

DATA WAREHOUSE DESIGN IN ACADEMIC ENVIRONMENT

Munawar

Fakultas Ilmu Komputer Universitas
Jalan Arjuna Utara no. 9, Kebun Jeruk, Jakarta 11510
munawar@esaunggul.ac.id

Abstract

The popularity of Data Warehouses (DW) for data analysis has grown tremendously, as conventional transaction processing systems have matured, becoming faster and more stable. Conventional systems in universities working in the same area, store daily activities and display or report these events on a regular basis. Therefore, the establishment of the DWs will help the top management in the development of strategies and take appropriate decisions in these universities. This paper addresses issues related to designing a DW for private university in Jakarta. The purpose of the project is to provide a process model that will enhance decision making capabilities in a private university in order to facilitate and improve academic activities.

Keywords: *requirements analysis, conceptual design, data warehouse, academic*

Introduction

The strategic use of information technology has become a fundamental issue for every business as information technology can enable the achievement of competitive and strategic advantage for the enterprise. Data warehousing technology has made a huge impact in the world of business, where with its help data turned to information that early adopters could leverage for enormous competitive advantages. A Data Warehouse (DW) is a collection of data from heterogeneous sources, integrated into a common repository and extended by summary information for the purpose of analysis to support decision making [29]

Higher education institutions, for the most part, are collecting more data than ever before. Most of these data are used to satisfy credentialing or reporting requirements rather than to address strategic questions, and much of the data collected are not used at all. In parallel with the business world, the pace of change is rapidly increasing. More than ever before, organizations need timely, accurate, and relevant information on which to base decisions, not only for the long term and for next year's planning, but on a daily basis. That needs a DW technology which is more appropriate and providing many tools for this purpose.

Higher educations are one of the important parts of our society and playing a vital role for growth and development of any nation. However, there is very little research on how data warehousing is used in higher education, especially as it relates to providing improved support for decision making. This is probably due to the fact that this area is not as commercially attractive as accounting, banking or other money related businesses. As a result, analysis needs for decision making is often ad hoc, and tends to be more resource and attention intensive than accounting systems. DW may be taken as good choice for maintaining the records of past history for the purpose of analysis.

One of the major barriers to develop DW in higher education is cost. Many institutions view DW as an expensive endeavor rather than as an investment. DWs has become a hot topic in higher education as more institutions show how DWs can be used to maximize strategic outcome. The deployment of a DW is, for most colleges and universities, a potentially significant cost-saving move to make. Depending on the kind of data an institution wishes to more readily manipulate than current conditions and reporting systems allow, a DW can better guide administrative decision-making, steer financial and academic planning, and more completely inform business units in a way that can fundamentally streamline institutional operations. It needs a lot of historical data to generate which the current OLTP system cannot support.

This study is relevant because it fills a gap in the knowledge about fact-based decision making in the environment of higher education. It is timely and of current interest, as demonstrated by increasing trend in information discovery in academic environment [5]. This paper shows the application of DW under the background of higher educational organization.

Data Warehouse and Higher Education

The DW provides effective business decision support [9] and business analyses [8] by integrating data from multiple, incompatible systems into a consolidated database. DW technology is inherently complex [10], it requires huge capital spending [2] and consumes a lot of development time. The adoption of DW technology is not a simple activity of purchasing the required hardware and software, but rather a complex process to establish a sophisticated and integrated information system [2, 11]. Building a DW consists of a complex process involving data sourcing, data extraction and conversion [12-14].

Research into how DW and the impact of high-quality data within higher education systems has been relatively sparse. The literature that examines higher education DWs is limited to the professional papers that appear on vendor web-sites and the publications of organizations. An overview of OLAP within higher education environment is explained in [2]. The modeling of students record data of the Georgetown University DW is addressed in [19, 20], while DW design and modeling experience of the University of Florida is presented by [21]. In addition, author of [6] has examined the challenges that universities confront in the management of information necessary for decision making. Issues related to modeling and implementing a DW for the higher education in Croatia addressed by [22]. Palmer [23] in his study focused on the development of a method to improve the quality of the data contained in the academic students' records component of a DW. Heise [24] has examined the use of Data warehousing in higher education and its role in decision making.

