DATA WAREHOUSE DESIGN FOR EDUCATIONAL DATA WITH DATA MINING APPLICATION

College of Arts and Sciences in partial
fulfillment of the requirements for the degree Master of Science (IS)
Universiti Utara Malaysia

By
NURA MUKHTAR
(Matric No: 804090)

© NURA MUKHTAR, 2010
All rights reserved

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

NURA MUKTHAR
(804090)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Intelligent System)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

DATA WAREHOUSE DESIGN FOR EDUCATIONAL DATA WITH DATA MINING APLICATION

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **ASSOC. PROF. FADZILAH SIRAJ**

Tandatangan
(Signature) : _____

PROF. MADYA FADZILAH SIRAJ
Pensyarah
Bidang Sains Gunaan
Kolej Sastera & Sains
Universiti Utara Malaysia

Tarikh
(Date) : 16/05/2010

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master of Science in IS degree from University Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in their absence by the Academic Dean College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Dean (Academic) College of Art and Sciences

University Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman.

Abstract

Large data is stored in the data bases in Secondary Schools, which contain student's demographic and Examination Information. There is a need of these data to be integrated in one place. This study analyses the application of Data warehouse and Data mining application on data of student's previous performance on subject they have took. Where by the data from three Secondary schools were extracted, transformed and loaded into data warehouse. This study also builds student's performance multidimensional cube for each school, a Data mining model was build based on multidimensional cube designed using Microsoft Neural Network as a Data mining tool to predict the Student's performance in their SPM Exam based on their previous subjects performance. The result shows that subject BI from Sama Gagah Secondary School have the highest prediction.

ACKNOWLEDGEMENTS

First, I would like to express my appreciation to Allah, the most merciful and, the most compassionate, who has granted me the ability and willing to start and complete this study. I do pray to His Greatness to inspire and enable me to continue the work for the benefits of humanity.

After that, my most profound thankfulness goes to my supervisor Assoc. Prof Fadzilah Siraj for her scientifically proven and creativity encouraging guidance and great support in this study. Also thank my evaluator Miss Thagirarani for her guidance and support.

Last, I wish to thank my parents and my lovely wife, who were always there for me by giving everything they have, my brothers and sisters for their love and support.

Thank you UUM.

Nura Mukhtar

May 13, 2010

TABLE OF CONTENTS

Page

CHAPTER 3: METHODOLOGY

CHAPTER 4: RESULTS AND DISCUSSIONS

List of Figures

Page

List of Tables

Page

List of Appendices

Page

CHAPTER ONE

INTRODUCTION

This chapter presents the background of the study, in which it explain the motivation behind the project and the domain on which the project is based on. This chapter also describes the problem statement, the objective to accomplished, significance and scope of the study. Finally it highlighted the way subsequent chapters will be organized.

1.1 Background

In the modern world, a lot of data is being received from various sources such as internet, barcode reading, remote sense and organisational database. Therefore huge amount of data is stored in a database system, which makes it difficult for organisations to properly organise and extract meaningful information. Although, traditional operational database store enormous data but only record and a real-time transactional data are managed. The operational database cannot answer a query such as what is the probable cause of profit decrease by 10% last year. This implies that operational database cannot help in making decision. This leads to the development of techniques called data warehouse and data mining which can handle this kind of situation.

As a result of increasing data in operational databases, these data needs to clean, transform and integrated into one centre called data warehouse, Online Analytical Processing (OLAP) is a good way to analyse data in a data warehouse and data mining is used to uncover useful information. Data warehouse is a subject oriented, integrated, time-varying and non-volatile database system for decision support (Vincent & Liu, 1998). The main purpose of data warehouse is to extract data from several sources i.e. external and internal sources, transform the data and make it in a

The contents of the thesis is for internal user only

REFERENCES:

Nikalanta, S., Scheiba, K., & Rai, A. (2008). Dimensional Issues in Agricultural Data warehouse Design. *Computer and Electronics in Agriculture,* 60, 263-278.

Wai, T. T., & Aung, S. S. (2009). Meta data Based Student Data Extraction form Universities Data warehouse. *International Conference on Signal Processing.*

Sahama, T. R., & Croll, P.R. (2007). A Data warehouse Architecture for Clinical Data warehousing. I*n a Conference on Research and Practice in Information Technology, Australia,* 68.

Dan-Ping, Z. (2009). A Data warehouse Based on University Human Resources Management of Performance Evaluation. *Proceedings of IEEE of International Forum on Information Technology and Applications.*

Zhao, H. (2008). Application of OLAP to the Analysis of the Curriculum Chosen by Students. *Proceedings of IEEE on International Conference on Anti-Counter Feting, security and Identification,* 97-100.

Zhenliang, L., & Shufu, D. (2009). Establishment of Water Environmental Data Mart. *Proceedings of IEEE on International Conference on Bioinformatics and Biomedical Engineering,* 1-4.

Tremblay, M. C., Fuller, R., Berndt, D., & Studnick, D. (2006). Doing More With More Information: Changing Healthcare Planning with OLAP tools. *Decision Support Systems,* 43, 1305-1320.

McDonald, J. D., Rajagoplan, S., Waizenegger, J. R., & Pardo, F. (2008). *Realising the Power of Data Marts. Journal of IEEE on Power and Energy Magazine,* 5, 57-66.

