

DESIGN DATA WAREHOUSE FOR MEDICAL DATA

MUSTAFA MUSA JABER (801607)

UNIVERSITI UTARA MALAYSIA 2010

DESIGN DATA WAREHOUSE FOR MEDICAL DATA

**A Thesis submitted to Faculty of Information Technology in partial
fulfillment of the requirements for the degree Master
(Information and Communication Technology),
Universiti Utara Malaysia**

By

MUSTAFA MUSA JABER (801607)

© MUSTAFA MUSA JABER, 2010. All rights reserved.

FADZILAH SIRAJ



**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)

MUSTAFA MUSA JABER
(801607)

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)

DESIGN DATA WAREHOUSE FOR MEDICAL DATA

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**

PROF. MADYA FADZILAH SIRAJ

Pensyarah
Bidang Sains Gunaan
Kolej Sastera & Sains
Universiti Utara Malaysia

Tandatangan
(Signature)

: 

Tarikh
(Date)

: 16 / 5 / 2010

PERMISSION TO USE

In presenting this project of the requirements for a Master of Science in Information Technology (MSc. IT) 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 project paper in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or in their absence, by the Dean of Graduate School. It is understood that any copying or publication or use of this project 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 my project paper.

Request for permission to copy or make other use of materials in this project, in whole or in part, should be addressed to:

Dean of Graduate School

University Utara Malaysia

06010 Sintok

Kedah Darul Aman

Malaysia

ABSTRACT

Organizing and managing the database relations in term of data warehouse technology has been addressed widely in different complex environments. The data warehouse contains a source of valuable data mining. The data contained in the data warehouse is cleaned, integrated, and organized. This study highlighted the existing issues on the medical databases which present a huge number of information across various departments, managing this type of data require time, and laborious tasks to separately access and integrate reliably. Hence, this study aimed to model new medical data warehouse architecture for managing and organizing the medical dataset operation into data warehouse. Technically OLAP has been used to design the proposed architecture, for the hospitable administrators, and top manager and/or sophisticated user can use MDW by using Microsoft SQL Server 2005. Building the proposed architecture adopted by using Microsoft Visual Studio for performing the OLE database operations. The performing process has been tested through the using of use test case technique.

ACKNOWLEDGEMENTS

First and for most, my solemn gratitude and sincere appreciation goes to the Almighty Creator, the giver of life, health and knowledge, for blessing me with the gift of life. Secondly, my profound gratitude goes to my supervisor Assoc. Prof. Fadzilah Binti Hj. Siraj for her helpful guidance and most importantly giving me the inspiration during the course of this study. Moreover, I would also like to thank her for the opportunities that she has made available to me.

I am always grateful to all members of my family, first being my parents, I am thanking them for every prayer and supplication they made for me and as well for their love and most importantly being my parents. Equal gratitude goes out to my siblings, brothers, and the friend and the brother Mohammed Aal Mohammed for caring me when I was alone overseas. Finally I want to thank the friend and the sister Hasnida Hajitoleh for her supporting to me.

I would like to conclude my acknowledgment, by once again thanking the Almighty God.

Thursday, May 13, 2010, Mustafa Musa Jaber.

