UNCOVERING HIDDEN INFORMATION WITHIN UUM STUDENTS' DATA USING OLAP

A thesis submitted to the Academic Dean Office in partial
fulfilment of the requirements for the degree
Master of Science (Information Technology)
Universiti Utara Malaysia

By
Nur Hani Binti Zulkifli Abai

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)

NUR HANI ZULKIFLI ABAI
(801831)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

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

UNCOVERING HIDDEN INFORMATION WITHIN UUM STUDENTS' DATA USING OLAP

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) : Fadzilah

Tarikh
(Date) : 17/5/2010

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a postgraduate degree from Universiti 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(s) or, in their absence by the Dean of the Academic Office. 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 Universiti Utara Malaysia for any scholarly use which may be made of any material from my 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 of Academic Office

UUM CAS

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman.

ABSTRAK

Penggunaan "online analytical processing" (OLAP) dalam membantu mempercepatkan proses membuat keputusan semakin berkembang di masa kini termasuklah di dalam bidang pendidikan. Banyak produk-produk berkenaan OLAP dan "Business Intelligence" (BI) dipasarkan namun perlaksanaannya di dalam institusi pendidikan perlu dipertingkatkan. Kajian ini menunjukkan perlaksaan OLAP di Universiti Utara Malaysia bagi mengenalpasti maklumat berharga yang terdapat di dalam data pelajar. Ia menekankan mengenai teknik pemilihan dan integrasi data bagi membolehkan pihak pengurusan universiti membuat analisa secara atas talian ke atas data yang telah diintegrasikan. Antara isu yang dibincangkan merangkumi proses pembersihan data bagi mendapatkan data yang berkualiti sebelum sebarang analisa dibuat serta teknik pembangunan OLAP cube menggunakan SQL Server Analisis Services. Kajian ini juga mengupas hasil analisa menggunakan OLAP cube yang membuat capaian data pelajar terus kepada pelayan.

ABSTRACT

The implementation of online analytical processing (OLAP) in speeding decision making process has increased especially in educational industry. Lots of OLAP and Business Intelligence (BI) products have been commercialized, but its implementation in educational industry should be widen up. This study shows the implementation of OLAP in Universiti Utara Malaysia to discover hidden information within UUM students' data. It discussed techniques of data selection and integration to enable university management analyzed integrated students' data online. Issues that have been discussed include data cleaning techniques to ensure data quality before analysis being done and OLAP cube generation technique using SQL Server Analysis Services. This study also elaborates analysis results using OLAP cube that retrieves data online, direct to the server.

ACKNOWLEDGEMENTS

In the name of Allah, the most gracious and the most merciful.

First of all, I would like to express my gratitude and special thanks to my supervisor, Associate Professor Fadzilah Siraj, for her generous comments, encouragement, guidance and most of all her support given while supervising this dissertation.

I also wish to acknowledge my indebtedness to my lovely husband, Loa'i Naser Mahmoud Al-Hawamdeh and my mom, Khamariah Samsudin for their fulfilling support and understanding as well as their prayers. Grateful thanks to my course-mates in the Master of Science (Information Technology), for all the sweet memories shared between us.

I would also like to thank to the Dissertation Committee, Tuan Zalizam for the comments and ideas to improve this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

KDD	-	Knowledge Discovery in Database
DW	-	Data Warehouse
DM	-	Data Mining
OLAP	-	Online Analytical Processing
OLTP	-	Online Transactional Processing
BI	-	Business Intelligence
VC	-	Vice Chancellor
DVC	-	Deputy Vice Chancellor
AVC	-	Assistant Vice Chancellor
HAD	-	Head of Academic Division

CHAPTER 1

INTRODUCTION

This chapter consists of a study on uncovering useful information within UUM students' data using OLAP technology. Students' information from several UUM student databases that have been located in different servers and platforms were integrated for analysis. This chapter highlights research background, problem statements, project's objectives, scope, research questions and significance of the study.

1.1 RESEARCH BACKGROUND

Development of software applications in the whole world has witnessed high data collection activities. Since 1980's, many organizations in various fields were interested in applying operational system application to ease them in daily works and keep all transaction data into their database. This leads to high demand on data storage for keeping daily transaction data. To date, some organizations need trillion byte of data storage to support their needs. Human's capability to understand data is far behind than data storage capabilities. New technique and method that could help us support the requirement of digging useful information in fast grown volume of data

The contents of the thesis is for internal user only

REFERENCES

Anandarajan, M., Anandarajan, A., & Srinivasan, C. A. (2004). *Business Intelligence techniques: A Perspective from Accounting and Finance.* Germany:Springer.

