

Design and Implementation of Data Warehouse for a Higher Educational Institute in Sri Lanka

C.U. Serasinghe
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information
Technology
Colombo, Sri Lanka
chathraserasinghe89@gmail.com

D.C.R. Jayakody
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information
Technology
Colombo, Sri Lanka
chrishan@outlook.com

K.T.M.N. Dayananda
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information
Technology
Colombo, Sri Lanka
melanid111@gmail.com

Dinesh Asanka
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information
Technology
Colombo, Sri Lanka
dineshasanka@gmail.com

Abstract— In any organization, the leadership is responsible for taking decisions that will lift the said organization to a better place. The problem-solving abilities of the management are mostly depending on the ability to grasp all the required information in a clean and actionable format. Building a well-designed data warehouse leads to answer that problem. When data sourced from different sourcing systems, it's very important that the aggregated data is relevant and supporting the decision-making by the leadership. This study aims at mitigating the issues that are hindering such organizations to make correct decisions.

Keywords—Data Warehouse, Higher Education, ETL, OLAP Cube

I. INTRODUCTION

Higher Education System in Sri Lanka consists of three main categories of higher education institutes naming Universities and Higher Education Institutes under the Ministry of Higher Education, Universities and Higher Education Institutes under other ministries, Private Universities and Private Higher Educational Institutes from which private education has been categorized to degree-awarding institutions, professional institutions and cross-border institutions [1].

A private sector higher education institute which has multiple branches across the country with more than 100 educational programs has been selected for the study and it offers qualifications awarded from well-recognized national and international educational bodies. The courses offered are divided among several departments according to the field of the course. Namely 1) Diploma Division 2) School of Computing 3) School of Business 4) School of Tourism and Hospitality. These departments utilize different types of information systems for sales and marketing, student enrolment, academic and student administration, facilities management, human resource, finance, strategic planning and management purposes. Some of the systems include 1) Inquiry Management System 2) Facilities Management System 3) Customer Relationship Management 4) Attendance Management System & 5) Student information management system. Out of the mentioned systems, the most critical system for operations management is the student

information management system which consists of academic & payment details of the students.

The purpose of this study is to find solutions for critical issues faced by the said organization. 1) Information exists in a standalone data silo. Therefore, useful information for decision-making is not shared among management. For example, the academic division has the details of students such as qualifications, experience & employment details of students that are considered as important factors for sales & marketing teams, but they are unable to access cross-organizational data. 2) Direct access to cross-organizational systems can be at high risk of breaching. Therefore, giving access to confidential student data to the sales and marketing team is not possible in the current situation. 3) top management and the leadership have to access multiple systems on daily basis for analytical purposes whereas inquiries and registrations are runs on two different systems and so data transfer needed to be done manually consuming a lot of labor with a high chance of inconsistency. Another fine example would be setting up sales targets for sales teams based on sales volumes and the number of heads for the annual intake plan. These error-prone manual systems are slowing down the decision-making for the upper management.

Most of the problems in standing alone systems arise when individuals are responsible to create reports on excel and other formats that can be manipulated as per individual needs. These reports may lead to miscommunication and errors in decision making and centralizing systems into a warehouse may reduce the dependency on individual reporting and analysis in this project of designing a data warehouse for the selected higher educational institute.

II. LITERATURE REVIEW

Many of the former studies done similarly in the effect of using data warehouse technologies for decision making in various types of organization has discussed the importance of the design on the data warehouse. According to researchers in the research paper "Materialized view construction in the data warehouse for decision making", the importance of view constructions for decision making using analytical views of data warehouse systems have been discussed [2].

Experiments have resulted indicating that materialized views increase the capability of responding to future queries and reduce the response time.

R. R. Nadikattu in his research paper “Data Warehouse Architecture – Leading the Next Generation Data” describe the magnificent results of using Big Data [3] in educational institutes which have already had a massive impact on customization of learnings. Using historical data of students gathered from student information management system, the analytical system predicts the best courses, student should be taking in upcoming semester to achieve the best results that will enlighten their career paths. This has helped the institutes to improve their student passing out percentage which automatically increased new student enrolments.

