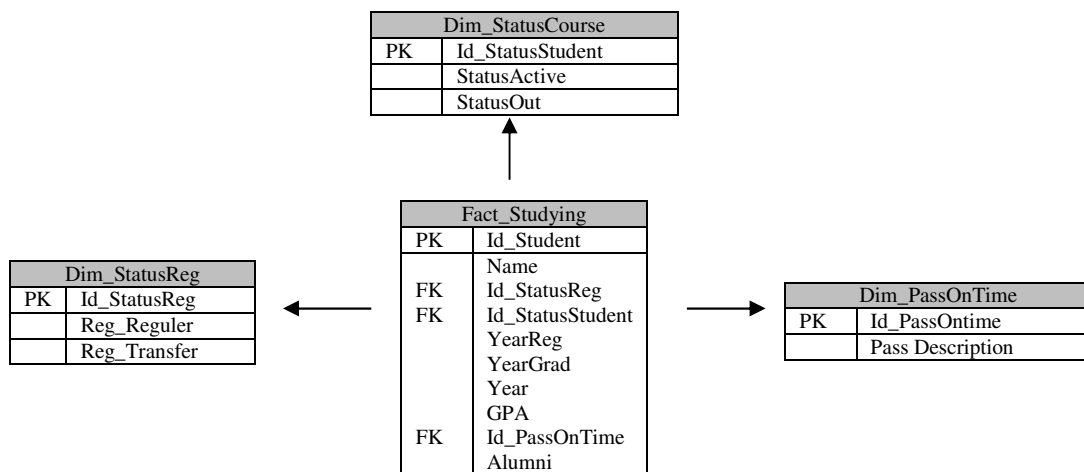


Figure 2 Star schema for Studying

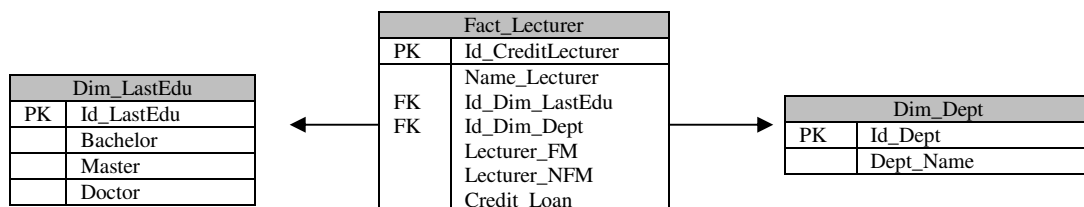


Figure 3 Star Schema Lecturer

To ease building data warehouse, the process ETL of all fact table and its dimension have to be done.

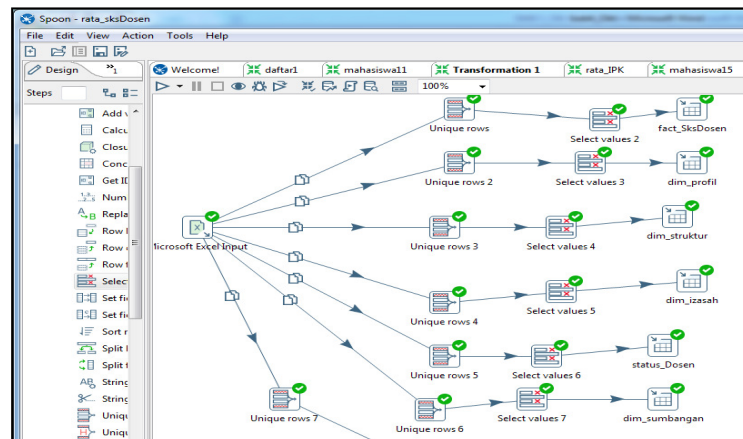
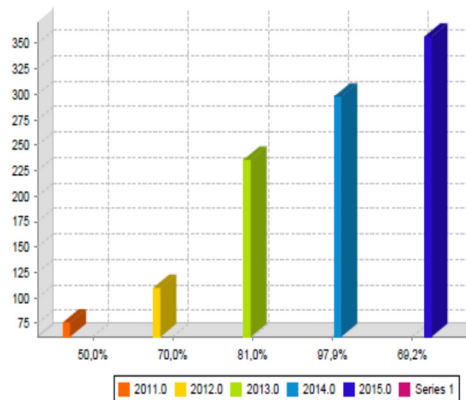


