ACADEMIC BUSINESS INTELLIGENCE SYSTEM DEVELOPMENT USING SAS TOOLS

AZMAN TA'A
Faculty of Information Technology
Universiti Utara Malaysia
azman@uum.edu.my

MUHAMAD SHAHBANI ABU BAKAR
Faculty of Information Technology
Universiti Utara Malaysia
shahbani@uum.edu.my

ABDUL RAZAK SALEH
Faculty of Quantitative Sciences
Universiti Utara Malaysia
razak289@uum.edu.my

Managing an organization requires access to information in order to monitor the activities and to assess the performance. In addition, information demand, data volumes, and customer populations are growing and will continue to grow exponentially. Business Intelligence (BI) solutions provide organizations with timely, integrated information that is crucial to the understanding the business environment and customer needs. These BI systems allow an organization to gather, store, access and analyze corporate data sources for business planning and decision-making. In academic institutions, the management demands more information for academic resources planning and academic excellence. This study focuses in defining and developing data warehouse (DW) for academic domain in Universiti Utara Malaysia (UUM). The dimensional model (DM) of DW in student subject area has been defined before the DW was created. A prototype of BI application based on the proposed DW model was developed and linked to the UUM information portal for accessibility. In particular, the proposed model becomes a guideline for the practitioners to develop BI system for academic domain and BI system in general. SAS ETL and SAS Enterprise Guide have been used successfully in developing the prototype of BI system. Both approaches and techniques in data gathering, transformation, loading and analyzing by using SAS tools will be focused and explained.

Keyword: Business Intelligence, Data Warehouse, SAS Application, SAS Data Integration, SAS Enterprise Guide.

1. Introduction

Today's academic management knows that they need to manage vast amounts of data for board of director, administrators, deans, lecturers and students to be able to access the right information so they can satisfy their unique reporting and analysis requirements. Turning that data into information can increase its values and give academic organizations the necessary edge to not just stay in the game, but to stay ahead of it. Decision making by academic management lays the groundwork for lower levels to develop policies and procedures for various academic activities. However, the potential academic strategy of these activities depends on the quality of the decisions and, in turn, on the quality of the data used by them. Some inputs are judgmental, others are from transactional systems, while others are from external sources, but every source collected must have a level of quality appropriate for the decisions making. The competitive forces prevailing in the world of academic institutional today required the Institute of Public Higher Learning (IPHL) to operate as efficiently and productively as possible in order to maintain and enhance students' management, human resources, teaching, researching,

marketing, financing and strategic planning. This will be the motivation of BI system development in academic domain and presented the process of design and develop the BI system for UUM student affairs using SAS tools.

2. Information Requirements in Academic Domain

IPHL is organization focuses on academic businesses. All the business activities are implemented towards achieving academic excellence. Pertaining to that goal, IPHL requires the quality of information from the workable information system which should support the necessary IPHL business processes. In general, there are five (5) main business processes in academic domain defined by Ministry of Higher Education (MOHE)[1] namely i) Student Affairs ii) Staff Affairs iii) Finance Affairs iv) Research and Development Affairs v) Infrastructure Affairs. From all the business processes, many application systems have been developed and collecting huge of data that very useful for the IPHL. The application system such as Student Record System, Student Affair System, Research and Consultation System, Human Resource System, Finance and Accounting System and many more provide an invaluable of information to the IPHL management for planning, forecasting, and implementing. For example, UUM needs to manage a large volume of data and system as shown in Figure 1.

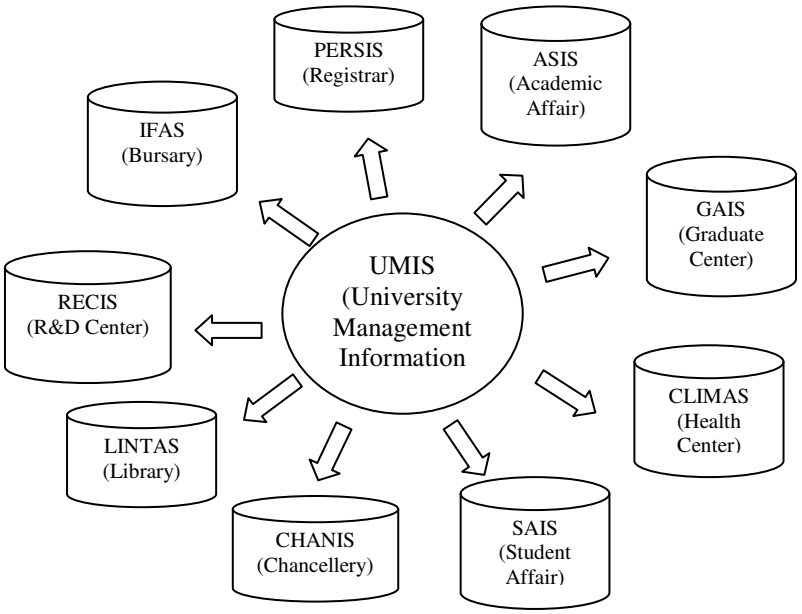

Figure 1: University Management Information System (UMIS)