Business Intelligence (BI) takes an important trend in the world of business where Educational Intelligence (EI) has become a novel trend in the higher education industry. Education Intelligence Systems are designed to combine different sources of records in a solitary data warehouse. OLAP design sets storage to multidimensional and hierarchical data enabling constant time laps to answer queries [4]. According to researchers, the educational data warehouse contains dimensions such as Schema Proposal, Analysis of the User Requirement, Analysis of Business Requirement, Implementation, Effectiveness and Data Analysis which are performed by identifying that the Star Schema is the highest applied approach [5].

R. Rahutomo, R. Putriand B Pardamean, in their research paper “Building Datawarehouse for Educational Institutions in 9 Steps” describe a practical guide to building a data warehouse for an educational institute [6]. In here they mainly focused on improving the decision making based on the 9-step methodology proposed by Kimball [7]. Using formal observations as well as interviews and questionnaires for few dozens of stakeholders of the system, they were able to identify the real issues that prevent these stakeholders to take quick decisions to improve their day-to-day activities. Using the responses, they were able to come up with a better data warehouse model that will support the stakeholders to make better decisions.

III. NEED OF DATA WAREHOUSE

Higher Education Industry today is extremely competitive due to the privatization of higher education and the availability of alternative qualification pathways that are offered by foreign universities [8]. It has been identified by a survey that students’ enrolment decision is influenced by different factors including recognition of the qualification, course modules, course duration, assessment structure, investment and payment options, etc. Effectiveness of marketing campaigns is the driving factor of organizational success where most institutions have moved from traditional marketing methods to digital marketing approaches. Due to the competition faced by institutions, it is required for organizations to make better decisions.

Using a data warehouse will support the institute to overcome issues that are related to the sales and marketing domain which further connects to student enrolment. In this study, it is planned to understand how the following questions can be answered using concepts in data warehouse.

Q01. What qualifications and professional backgrounds are accepted by the University for Student Registrations of Masters programs? How it can be applied in digital marketing campaigns considering industries that the students work, average years of total experience and list of qualifications to increase the number of inquiries?

Q02. What is the trend of student registrations when institute offers scholarships and discounts?

Q03. What is the trend of student registrations when the institute offers monthly instalment plans?

Q04. How many students continue to register for MBA completing the course hierarchy? For example, number of students registered from L7 Postgraduate Diploma to MBA?

Institute receives thousands of inquiries across 150 courses and over 50 branches island-wide. Distributed OLTP databases retrieve small number of records quickly where data warehouse enables faster retrieval of large volumes of records. It is observed that the reporting and analysis spend more time gathering information from different information systems and manual datasheets before analysis is done. Top Management needs to have the access to timely reports that can be used as analysis to assist business decisions for the long-term success of the organization.

A. Data Sources

Existing data sources of the institute are identified from different functional areas and business processes. Information Systems including Inquiry Handling System, HRIS and MIS are system-based data sources. Other data sources include csv. files, excel files and manual records such as student application forms, copies of student qualification certificates, service letters and copy of payment slips.

B. Data Warehouse Requirements

Institute has different users including the Top Management, Head of Departments, Academic Heads, Branch Managers, Student Counselors, Marketing and Academic Executives etc. Each user has different requirements that are identified by categorizing them into functional and non-functional requirements.

1) Functional Requirements

These requirements define what is the institute expects from data warehouse includes the following:

Data Visualization: Different charts, graphs and comparison tables are demanded by the users based on the queries with expected results.

Data Browsing and Dynamic Queries: OLAP tools can create pivot tables from multidimensional cubes.

Data Integration: SSIS is used to perform ETL including extract data from multiple data sources, transform and load to data warehouse.

Data Validation: Results need to be logical and constant which demands testing and validation.