In an information-intensive climate, success is often determined by what and how data are used. Given this, the academic DW is one of a few tools available that can support the importance and increasing volume of data that decision makers require today and in the future. The primary beneficiaries of the academic DW are the key decision makers of the institution. Administrators will benefit from improved access to and analytical capabilities of historical and environmental data. Strategic planning teams, along with operational decision makers, will be more informed and able to guide the institution more effectively into the future. The majority of the faculty, staff, and students will benefit from the result of improved administrative decision making and planning [1]

Conceptual Design of the Academic Data Warehouse

We adopted integrated requirement analysis for designing data warehouse (IRADAH, for short) [3,4], for designing the academic DW. DW is a system composed of several sub systems and processes between them. A framework provides assistance to DW designers by linking the main components of DW architecture to a model of DQ that will be integrated. At the requirements analysis, the information needed to be maintained in the DW is discovered. The next phase, that is the conceptual design, can be said as abstraction of the users' request to some information structures, which act as the bridge connecting the real world and the machine world.

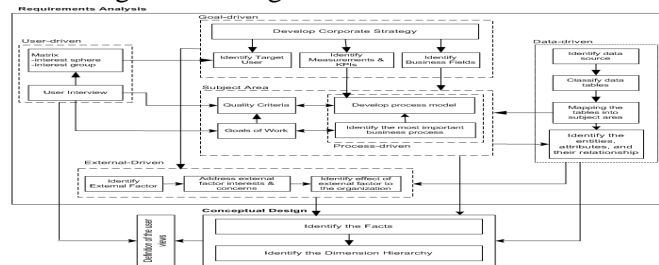


Figure 1
IRADAH Method [4,5]

Requirements Analysis

The requirements analysis for DWs is different than those which apply to other types of information systems. Requirements analysis for DW identifies which information is relevant to the decisional process by either considering the user needs or the actual availability of data in the operational sources.

Requirements can be classified into functional & non-functional requirements. In context of a DW, functional requirements specify what data is to be stored in it, while non-functional requirements specify how this information should be provided to facilitate reporting and analysis in a correct manner. There are several approaches how to determine the requirements for the development of DW.

- ✓ The data-driven (also called supply-driven) approach [25; 26] is a bottom-up technique that based on exploration of the data sources in order to identify all the available data. Specifically, data-driven approach tries to construct DW based only on operational system database schemas overlooking business goals and user needs. Organization needs are not identified, or are identified only partly. Based on existing documentation and database schema that already exist in database, existing class diagram for academic system can be seen in the following figure.

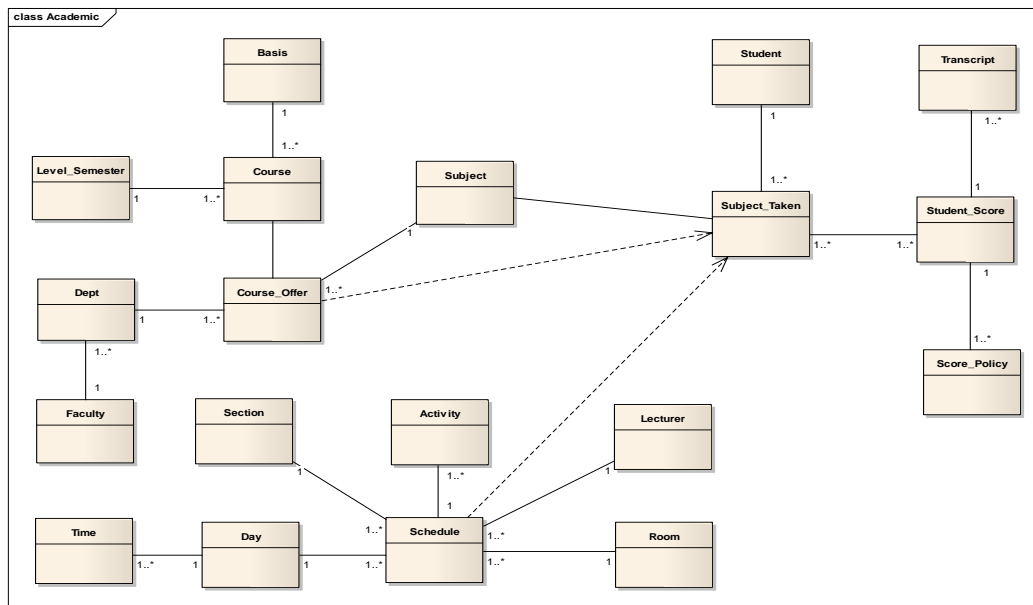


Figure 2
Class Diagram for academic

- ✓ The process-driven approach identifies the most important business process that requires measurement and control [27;28] and then aligns them with corporate strategies. A DW is designed according to a set of relevant business subjects [3] and each of these subjects centers around a business process.