All decisions must be made as quickly as possible based on the historical and current data through the implementation of the up-to-date tools that store, retrieve, manipulate, analyze and reporting. The IPHL management such as vice chancellor (VC) and deputy vice chancellor (DVC) are seeking synthesized and analyzed operation information to better devise and implement strategic planning based on university environment to more effectively manage available resources in order to meet the needs of the university's customers such as students, parents, stakeholders and government. However, most organizations and their manager encounter

[1]Workshop on Data Collection System for IPHL-MOHE, 13 – 14 December 2006, Awana Hotel Langkawi.

information overload and inaccessible. The availability of information was forcing the IPHL to develop an Information Portal in order to easily deliver the information to the various types of users. A portal becomes a one-stop client-oriented web site that personalizes the information to the user specific needs and characteristics. Thus, BI technology is appropriate to analyze the data, produce the reports and deliver the information accordingly. However, our purpose is to focuses on the business process in student affairs because the information is always being requested by the IPHL management and MOHE for effectively managing the resources of student's registration and enrolment each year.

3. Development of BI System

The development of BI system is usually challenging because it is constrained by information reality. The developer must understand the business requirements, the formats and deficiencies of the data sources, the existing system and the needs of end users. According to Kimball and Caserta (2004), the whole flow for these processes can be viewed as followed:

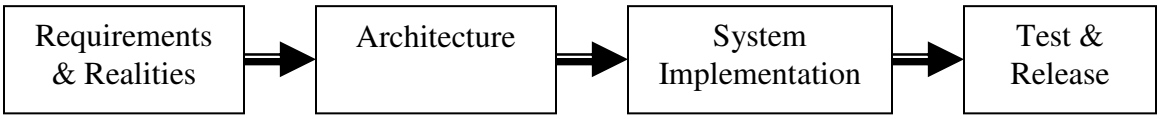

Figure 2: Process Flow for BI System

Basically, these processes are based on generic software development life cycle (SDLC). Thus, it necessary to understand the current realities and defined the user requirements of DW and followed by designing the DW architecture for the academic domain. Since the DW modeling has been completed, the next process will be moved to the activities of extracting, transforming and loading (ETL). Normally, these activities will contribute 70 percent of the whole processes in the BI system implementation (Kimball and Caserta 2004). The model for BI system is adapted from the BI system life cycle (Gangadharan and Swami 2004). However, this study only focuses on process of analyzing user requirements, designing DW, and development of prototype BI application. An analysis user requirement is a process to identify what the users want and expect and checked with the data sources whether the data is available or not. The analyses of user requirements can be tabulated in Table 1.

Table 1: Analyses User Requirements

No.	Requirements	Dimension View	Frequency of view	Users
1.	Number of students register each semester	Category, Program, Race, Nationality, Gender	Each semester	VC,DVC, Dean, Deputy Dean
2.	Number of students enroll each semester	Category, Program, Race, Nationality, Gender	Each semester	VC,DVC, Dean, Deputy Dean, Academic Director
3.	Number of students graduate each year.	Category, Program, Race, Nationality, Gender	Each year	VC,DVC, Dean, Deputy Dean, Academic Director
4.	Number of students stay in hostel each semester.	Category, Program, Race, Nationality, Gender	Each Semester	VC, DVC, Hostel Manager, Affair Director

In design stage, the DW is modeled and designed based on user requirements and data sources structure by applying Dimension Modeling (DM) approach. DM is a logical design techniques often used for DW design (Kimball 1996). Each DM is composed of one table with a multipart key called the *Fact Table* and a set of smaller tables called *Dimension Table*. This well-known structure often called *star schema* or *star join*. The fact table was defined from the measurement required by the users. The measurements were derived from the Table 1 as number of student registered, number of student enrolled, number of student graduated, number of student stayed in hostel and number of student for first class, second class upper, second class lower and failed. From the information provided, the DM for academic domain can be depicted in Figure 3.