Remote Access: Many users need to have access to the internet from multiple locations.

Security: Each user will be recognized for authorization based on the designation, job role and level of authority based on the job position.

2) Non-Functional Requirements

These requirements are considered to have less importance, includes;

Simplicity: Design needs to be simple and it is requested not to have any specific knowledge areas or skills.

Scalability: Ability to expand the use of data warehouse in terms of storage and future analysis.

Maintainability: Design should be able to easily updated and continuously improve for future requirements.

IV. DATA WAREHOUSE DESIGN

A. Kimball's Bottom-Up-Approach

As mentioned in Kimball's Approach data warehouse can be defined as "A copy of transactional data that is precisely structured for query and analysis" [9]. Dimensional data marts are created before data warehouse that provides capabilities of reporting and analysis which can be applied to specific business domains such as marketing, academic administration and customer relationship management etc.

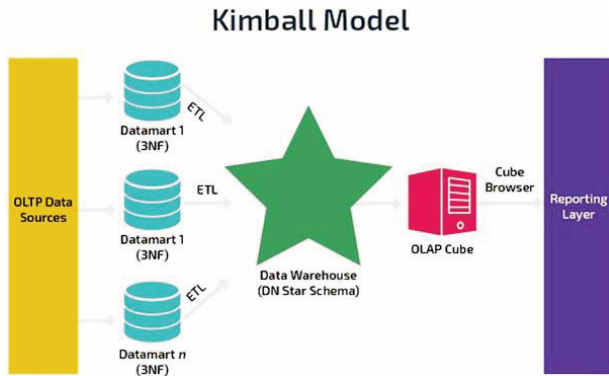


Fig. 1. Kimball's Bottom-Up Approach

9 Step Methodology

Kimball's methodology can be identified as a process approach including 9 steps [6] as following;

Step 01: Choosing the Process means selecting the relevant data warehouse model to the identified business scenario. Student Registrations, Sales and Marketing is part of end-to-end process of the business where it represents multiple complex object collections.

Step 02: Selecting the Grain is where the fact table represents a specific data source [10]. Fact Table is determined as "Student Registrations" which is identified as a *Fact-less Fact Table*.

Step 03: Identifying and confirming Dimensions is an important part of data warehouse design where it determines the dimensions of the fact table. Table 1 shows the list of identified dimensions.

TABLE I. DIMENSIONS TABLE

Dimension Name	Description
Dim_Course	Information that distinguishes each course on the areas such as modules delivered, course leader and course fee etc
Dim_Batch	Information related to cohort segmentation based on the intake

Dim_Student	Information of student includes personal information, professional information and payment details
Dim_Date	Information about timelines such as year, month and semester
Dim_Scholarship	Information that relates to discounted payment offers such as 25%, 50%, 75% or waiving of fees

Step 04: Choosing Facts is another important step in designing the data warehouse. There are multiple grains identified in the study such as Fact_Payment, Fact_inquiries and Fact_CounselorPerformance. Fact_StudentRegistration has been chosen for the data warehouse design.

Step 05: Sorting Pre-calculation in the fact table is vital for the design. When considering the fact_student registrations, it is identified as pre-calculation is not required since it is fact-less fact table.

Step 06: Rounding out the Dimension Table is the next to perform where attributes are identified for each dimension of the chosen fact. Table 2 shows the identified attributes for each dimension.

TABLE II. ATTRIBUTES OF DIMENSIONS

Dimension Name	Attributes
Dim_Course	Course_No, Course_Name, Course_Fee, Course_Leader, Row_Effective_Date, Row_Expire_Date, Is_Current
Dim_Batch	Batch_No, Batch_Status
Dim_Student	Student_Id, Student_Name, Profession, Max_Qualification
Dim_Date	Daynoofweek, Englishdaynoofweek, daynoofmonth, daynoofyear etc
Dim_Scholarship	Scholarship_Type

Step 07: Choosing the duration of the database is how long the data will be retained. Proposed data warehouse is designed for a duration of 10 years.