TABLE OF CONTENTS

PERMISSION TO USE	I
ABSTRACT	II
ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
LIST OF TABLES	VI
LIST OF FIGURES	VII
CHAPTER ONE	
	<i>Page Num</i>
INTRODUCTION	1
1.1 BACKGROUND	1
1.2 PROBLEM STATEMENT	3
1.3 RESEARCH QUESTION	4
1.4 OBJECTIVE OF STUDY	4
1.5 SIGNIFICANCE OF STUDY	4
1.6 SCOPE OF THE STUDY	5
CHAPTER TWO	
LITERATURE REVIEW	6
2.1 DATA WAREHOUSE	6
2.2 A NOVEL THREE-LEVEL ARCHITECTURE	7
2.3 MULTI-LAYER ARCHITECTURE OF DATA WAREHOUSE	8
2.4 DATA WAREHOUSE ARCHITECTURE FOR GIS APPLICATIONS	9
2.5 DATA WAREHOUSE ARCHITECTURE FOR NETWORK/ WEB SERVICES	10
2.6 WAREHOUSING ARCHITECTURE FOR CLINICAL DATA	11
2.7 DATA WAREHOUSE ARCHITECTURE	14
2.8 ONLINE ANALYTICAL PROCESSING (OLAP)	16
2.9 RELATED WORKS	19
2.10 SUMMARY	22
CHAPTER THREE	
RESEARCH METHODOLOGY	23
3.1 DESIGN RESEARCH METHODOLOGY	23
3.1.1 AWARENESS OF THE PROBLEM	25
3.1.2 SUGGESTION	25
3.1.3 DEVELOPMENT	26
3.1.3.1 Data Sources	27
3.1.3.2 Integration	29

3.1.3.3 Data Warehouse	30
3.1.3.4 Customization	31
3.1.3.5 Application	32
3.1.4 EVALUATION	32
3.1.5 CONCLUSION	33
3.2 SUMMARY	33

CHAPTER FOUR

ANALYSIS AND RESULT	34
4.1 INTRODUCTION	34
4.2 FUNCTIONAL REQUIREMENTS	35
4.3 MEDICAL DATA WAREHOUSE USE CASE	36
4.4 SEQUENCE DIAGRAM AND COLLABORATION DIAGRAM	37
4.5 DATA WAREHOUSE COMPONENTS	47
4.5.1 Data Warehouse	47
4.5.2 Data Marts	48
4.5.3 Data Extraction-Transformation-Load (ETL)	49
4.5.4 Metadata and the Metadata Repository	50
4.6 DISCRIPTIVE ANALYSIS	50
4.7 APPLICATION TESTING	51
4.7.1 Select Data Set Use Test Case	51
4.7.2 Create OLE Database Use Test Case	52
4.7.3 Analyze Data Set Use Test Case	52
4.7.4 View Database Clusters Use Test Case	53
4.7.5 View Database Statistics Use Test Case	53
4.8 LIMITATION	54
4.9 CONCLUSION	54

CHAPTER FIVE

CONCLUSION	55
5.1 INTRODUCTION	55
5.2 ARCHITECTURE STRENGTH	56
5.3 ARCHITECTURE LIMITATION	56
5.4 CONCLUSION	57
REFERENCES	58

LIST OF TABLES

TABLES TITLE	Page Num
Table 2.1: Related works	20
Table 3.1: Mapping between research objects and research methods	25
Table 4.1: Functional and non-functional requirements	35
Table 4.2: Select dataset use test case	52
Table 4.3: Create OLE database use test case	52
Table 4.4: Analyze dataset use test case	53
Table 4.5: View database clusters use test case	53
Table 4.6: View database statistic use test case	53

LIST OF FIGURES

Figure Title	Page Num
Figure 2.1: Three Level Architecture	7
Figure 2.2: Multi-layer architecture of data warehouses customization	9
Figure 2.3: Data Warehouse Architecture for GIS Applications	10
Figure 2.4: Architecture of the Data Warehouse System	11
Figure 2.5: Different types of DW Architectures	12
Figure 2.6: Enterprise DW Architecture	13
Figure 2.7: Distributed DW Architecture	13
Figure 2.8: Data Mart Architecture	14
Figure 2.9: Data Warehouse Architecture	15
Figure 2.10: Data Model for OLTP	18
Figure 2.11: Star Data Model for OLAP	18
Figure 2.12: OLAP Cube with Time, Customer and Product Dimensions	19
Figure 3.1: The General Methodology of Design Research Vaishnavi and Kuechler	24
Figure 3.2: Enterprise DW Architecture	26
Figure 3.3: Medical Warehouse Architecture	27
Figure 3.4: Medical Data Warehouse use case	28
Figure 3.5: Create Table in SQL Server 2005	28
Figure 3.6: Creating ETL Tool Connection	29
Figure 3.7: Creating Data Warehouse	30
Figure 4.1: Medical Data Warehouse use case	36
Figure 4.2: Select Data-Set Sequence Diagram	37
Figure 4.3: Select Data-Set Collaboration Diagram	38
Figure 4.4: Create OLE Database Sequence Diagram	39
Figure 4.5: Create OLE Database Collaboration Diagram	40
Figure 4.6: Analyze OLE DB Sequence Diagram	41