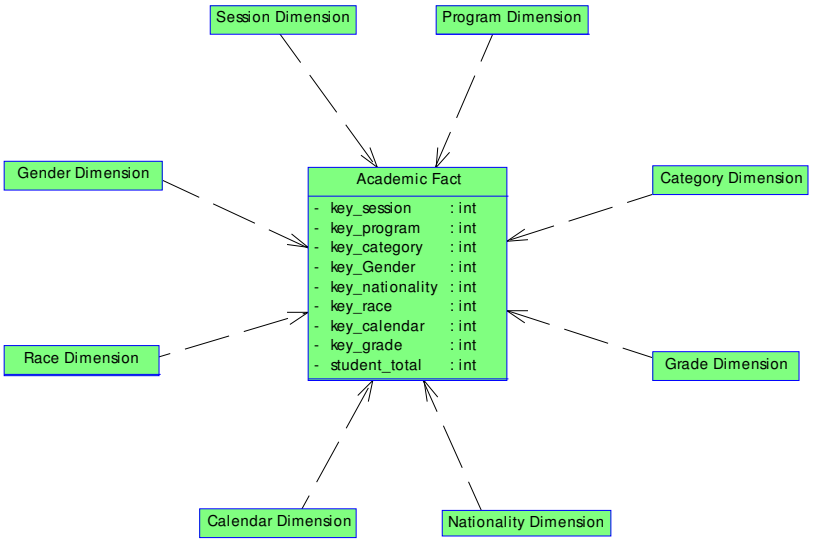

Figure 3: A Dimensional Model for Academic Domain

The measurement value always refers to a total number of students (fact value). As noticed, a dimension table should always be built with a single primary key field that is a simple meaningless integer assigned automatically during the ETL processes called *surrogate keys*. With the huge of data handling by DW, surrogate keys play the main role in order to support the *slowly changing dimension* (SCD) capabilities to maintain the precision of DW. Thus, designed the mapping tables for surrogate keys and corresponding natural keys from the disparate data source become an efficient way to maintain surrogate keys in DW.

In prototype development, the full process of data flow across the BI system have to be documented. This involved storing what type of metadata to capture and delivery to the users. The repository metadata using SAS database structure based on SAS schema has been implemented. The SAS ETL tools has been used for data extracting, cleansing and data transformation. Moreover, in ETL processes, the required data was gathered from the operational data sources and cleaning the data during staging processes. Inside the staging processes, the data will be conformed and transformed to the final DW. Once the final DW is ready, BI application will be developed to analyze and deliver the information to the users.

4. Using SAS Tools in BI System Development

The development and implementation of the ETL processes and BI application need the use of suitable tools, database and platform. Based on BI technology recently, SAS software is among the best tools available in market compare with others technology such as Cognos, Hyperion, TerraData, Informatica and many more. For the fourth year in a row, analyst firm IDC lists SAS as one of the top three leading BI vendors by revenue. IDC's recently released *Worldwide DW Tools*

2005 Vendor Shares report shows SAS is the No. 2 BI vendor worldwide. This factor highly influences us to choose SAS as tools besides the SAS software is available in UUM and always been supported by SAS Vendor. SAS Data Integration (particularly refer to SAS ETL Studio 3.3) provides an intuitive point-and-click process design that allows the designers to easily build logical process workflows for data extraction, cleansing and loading, quickly identify the input and output data flows, and create business rules in metadata for rapidly generate the DW. The entire of ETL process can be divided into three (3) main processes namely extract data sources, transform data process and load to target DW. In detail, the processes can be executed in a series and systematic ways.

Define the data sources is the main step in ETL processes before the next process can be implemented. The data sources provide the data needed by the BI system and being pulling into the staging area through source designer function in SAS ETL tool. By referring to the DW structure as explained in DM, we can detail the architecture design by determine the fact and dimension table specification such as relationship and attributes name. Each of the data/table sources has to define the source properties such as table name, column data, index column, key column, physical storage (which SAS database server located), parameters, extended attributes and notes. The data sources are defined from the Academic Student Information System (ASIS), the Graduate Academic Information System (GAIS), the Student Affair Information System (SAIS) and the Personnel Information System (PERSIS). The snapshot of data sources setting is showed in Figure 4.