Latif, A., Javed, M. Y., & Khan, S. (2008). Semi-Automated Approach for Converting ERD to Semi- Star Schema. *In Proceedings of IEEE in International Conference on Emerging Technologies*, Pakistan.

Singhal, A. (2003). Design of Data warehouse System for Network/Web Services *CIKM, Washington Dc, USA.*

Shi, Z., Huang, Y., Qing, H., Xu, L., Liu, S., Qin, L., Jia, Z., Li, J., Huang, H., & Zhao, L. (2004) MSMiner – A Developing Platform for OLAP. *Decision Support Systems – Elsvier*, Vol. 42, pp. 2016-2028.

Bonifati, A., Cattanio, F., Stefano, C., Fuggetta, A., & Paraboschi, S. (2001). Designing Data Marts for Data warehouses. *ACM Transitions on Software Engineering and Methodology,* 10, 452-483.

Qian, Z., & Qing, X. (2009). The Study on Data warehouse Modelling and OLAP for Highway Management. *International Conference on Measuring Technology and Mechnotronics Automation, IEEE*, 587-590.

Malinowski, E., & Zimanyi, E. (2006). A Conceptual Solution for Representing Time in Data warehouse Dimensions. *Third Asia-Pacific Conference on Conceptual Modelling*, 53.

Koehler, J., Schewe, K. D., & Zhao, J. (2007). Dynamic Data warehouse Design as a Refinement in ASM-Based Approach. *Conference in Research and Practice in Formation Technology*, 67.

Rifaie, M., Blas, E. S., Muhsen, A. M., Mok, T. T. H., Kianmehr, K., Alhajj, R., & Ridley, M. J. (2008) Data warehouse Architecture for GIS Application. *Proceedings of iiWA,* Linz, Australia.

Han, M., & Ju, C. (2008). Research and Application on OLAP-Based Form Product Examination Model. *International Symposium on Electronic Commerce and Security*, 858-861.

Vincent, M., & Liu. J. (1998). AN Architecture for Data Warehouse Systems. *IEEE Conference on Global Connectivity, Computer, Communication and Control,* 1, 107-110.

Milanovic, N., Soskic, G., & Petkovic, A. (2009). Data Warehouse Design for Croatian Students' Nourishment Information System. *IEEE proceedings of Information Technology Interfaces, 193-198*

Jin, H., Wu, T., Lui, Z. & Yan,J. (2009). Application of Visual Data Mining in Higher education Evaluation System. *First International Workshop on Education Technology and Computer Science,* 101-104

Siraj, F., & Abdoulha, M. A. (2009). Uncovering Hidden Information within University's Student Enrolment Data Using Data Mining. *IEEE International Conference on Modelling and Similation,* 413 – 418.

Wook, M., Yahaya, Y. H., Wahab, N., Isa, M. R. M., Awang, N. F., & Seong, H. Y. (2009). Predicting NDUM Student's Academic Performance Using Data Mining Techniques. *An IEEE International Conference on Computer and Electrical Engineering.* 357 – 361.

Elayidom, S., Idikkula, S.M., Alexander, J., & Ojha, A. (2009). Applying Data Mining Techniques for Placement Chance prediction. *IEEE International conference on Advances in Computing, Control and Telecommunication Technologies,* 669 – 671.

Vranic, M., Pintar, D., & Skocir, Z. (2007). The Use of Data Mining in Education Environment. *IEEE International Conference on Telecommunications,* 243 – 250.

Chen, Chien – Ming., Ma, Chen- Hao., Jang, Bin – Shyan., Hsia, Yen – Ten., & Lin, Tsong – Wuu. (2008). Using Data mining to discover the Correlation between Web Laerning Portoforlios and Achievements. F2F-9 – F2F-14.

Kiang, M.Y., Fisher, S.A., Fisher, D.M., & Chi, R.T. (2007). Selecting The Right Peer schools for AACSB Accreditation – A Data Mining Application. *Proceedings of IEEE Symposium on Computational Intelligence and Data Mining.* 31-34

Minaei – Bidigoli,B., KAshy, D. A., Kortemeyer, G., & Punch, W.F. (2003). Predicting Student's Performance: An Application of Data Mining Methods with Educational Web – Based System. *IEEE/ASEE Conference on Frontiers in education,* T2A-13-18.

Ma, Y., Liu, B., Wing, C.K., Yu, P.S., and Lee, S.M (2000). Targeting The Right Students Using data Mining.

Bresfelean, V.S. (2007). Analysis and Predictions on Student's Behavior Using Decision Trees in Weka Environment. *ITI Proceedings for International Conference on Information Technology Interfaces,* 51 – 56.

Pinentel, E. P., & Omar, N. (2005). Towards a Model for Organising and Measuring Knowledge Upgrade in Education with Data mining. *IEEE International Conference on Information Reuse and Integration,* 56- 60.

R. Barquin, H. Edelstein, Planning & Designing the Data Warehouse, Prentice Hall PTR, Upper Saddle River, New Jersey, 1997.

W. H. Inmon, Building the Data Warehouse, Second Edition, John Wiley & Sons, New York, 1996.

P. Gray, H. J. Watson, Decision Support in the Data Warehouse, Prentice Hall PTR, Upper Saddle River,New Jersey, 1998.

Shahzad, K., & Johannesson, P. (2009). An Evaluation of Process Warehousing Approaches for Business Process Analysis. *Proceedings of ECOMAS.* 200-214.