Figure 4.7: Analyze OLE DB Collaboration Diagram	42
Figure 4.8: View DB Cluster Sequence Diagram	43
Figure 4.9: View DB Cluster Collaboration Diagram	44
Figure 4.10: View Statistic Sequence Diagram	45
Figure 4.11: View Statistic Collaboration Diagram	46
Figure 4.12: Getting Data Set from UCI	47
Figure 4.13: Create Table in SQL Server 2005	48
Figure 4.14: Creating Medical Data Warehouse	49

CHAPTER ONE

INTRODUCTION

This chapter presents the background of the study, the problem statement, the research question, the objective of study, the significance of study, and scope of the study.

1. 1 BACKGROUND

Inmon (2002) defines data warehouse as the heart of the architected environment, and is the foundation of all decision support system (DSS) treatment. The work of the DSS analyst in the environmental data warehouse is much easier than in the environment of the classical heritage because there is a single integrated source of data (the data warehouse) and because the granular data in data warehouse is easily accessible.

The data warehouse contains a source of valuable data mining. The data contained in the data warehouse is cleaned, integrated, and organized. And the data are historical. It should be noted that although the data warehouse is an excellent source of data for the minor and the explorer, the data warehouse is often the only source. External data and other data can be freely mixed with data warehouse data as part of their exploration and mining (Akoka *et al.*, 2007).

The contents of
the thesis is for
internal user
only

REFERENCES

- Adrian, T. & Boussaid, F. (2006). Towards Complex Data Warehousing : A new approach for integrating and modeling complex data. Retrieved March, 27, 2010, from www.phd.at8.fr/papers/MCO04_Towards_Complex_Data_Warehousing.pdf.
- Adrian, T. & Boussaid, O. (2003). CDO2XML - Complex Data Object to XML - XML documents generation prototype. Retrieved March, 23, 2010, from <http://www.bat710.univ-lyon1.fr/~atanases/CDO/install.en.zip>.
- Aisha, L., Younus, J. & Ayesha, A. (2009). Hybrid Technique of using ANN in Semi-Star Schema Generation. *International Journal of Hybrid Information Technology* 2(2).
- Akoka, J., Berti, L., Boucelma, O. & Bouzeghoub, O. (2007). A Framework for Quality Evaluation in Data Integration Systems.
- Bose, I. & Mahapatra, R. (2001). Business data mining - a machine learning perspective. *Information and Management*, 39(3), pp.211-225.
- Cassandra, P. & Karen, D. (2002). Automating Data Warehouse Conceptual Schema Design and Evaluation. Retrieved March, 22, 2010, from www.lrrb.org/pdf/200933TS.pdf.
- Chau, K. & Ying, C. (2002). The Application of data warehouse and decision support system in construction management, *Automation in Construction* 12, 213-224.
- Chaudhuri, S. & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26(1), 65-74.
- Chen, X., Neil, P. & Neil, E. (2008). Adjoined Dimension Column Clustering to Improve Data Warehouse Query Performance. *ICDE 2008*, pp. 1409-1411.