Baker, R. S. J. D., & Yacef, K. (2009). *The State of Educational Data Mining in 2009: A Review and Future Visions.* Retrieved March 20th 2010 from http://www.educationaldatamining.org/JEDM/images/articles/vol1/issue1/JEDMVol1Issue1_BakerYacef.pdf

Bellatreche, L., & Mohania, M. (2008). Physical Data Warehousing Design. In Wang, J.(Ed), *Data Warehousing and Minng : Concepts, Methodology, Tools and Applications (Volume 1).* 18- 25. New York: Information Science Reference, Hershey.

Bernardino J. (2002) Approximate Query Answering Using Data Warehouse Striping. *Journal of Intelligent Information Systems*, 19:2, pp. 145-167.

Boyle, R., Carter, J., & Clark, M. (2002). What makes them succeed? Entry, progression and graduation in Computer Science. *Journal of Further and Higher Education, 26*(1), 3-18.

Burdict, D., Deshpande, P. M., & Jayram, T. S. (2007). OLAP over uncertain and imprecise data. *The International Journal on Very Large Data Bases,* 16(1), 123-144. Retrieved 23rd March 2010 at http://vldb.idi.ntnu.no/program/paper/fri/p970-burdick.pdf

Cao, L.(2009). Introduction to Domain Driven Data Mining. In Cao, L., Yu, P.S., Zhang,C. & Zhang, H.(Eds). *Data Mining for business applications.* pp : 3 – 10. Springer : USA.

Chaudhuri, S., & Dayal, U.(1997). An Overview of Data Warehousing and OLAP Technology. SIGMOD Record, 26(1), 65-74.

Codd, E. F., Codd, S. B., & Salley C. T. (1993). *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate.* Codd & Date, Inc.

Delavari, N., Beikzadeh, M. R., & Phon-Amnuaisuk, S. (2005). Application of enhanced analysis model for data mining processes in higher educational system. In *6th Annual International Conference (ITHET),* (1-6). Juan Dolio, Dominican Republic.

Deshpande, P. M., & Ramasamy, K. (2008). Data Warehousing, Multi-Dimensional Data Models and OLAP. In Wang, J.(Ed), *Data Warehousing and Minng : Concepts, Methodology, Tools and Applications (Volume 1).* pp: 179- 186. New York: Information Science Reference, Hershey.

EDM Working Group (2009). *International Working Group on Educational Data Mining.* Retrieved 19th January 2010. Website: http://www.educationaldatamining.org/index.html

Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceeding of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96).* Portland, Oregon. August 2-4, 1996.

Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases.* American Association for Artificial Intelligent.

Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). *The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of The ACM*, 39(11), 27-34.

Gilad, B., & Gilad, T. (1988). *The Business Intelligence System : A New Tools for Competitive Advantage*. American Management Association.

Giudici, P. (2003). *Applied Data Mining : Statistical methods for business and industry*. John Wiley & Sons Inc : England.

Hancock, J. C., & Toren, R. (2007). *Practical Business Intelligence with SQL Server 2005.* Addison-Wesley, Pearson Education, Inc. USA : Upper Saddle River.

Hao, Y., & Xing-chun, D. (2008). The Design and Implementation of Data Cleaning Knowledge Model. *International Symposium on Knowledge Acquisition and Modeling 2008.* Retrieved on 23rd March 2010 from http://ieeexplore.ieee.org.eserv.uum.edu.my/stamp/stamp.jsp?tp=&arnumber=4732810

Hasan, H., & Hyland, P. (2001). Using OLAP and Multidimensional Data for Decision Making. *IT Pro*, pp. 1-7.

Hernandez-Orallo, J. (2008). Data Warehousing and OLAP. In Wang, J. (Ed), *Data Warehousing and Mining : Concepts, Methodology, Tools and Applications (Volume 1)*. pp: 169- 178. New York, Information Science Reference, Hershey

Inmon, W. H. (1996). *Building the Data Warehouse*. John Wiley & Sons, Inc. 2nd Edition, 1996.

Inmon, W. H., Strauss, D., & Neushloss,G. (2008). *DW2.0 : The Architecture for the Next Generation of Data Warehousing*. Morgan Kaufmann Publishers. United States.

Jarke, M., Lenserini, M., Vassiliou, Y., & Vassiliados, P. (2003). *Fundmentals of Data* Warehouses: Second, Revised and Extended Edition. Germany. Springer-Verlag Berlin Heidelberg

Klosgen, W., & Zytkow, J. M. (2002). *Handbook of data mining and knowledge discovery*. New York : Oxford.

LearnDataModeliing.Com.(2008). *Designing Start Schema*. Retrieved 21st March 2010 at http://www.learndatamodeling.com/star.htm

Luan, J. (2004). *Data Mining Applications in Higher Education*. SPSS Executive Report.