Gathering the information related to processes from different sources, monitoring these processes, and aligning them with corporate strategies and high-level goals is a major issue for decision makers. The greatest potential benefit of the DW occurs when it is used to redesign business process and to support strategic business objectives [28].

The Semantic Object Model (SOM) is a comprehensive and integrated methodology for business engineering [27]. It supports sound modeling of business systems and can be used for analysis and design. The following is task-event schema of the academic business process in a private university.

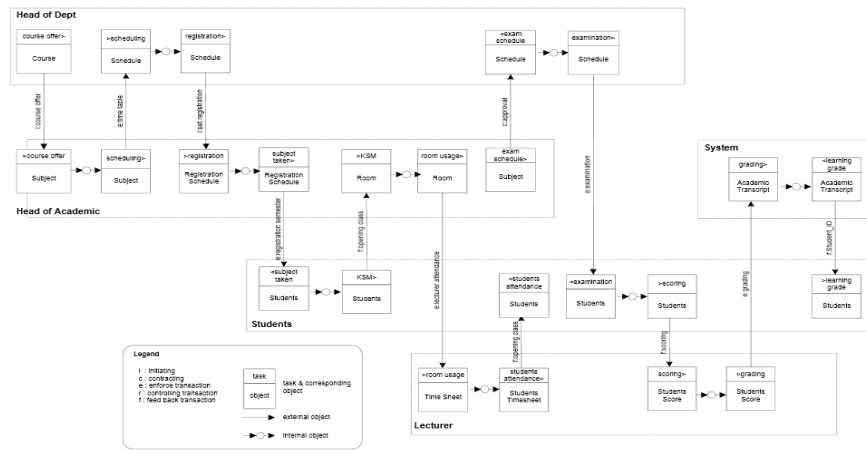


Figure 3

Task-event schema for the academic process in a private university

✓ The goal-driven and the user-driven approach

The goal-driven approach is typically top-down. It emphasizes on the need to align DW with corporate strategy and business objectives [7] by interviewing the top-management. Different visions are then analyzed and merged in order to obtain a consistent picture and finally translated into quantifiable KPIs.

The user-driven approach is a bottom-up technique that stresses involvement of end users in DW [18;17] to determine the information requirements of different business users. Their points of view are then integrated and made consistent in order to obtain a unique set of multidimensional schemata.

A scenario-based design method is very useful to determine what functionalities are required for the system [8]. Scenarios can be used to systematically derive all the requirements. Varied situations can be easily exemplified using scenarios. Also, implementation of the DW becomes easier due to the clear structure of scenarios.

Based on interviews using Kimball template, and the similar answers result then be grouped into matrix interest sphere, the scenario of academic can be seen in the following figure.

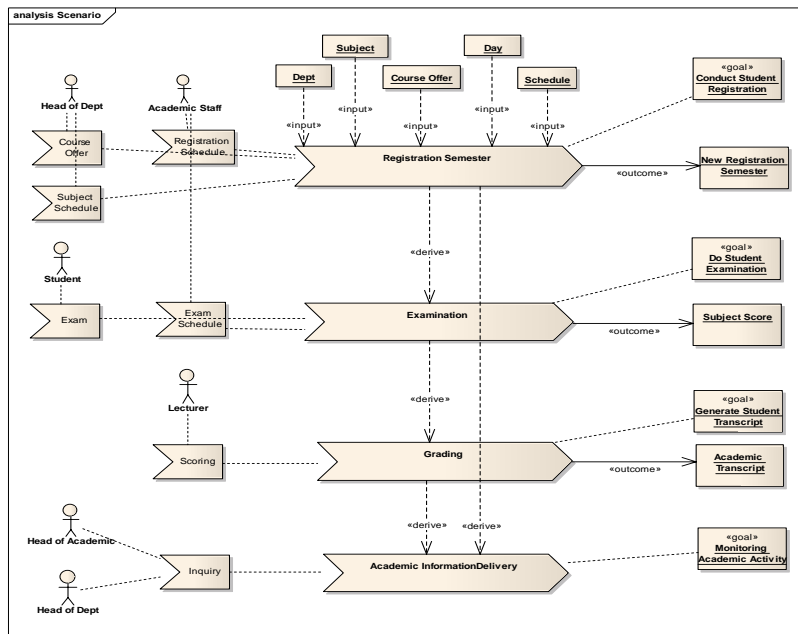


Figure 4
Academic Scenario

- ✓ The external-driven approach is the need for handling the pressure to comply with governmental regulations (such as bank central regulation for banking industry) and others external pressure (such as regulation for public listed company) which require a real-time disclosure about business operations [15]. This approach is typically top-down and the main reason behind the current growth rate of such DW initiatives.