Figure 4 Process of ETL of Fact Registration Student

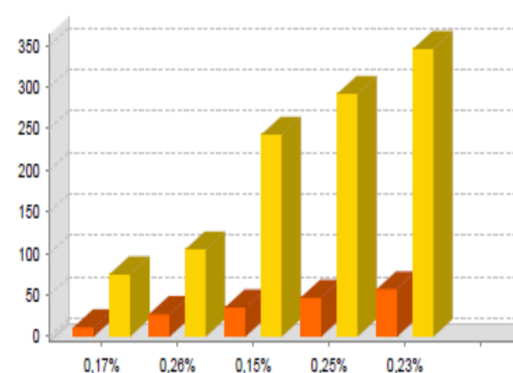
In this study, the data extraction process is conducted by using a tool called *Pentaho Data Integration / Kettle*. *Kettle* is a part of *Pentaho* application. The data source are obtained from operational data. Figure 4 shows one of process ETL (Fact Registration Student).

4. Analysis Result

This section describes the data which is provided by data warehouse to fulfil the form of BAN-PT. This data is taken 2011-2015. Some of the data are shown as follows. Figure 5 shows Percentage of capacity for new students. This chart can fulfil the requirement standard3 (a). Each academic year conducting new admissions, the capacity of capacity of 2011 = 150 students, the year 2012 = 150 students, 2013 = 300 students, 2014 = 300 students, the year 2011 = 500 students. The ratio of the total number of new students with a capacity that is in 2011 = 50.0%, in 2012 = 70.0%, 2013 = 81.0%, 2014 = 97.9%, 2015 = 60.2%.



(a)



(b)

Figure. 5(a) Capacity percentage candidate students and (b) Ratio percentage transfer students compared to regular students

Each year the university organizes new admissions for some transfer students. Comparison of registered students regularly with transfers students can be seen on Figure 5b. This chart can answer the requirement standard 3(c). Figure 6 (a) shows the graph of the average GPA of students in the last five years (Standard3 (d)). GPA (grade point average) is obtained from the total value of the number divided by the total number of credits of all subjects.

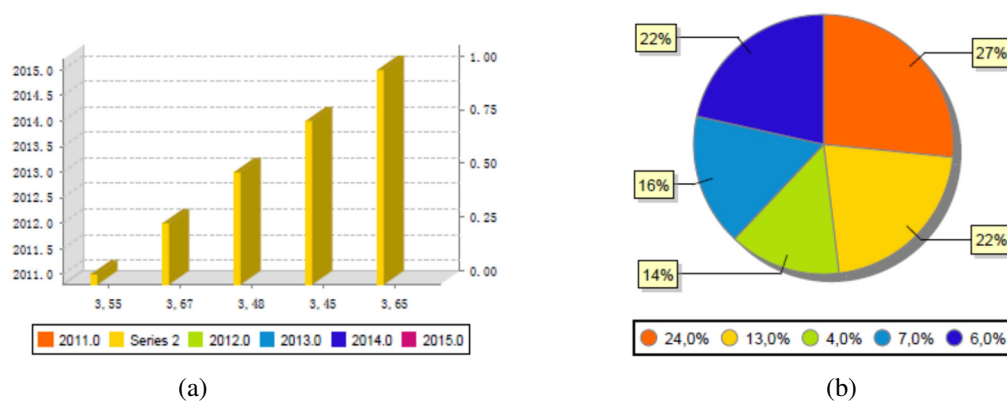


Figure 6 (a) Average GPA, (b) Percentage of drop out students

Figure 6 (b) shows the percentage of students drop out or resign. This chart is useful for Standard3 (f)

5. Conclusion

In this paper, we design a data warehouse to produce reports and charts as the representation activity in university to fulfil the requirement of accreditation council-BAN PT. The nine step methodology is used to develop the data warehouse. This data is taken from the data operational of management system on STMIK Pelita Nusantara, Medan. By using the ETL process, the data OLTP can be integrated into data warehouse.