Step 08: Tracking changes in dimensions with gradual change is slowly changing dimensions. In this study, course fee and student location have been identified as slowly changing dimensions.

Step 09: Determining priorities and query mode is to sort by primary key of each table. It is also automatically indexed by the primary key.

B. Data Warehouse Architecture

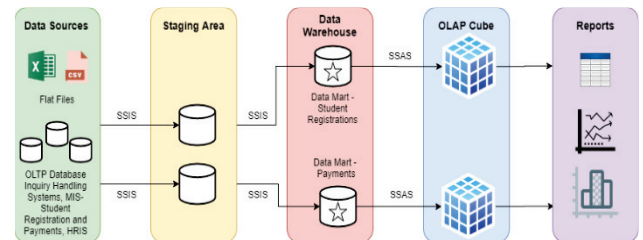


Fig. 2. Data Warehouse Architecture

Figure 2 demonstrates the physical implementation of the data warehouse where it consists of data marts that are formed based on business functions. This design is closely

aligned with Kimball’s Enterprise Data Warehouse Bus Architecture.

C. Proposed Design

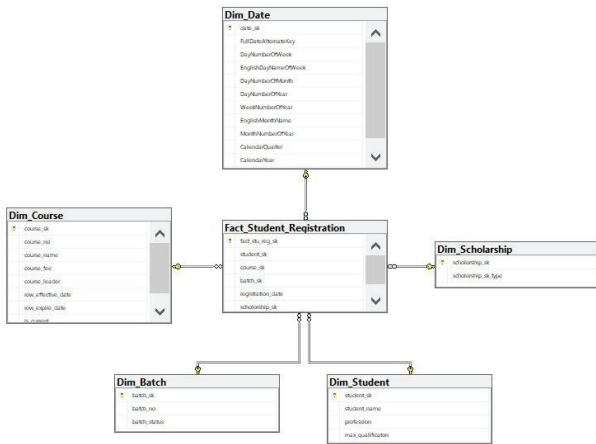


Fig. 3. Proposed Design

The logical design of the Datawarehouse is shown in Figure 3 including relationships among facts and dimensions. It has been identified that the Star Schema is more suitable for this scenario due to its less complexity and faster query response for report generation.

V. EXTRACT-TRANSFORM-LOAD PROCESS

Extract-Transform-Load (ETL) process is considered as the core process of data integration associated with data warehousing. It extracts the data from heterogeneous data sources, transforms it to usable formats to fit in to business requirements and loading it to data warehouse [11].

ETL tools perform cleansing, reformatting, insertion, customization and integration where ETL process considered as a major process in creating a data warehouse which is complex and time-consuming [12]. It demands high implementation efforts, resources and costs.

First, it was required to load the data to staging database from heterogeneous data sources using Microsoft SQL Server Integration Services (SSIS) by forming necessary packages. Data cleansing also applied at this stage. Another SSIS project was created to load the data from staging to data warehouse by creating necessary packages for each dimension and fact tables. Packages consist of ‘data flow tasks’ and ‘execute SQL tasks. Data has been sorted when loading to the data warehouse. Many transforming operations have been used such as data type conversion, splitting, sorting, derived columns and string functions etc. When loading to the destination data warehouse the basic audit logs can be visible in SSIS output. Audit logs include how many records were inserted during login and the execution time. It is important to define error handling accurately in ETL packages.