In case of academic, external-driven mainly focus on identifying the need for external reporting especially for ministry of higher education. Forlap (reporting format for higher education) has to be submitted to the ministry of higher education every semester. Inclusion in this report are students profile, lecturers profile, academic schedule and grading.

From Requirements Analysis to Conceptual Design

Conceptual design is widely recognize to be necessary foundation for building a database that fully satisfies user requirements. In particular, from the designer point of view the availability of a conceptual model provides a higher level of abstraction in describing the warehousing process and its architecture in all its aspects.

Conceptual model of a DW consist of a set of fact schemes. The basic components of fact schemes are facts, dimensions and hierarchies. A fact is a focus of interest for the enterprise and usually measures of business processes; a dimension determines the granularity adopted for representing facts; a hierarchy determines how fact instances may be aggregated and selected significantly for the decision-making process.

Facts are usually measures of business processes and dimensions represent the context for analyzing these measures. Dimensions offer the key to understanding fact measures by allowing the user to view data through different viewpoints. Dimensional Fact Model (DFM) is an appropriate tool for conceptual design in a DW [16]. Below is DFM for academic system.

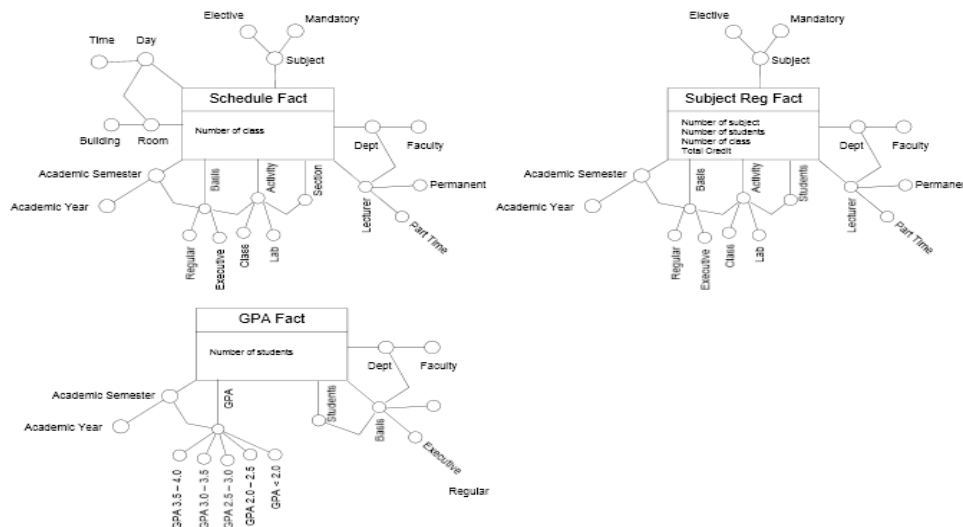


Figure 5
Dimensional Fact Model for Academic Systems

Conclusion

In this paper, conceptual design for academic activities in private university is presented. This paper addresses issues related to designing a DW for private university in Jakarta. The purpose of the project is to provide a conceptual model that will enhance decision making capabilities in a private university in order to facilitate and improve academic activities. Our future works will be devoted to developing the methodology for integrating quality into the logical design and physical design.