This data warehouse is designed only for fulfil the standard3 and standard4. In future work, the data warehouse can be extended for the other standards. Also, the data which is created by data warehouse can be used for mining research such as prediction, association rule and so forth.

REFERENCES

- [1] Bassil Y, "A Data Warehouse Design for A Typical University Information System," J. Comput. Sci. Res., vol. 1, no. 6, pp. 12–17, 2012.
- [2] El-Sappagh S H A, Hendawi A M A, and El Bastawissy A H, "A proposed model for data warehouse ETL processes," J. King Saud Univ. - Comput. Inf. Sci., vol. 23, no. 2, pp. 91–104, 2011.
- [3] K. Rabuzin K, "The Use of a Data Warehouse For Analyzing Exams at the University," 3rd Int. Conf. Data Min. Intell. Inf. Technol. Appl., pp. 145–148, 2011.
- [4] Kurniawan Y, and Halim E, "Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits)," in Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2013, 2013, pp. 98–103.
- [5] Jindal R, and Taneja S, "Comparative study of data warehouse design approaches: a survey," Int. J. Database Manag. Syst., vol. 4, no. 1, pp. 33–45, 2012.
- [6] Luan X M and Jiang H, "Design of Books Analysis System in University Library Based on Data Warehouse," Appl. Mech. Mater., vol. 513–517, pp. 2121–2124, 2014.
- [7] Namusonge K G K, "Role of Information Technology on Warehouse Management in Kenya: A Case Study of Jomo Kenyatta University of Agriculture and Technology," Int. J. Acad. Res. Bus. Soc. Sci., vol. 4, no. 11, pp. 2222–6990, 2014.
- [8] Gupta D S L, Pahwa D P, and Mathur M S, "Classification Of Data Warehouse Testing Approaches," International Journal Of Computers & Technology, vol. 3, no. 3a, pp. 381–386, 2012.
- [9] Vassiliadis P, "Data Warehouse Modeling and Quality Issues," Natl. Tech., no. June, pp. 1–129, 2000.
- [10] Hajmoosaei A, Kashfi M, and Kailasam P, "Comparison plan for data warehouse system architectures," in 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA), 2011, pp. 290–293.
- [11] Aggarwal C C, Data Mining: The Textbook. 2015.
- [12] Gosain A and Arora A, "Security issues in data warehouse: A systematic review," in Procedia Computer Science, 2015, vol. 48, no. C, pp. 149–157.
- [13] Viqarunnisa P, H Laksmiwati, and Azizah F N, "Generic data model pattern for data warehouse," in

- Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI, 2011.
- [14] Muñoz L, Mazón J N, and J. Trujillo, "ETL Process Modeling Conceptual for Data Warehouses : A Systematic Mapping Study," *America (NY).*, vol. 9, no. 3, pp. 358–363, 2011.
 - [15] Radhakrishna V, Sravankiran V, and Ravikiran K, "Automating ETL process with scripting technology," in *3rd Nirma University International Conference on Engineering, NUiCONE 2012*, 2012.
 - [16] Wilkinson K, Simitsis A, Castellanos M, and Dayal U, "for ETL Design," *Business*, vol. 6412, pp. 15–30, 2010.
 - [17] D. Warehousing and B. Intelligence, Kimball Group. .
 - [18] George S, "Inmon vs. Kimball: Which approach is suitable for your data warehouse?," *Data Warehous.*, pp. 1–12, 2012.
 - [19] Baker P, and Canessa M, "Warehouse design: A structured approach," *Eur. J. Oper. Res.*, vol. 193, no. 2, pp. 425–436, 2009.
 - [20] S. H. A. El-sappagh, A. Hamed, E. Bastawissy, and A. M. Ahmed, "Original Article A proposed model for data warehouse ETL processes," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 23, no. 2, pp. 91–104, 2011.
 - [21] L. Dokumen, "With Pentaho Community Edition."
 - [22] "Introducing the Pentaho BI Suite 3.5 Community Edition."
 - [23] Sandoval L J, "Design of business intelligence applications using big data technology," in *2015 IEEE Thirty Fifth Central American and Panama Convention (Concapan XXXV)*, 2015, no. Concapan Xxxv, pp. 1–6.