VI. ANALYSIS

Schemas of the data warehouse are created based on the proposed OLAP design as mentioned in Figure 2. In this stage Microsoft Server Analysis Services (SSAS) is used to

create the cube by loading the above facts and dimension tables to the cube. Once the cube is deployed it can be

browsed where the users can see pivot table in an excel file that enables creating hierarchies, generating necessary reports including;

TABLE III. POTENTIAL REPORTS FOR BUSINESS PROBLEMS

Business Problem	Reports
Q01. What qualifications and professional backgrounds are accepted by the University for Student Registrations of Masters programs? How it can be applied in digital marketing campaigns considering industries that the students work, average years of total experience and list of qualifications to increase number of inquiries?	No.of Student Registrations by intake, No.of Student Registrations by batch, Batchwise student qualifications, Intake wise student qualifications If Fact_inquiries are also loaded to data_warehouse with necessary dimensions, it is possible to analyze inquiry related measures such as; No.of inquiries and No.of student registrations by intake and batch.
Q02. What is the trend of student registrations when institute offers scholarships and discounts?	No.of Student Registrations by intake with scholarships, No.of Student Registrations by intake with discounts
Q03. What is the trend of student registrations when the institute offers monthly instalment plans?	No.of Student Registrations by intake with instalment plan, No.of Student Registrations by cohort with instalment plan
Q04. How many students continue to register for MBA completing the course hierarchy? For example, number of students registered from L7 Postgraduate Diploma to MBA?	No.of Student Registrations with qualification L7 by intake, No.of Student Registrations with qualification L7 by batch Comparison of student registrations from L7 to MBA programme, Conversion ratio by intake, Conversion ration by batch

Analysis of Student Registration Count by profession as shown in Fig 4 Human Resources, Sales and Marketing, IT, Airlines and Banking have a higher number of student registrations which can be used in digital marketing campaigns as potential industries to promote study programs.

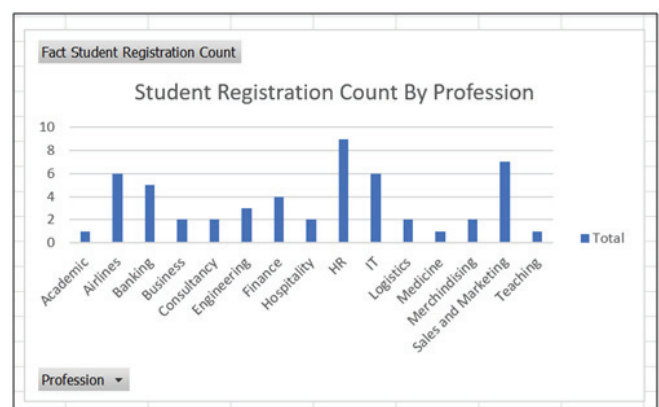


Fig. 4. Student Registration Count by Profession

As per the analysis highest number of student registrations hold bachelor qualifications and it has been identified that there are students with qualifications such as Diploma in Banking and Finance, Professional Qualification in HRM and L7 Postgraduate Diploma who are accepted by the university.

VIII. CONCLUSION AND FUTURE WORKS

Building a productive Datawarehouse for an educational institute involves a considerable amount of cost. Therefore, it is important to properly evaluate the return on investment and benefits to the education institute from data warehouse project. In this paper, the authors have described the model for implementing data warehouse to an educational institute considering the selected business process of student registration. It can be further implemented to other functional areas and business processes as well. Data that was spread across multiple sources were loaded into a single platform to create new knowledge and maintain historical data for future resolutions.