References

- Artz, J, "Data driven vs. metric driven data warehouse design", In Encyclopedia of Data Warehousing and Mining, pp. 223 – 227, Idea Group, 2005
- Boehnlein M., Vom Ende U, "A Business Process Oriented Development of Data Warehouse Structures", In Proceedings of Data Warehousing 2000, Physica Verlag, 2000
- C. dell'Aquila, F. Di Tria, E. Lefons, and F. Tangorra, "Business Intelligence Application for University Decision Makers", WSEAS Transactions on Computers. ISSN 1109-2750, Issue 7, Volume 7, July, 2008
- D. J. Berndt and R. K, "Satterfield, "Customer and household matching: Resolving entity identity in data warehouses", The International Society for Optical Engineering, vol. 4057, pp. 173-180, 2000
- D. L. Heise, " Data warehousing and decision making in higher education in the United States", Andrews University, 2006
- Frolick, M., & Ariyachandra, T, "Critical Success Factors in Business Performance Management- Striving for Success", Information Systems Management, 25, 113-120, 2006
- Giorgini, P., Rizzi, S., Garzetti, M.: GRAnD, "A goal-oriented approach to requirement analysis in data warehouses", Decision Support Systems 45(1), 4–21, 2008
- H. J. Watson, C. Fuller, and T. Ariyachandra, "Data warehouse governance: best practices at BlueCross and Blue Shield of North Carolina", Decision Support Systems, vol. 38, pp. 435-450, 2004
- H. Palmer, "A Data Warehouse Methodology and Model for Student Data in Higher Education", NovaSoutheastern University, Graduate School of Computer and Information Sciences, 2006
- J. Guan, W. Nunez, and J. F. Welsh, "Institutional strategy and information support: the role of data warehousing in higher education", Campus-Wide Information Systems, vol. 19, pp. 168-174, 2002
- Kaldeich, C., & Oliveira, J, "Data warehouse methodology: A process driven approach", In Proceedings of CAISE, LNCS, 3084, 536-549, 2004
- Kimball, R., Reeves, L., Ross, M., and Thornthwaite, W, "The Data Warehouse Lifecycle Toolkit", second edition, John Wiley & Sons, 1998
- List B., Bruckner R., Machaczek K., and Schiefer, J, "A comparison of data warehouse development methodologies: Case study of the process warehouse", In Proc. DEXA, 2002
- M. Baranovic, M. Madunic, and I. Mekterovic, "Data warehouse as a part of the higher education information system in Croatia", presented at 25th International Conference on Information Technology Interfaces, Cavtat, Croatia, 2003
- M. C. Lin, "University Data Warehouse Design Issues: A Case Study", presented at ASEE Annual Conference & Exposition, 2001

- M. Ester, H. P. Kriegel, J. Sander, M. Wimmer, and X. Xu, "*Incremental Clustering for Mining in a Data Warehousing Environment*", Proceedings of the 24rd International Conference on Very Large Data Bases, pp. 323-333, 1998
- Matteo Golfarelli, Dario Maio, Stefano Rizzi, "*Conceptual Design of data warehouses from E/R schemes*", Proceedings of the Hawaii International Conference on System Sciences, 1998
- Munawar, Naomie Salim, and Roliana Ibrahim, "*Toward Data Quality Integration into the Data Warehouse Development*", Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing. 978-0-7695-4612-4/11 © 2011 IEEE Computer Society. DOI 10.1109/DASC.2011.194, 2011
- Munawar, Naomie Salim, and Roliana Ibrahim, "*Toward Data Warehouse Quality through Integrated Requirements Analysis*", ICACSI 2011. ISBN: 978-979-1421-11-9, 2011
- P. Vassiliadis, "*Gulliver in the land of data warehousing: practical experiences and observations of a researcher*," Proc. 2nd Intl. Workshop on Design and Management of Data Warehouses (DMDW), pp. 12.1–12.16, 2002
- R. G. Allan and D. R. May, "*Data models for a registrar's data mart*", presented at The College and University Services Conference (CUMREC), 2000
- R. G. Allan, "*Data models for a registrar's data mart*", Journal of Data Warehousing, vol. 6, pp. 38-53, 2001
- R. G. Little and M. L. Gibson, "*Perceived influences on implementing data warehousing*," IEEE Transactions on Software Engineering, vol. 29, pp. 290-296, 2003
- R. G. Stephen, "*Building the data warehouse*," Commun. ACM, vol. 41, pp. 52-60, 1998
- S. M. Grotevant and D. Foth, "*The Power of Multidimensional Analysis (OLAP) in Higher Education Enterprise Reporting Strategie*", presented at The College and University Information Services Conference (CUMREC), 1999
- V. Poe, S. Brobst, and P. Klauer, "*Building a Data Warehouse for Decision Support: Prentice-Hall*", Inc. Upper Saddle River, NJ, USA, 1997
- W. H. Inmon, "*Building the data warehouse*", 4th ed, John Wiley & Sons, Inc, New York, USA, 2005
- Westerman, P, "*Data Warehousing using the Wal-Mart model*", p. 297, Morgan Kaufmann
- Wierschem, D., McMillen, J. and McBroom, R, "*What Academia Can Gain from Building a Data Warehouse*", Vol. 26, No. 1, EDUCAUSE Quarterly, 2003