Marshall, T. (2008). *Data Warehouse Success Story*. Retrieved April 1st, 2010 from http://www.uncg.edu/ism/ism611/Story.PDF

Merceron, A., & Yacef, K. (2005). Educational Data Mining: A Case Study. *Artificial Intelligence in Education (AIED2005)*, C.-K. LOOI, G. MCCALLA, B.BREDEWEG and J. BREUKER Eds. IOS Press, Amsterdam, The Netherlands, 467-474.

Microstrategy, Inc. (1995). *The Case for Relational OLAP*. Retrieved 21st March 2010 at http://www.cs.bgu.ac.il/~dbm031/dw032/Papers/microstrategy_211.pdf

Nguyen, T. N., Janecek, P., & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. In *37th annual Frontiers in education conference - global engineering: knowledge without borders, opportunities without passports (FIE '07)*, (pp. T2G-7 - T2G-12).

Oliveira, P., Rodrigues, F., & Henriques, P. (2009). SmartClean: An Incremental Data Cleaning Tool. *Ninth International Conference on Quality Software.*

Pimentel, E. P., & Nizam, O. (2005). Towards a model for organizing and measuring knowledge upgrade in education with data mining. In *IEEE International Conference on Information Reuse and Integration (IRI -2005),* 56-60. Nevada, USA.

Prabhu, C. S. R. (2002). *Data Warehousing:Concepts, techniques, products and Applications*. Second Edition. New Delhi, India: Prentise Hall.

Priebe, T. (2000). Towards OLAP Security Design. *Survey and Research Issues*, 33-41

Quafafou, M., Naouali, S., & Nachouki, G. (2005). Knowledge Datawarehouse: Web Usage OLAP Application. *International Conferenc on Web Intelligence.*

Raisinghani, M. (2004). *Business Intelligence in Digital Economy*. Hershey PA : The Idea Group.

Rizzi, S. (2008). Conceptual Modelling Solutions for the Data Warehouse. In Wang, J.(Ed), *Data Warehousing and Mining : Concepts, Methodologies, Tools and Applications (Volume 1)*. pp: 208- 227. New York, Information Science Reference, Hershey

Rusu, L. I., Rahayu, J. W., & Taniar, D. (2008). A Methodology for building XML Data Warehouse In Wang, J.(Ed), *Data Warehousing and Mining : Concepts, Methodologies, Tools and Applications (Volume 1).* 530- 555. New York, Information Science Reference, Hershey

Seifert, J. W. (2004). *Data Mining : An Overview*. CRG Report for Congress.

Shoshani, A. (1997). OLAP and statistical databases: Similarity and differences. *In proceeding of the sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 185-196.

Shumate, J. (2000). *A Practical Guide to Microsoft OLAP Server.* Addison-Wesley, Upper Saddle River, USA.

SPSS, Inc.(1997). *What is data mining*? Retrieved March 21st, 2010, from www.spss.com/datamine/define.htm.

Sriraam, N., Natasha, V., & Kaur, H. (2008). Data Mining Techniques and Medical Decision Making for Urology Dysfunction. In Wang, J.(Ed), *Data Warehousing and Mining : Concepts, Methodologies, Tools and Applications (Volume V)*. 2506 - 2516. New York, Information Science Reference, Hershey

Tissera, W. M. R., Athauda, R. I., & Fernando, H. C. (2006). Discovery of strongly related subjects in the undergraduate syllabi using data mining. *International Conference on Information and Automation (ICIA '06),* 57-62.

Turban, E., Sharda, R., Aronson, J. E., & King, D. (2008). *Business Intelligence : A managerial Approach.* Prentice Hall, Upper Saddle River, New Jersey.

Vranic, M., Pintar, D. & Skocir, Z. (2007). The use of data mining in education environment. In *9th International Conference on Telecommunications (ConTel 2007),* (pp. 243-250). Croatia.

Wah, T. Y., & Bakar, Z. A. (2003). Investigating the Status of Data Mining in Practice. Informing Science : *InSITE - "Where Parallels Intersect"*. 1405-1413.

Wrembel, R., & Koncilia, C. (2007). *Data Warehouses and OLAP : Concepts, Architecture and Solutions*. IRM Press. United States of America.

Xin, D. & Hin, J. (2007). Integrating OLAP and Ranking : The Rabking-Cube Methodology. *IEEE 23rd International Conference on Data Engineering Workshop,* 253-256.

Zarate, L. E., Nogueira, B. M., Santos, T. R. A., & Song, M. A. J. (2006). Technique of Missing Values Recovering in Imbalance Database : Application in Marketing Database with Massive Missing Data. *International Conference of Systems, Man and Cybernatic (2006)*, 2658-2664, Taipei, Taiwan