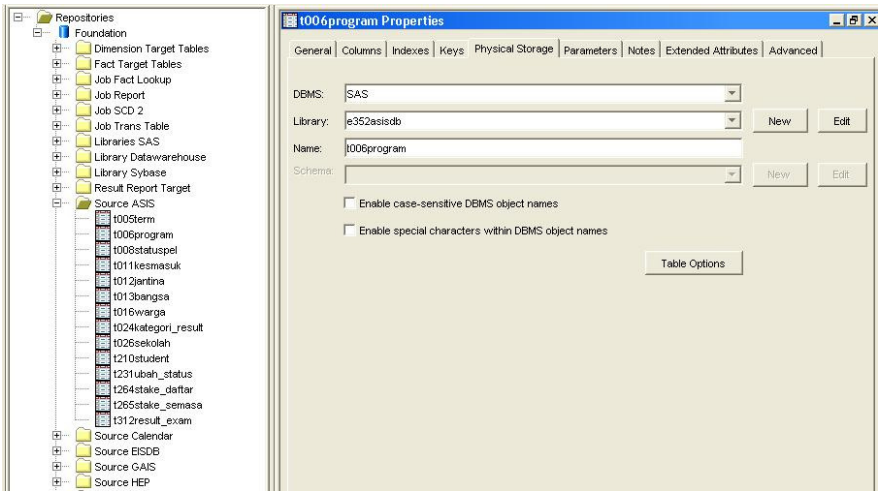

Figure 4: Setting the Data Sources

Define the target DW is important before loading data into final DW. The structure of the target table must be clearly defined and able to map the structure to the table source in staging area. The creation of fact table is done through target designer function. The *student register* fact table contains measurement of students registered in UUM. The fact table name is defined as *tf101masukpelajar* and put the setting into the table properties. Next, the design of transfer data process is created for transforming the data from one state to another. The state is refers to the transformation of data sources to staging data and finally to the DW. The process of transformation is done by the ETL jobs in SAS ETL tool. Specifically, ETL job is a metadata object that specifies processes to create an output for each transformation activity. The transformation activities are extract (*retrieve, filter*), transform (*conversion, join, merge*) and load (*insert*). To create the ETL jobs, we need to understand the process and acknowledge the data in output of transformation activities. For example, to create the *tf101masukpelajar* fact table the ETL job is created for extracting data from data sources (ASIS and GAIS) and store into temporary table (staging area). This can be viewed in snapshot as follows:

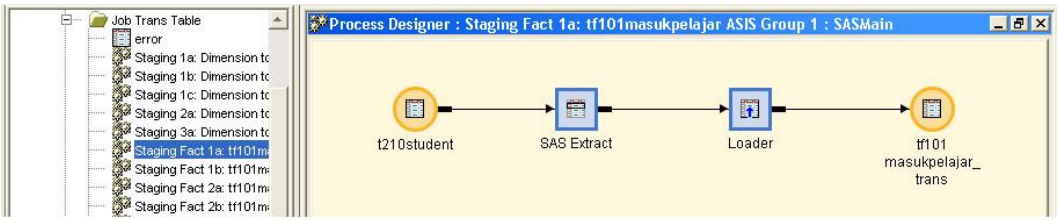

Figure 5: ETL job for data extracting

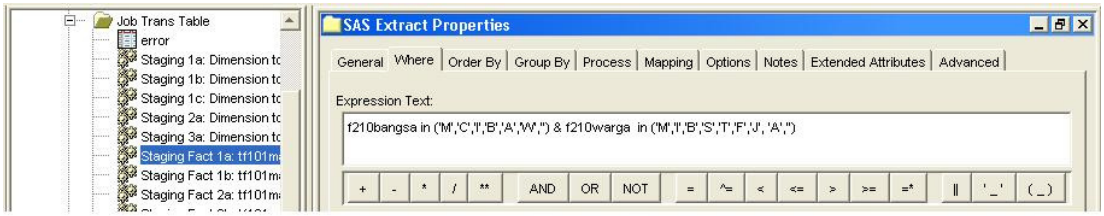

Figure 6: Condition for data extracting

Due to the limited of spaces was unable to explain the whole processes of data transformation and loading into DW. However, when the required data has been loading into DW, the SAS Enterprise Guide 3.0 tool will be used to analyze and design an analytical report for the BI system. A prototype of web-based BI system for academic domain can be tested and validated by the users.

5. Conclusion

Summing up, a major contribution of this study is to define the DW model for academic domain particularly in student affairs area and developing a prototype of BI application for the references of BI designers. The studies also analyzed and modeled the DW requirements and develop the ETL processes by utilizing the SAS ETL tool. Furthermore, the SAS Enterprise Guide also been used to analyze the DW and create the reporting for information presentation in BI system. Additionally, others significant contribution is to provide a platform for researchers and statisticians to perform data analysis and produce reports using BI system with SAS tools.

6. References

[1] Apte C., Liu B., Pednault E. & Smyth P. (2002). Business Applications of Data Mining, Communication of ACM, vol.45, no 8, 49-53.

[2] Kimball, R., & Caserta, J. (2004). The Data Warehouse ETL Toolkit. Practical Technique for Extracting, Cleaning, Conforming and Delivering Data.

[3] Kimball, R. (1996). The Data Warehouse Toolkit - Practical Techniques for Building Dimensional Data Warehouses: John Wiley & Son.

[4] Gangadharan, G. R., & Swami, S. N. (2004). Business Intelligence Systems: Design and Implementation Strategies. 26th ICITI'04, Croatia, 139-144.

[5] SAS Education Division (2006). Course Notes - Using SAS Data Integration Studio to Build Data Marts from Enterprise Data Sources: SAS Publishing Department, USA