REFERENCES

- [1] Asian Development Bank, "Innovative Strategies in Higher Education for Accelerated Human Resource Development in South East Asia," ADB, Manila, 2016.
- [2] K. Devi and V. Kumar, "Materialised view construction in data warehouse for decision making," *International Journal of Business Information Systems*, vol. 11, no. 4, p. 379, 2012.
- [3] R. R. Nadikattu, "Data Warehouse Architecture – Leading the Next Generation Data Science," *SSRN Electronic Journal*, vol. 67, no. 09, pp. 78-80, 2019.
- [4] Z. Abdullah and T. Obaid, "Design and Implementation of Educational Data Warehouse Using OLAP," *International Journal of Computer Science and Network*, vol. 5, no. 5, pp. 2277-5420, 2016.
- [5] O. Moscoso-Zea, J. Paredes-Gualtor and S. Luján-Mora, "A Holistic View of Data Warehousing in Education," *IEEE Access*, 2018.
- [6] R. Rahutomo, R. A. Putri and B. Pardamean, "Building Datawarehouse for Educational Institutions in 9 Steps," in *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, 2018.
- [7] R. Kimball and M. Ross, *The data warehouse toolkit*, 3rd ed., John Wiley & Sons, Inc., 2013.
- [8] Shaweta, "Critical Need of the Data Warehouse for an Educational Institution and Its Challenges," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 4556-4559, 2014.
- [9] S. Kandawala et al, "A Case Study of Sri Lanka Oil Price Fluctuations and Its Influencing Factors using Predictive Analytics," in *IEEE International Conference on Information and Automation for Sustainability (ICIAFS)*, 2016.
- [10] R. Kimball and M. Ross, *Fact Table Core Concepts in The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*, The Kimball Group, 2016.
- [11] V. Gour, G. Tanwar, A. Sharma and S. Sarangdevot, "Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse," *International Journal on Computer Science and Engineering*, vol. 2, no. 3, 2010.
- [12] S. H. A. El-Sappagha, A. M. A. Hendawib and A. H. E. Bastawissy, "A proposed model for data warehouse ETL processes," *Journal of King Saud University –Computer and Information Sciences*, vol. 23, pp. 91-104, 2011.
- [13] J. Srivastava and I. Y. Chen, "Warehouse Creation- A Potential Road block to Data Warehousing," in *IEEE transactions on knowledge and data education*, 1999.

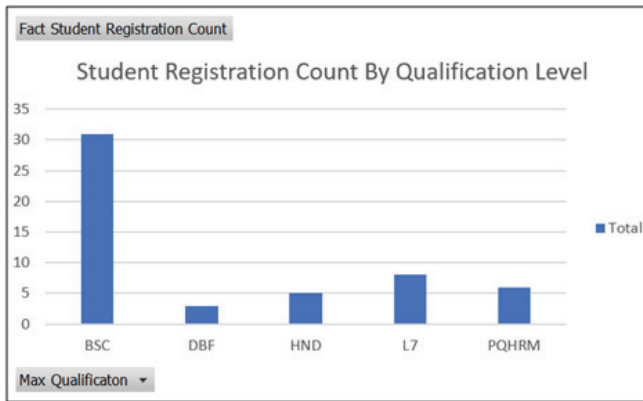


Fig. 5. Student Registration Count by Qualification Level

Scholarships are offered by the institute to encourage student registrations to increase the conversion ratios from inquiries to student registrations. As shown in Fig 6 Student Registration Count by Scholarship is measured.

VII. CHALLENGES

Some of the issues identified related to data warehouse design are related to creation and operationalization. Converting existing databases to the data warehouse design and schema integration is challenging where real-time maintenance also brought concerns to the design which is one of the important factors in educational institute operations [13]. Number of users and a higher number of different end-user types brought difficulties in implementation as different users demand reports and analysis in different forms. Mismatch of data and inconsistency of data are identified as major challenges in data conversion.

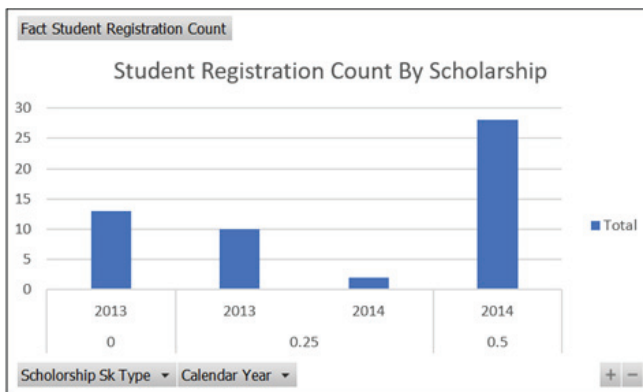


Fig. 6. Student Registration Count by Scholarship