- Delin, Q. (2009). Design of Medical Insurance Supervision System Based on Active Data Warehouse and SOA. Computer Science and Information Engineering, WRI World Congress, China.
- Fayyad, U., Piatetsky-Shapiro G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17, pp. 37-54.
- Fang, R. & Tuladhar, S. (2006). Teaching Data Warehousing and Data Mining in a Graduate Program in Information Technology. *Journal of Computing Sciences in Colleges*, 21(5), pp. 137-144.
- Holsheimer, M. (2009). Data mining by business users: integrating data mining in business processes. In: J. Han (ed.) *Tutorial Notes of the 5th ACM International Conference on Knowledge Discovery and Data Mining*. ACM, New York, pp. 266-291.
- Hornick, M., Marcad, E. & Venkayala, S. (2006). *Java Data Mining: Strategy, Standard, and Practice*. Morgan Kaufmann, San Francisco.
- Husemann, B., Lechtenborger, J. & Vossen, G. (2000). Conceptual data warehouse design. In *Proceedings of International Workshop on Design and Management of Data Warehouses*, Stockholm.
- Jane, Z. (2006). Designing Distributed Data Warehouses and OLAP Systems. Retrieved February, 21, 2010, from www.subs.emis.de/LNI/Proceedings/.../GI-Proceedings.63-21.pdf.
- Jianping, Z., Tianyi, M. & Qiping, S. (2004). Application of Data Warehouse and Data Mining in Construction Management. Retrieved March, 3, 2010, from e-pub.uni-weimar.de/volltexte/2004/127/pdf/icccbe-x_182.pdf.
- Inmon, W. (2002). *Building the Data Warehouse* (3 ed). John Wiley & Sons.

- Kang (2002). Exploiting Versions for On-Line Data Warehouse Maintenance in MOLAP Servers, Proc. Of VLDB Conference, China.
- Kasinadh, D. & Krishna, P. (2007). Building fuzzy olap using multi-atribute summarization. International Conference on Computational Intelligence and Multimedia Applications.
- Kimball, R. & Ross, M. (2002). The Data Warehouse Toolkit. Wiley Publishing, Inc.
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. & Becker, B. (2008). The Data Warehouse Lifecycle Toolkit. Wiley Publishing, Inc.
- Kimball, R., Ross, M. & Merz, R. (2002). The data warehouse toolkit: the complete guide to dimensional modeling: Wiley.
- Lawrence, H. & Meral, O. (2001). Data Mining, Data Warehousing & OLAP and E-Commerce Breakout Group Report. Retrieved March, 23, 2010, from itlab.uta.edu/idm01/FinalReports/DataMning.pdf. NSF IDM Workshop.
- Lenz, H. & Shoshani, A. (1997). Summarizability in olap and statistical data bases. Statistical and Scientific Database Management, 1997.
- Meier, A., Schindler, G. & Werro, N. (2008). Handbook of Research on Fuzzy Information Processing in Databases. (J. Galindo, Ed.) IGI Global. , 2, 586-614.
- Mailvaganam, H. (2007). Introduction to OLAP [Electronic Version]. Retrieved February, 23, 2010 from http://www.dwreview.com/OLAP/Introduction_OLAP.html.

- Mohammad, R., Erwin, J. B., Abdel Rahman, M. M., Terrance, T. H. M., Keivan, K. & Reda, A., (2008). Data warehouse architecture for GIS applications. Paper presented at the Proceedings of the 10th International Conference on Information Integration and Web-based Applications Services.
- Mohammad, A. & Michael, E. (2007). Case Projects in Data Warehousing And Data Mining. *Issues in Information Systems*. 8(1).
- Mukesh, M., Kamalakar, K. & Yahiko, K. (1999). Invited Talk: Data Warehouse Design and Maintenance through View Normalization. *Database and Expert Systems Applications*. 1677.
- Marbn, O., Segovia, J., Menasalvas, E. & Fernandez-Baizn, C. (2009). Toward data mining engineering: A software engineering approach. *Information Systems* 34 (1).
- Marko, B., Min, T. & Nevena, S. (2007). Integrating different grain levels in a medical data warehouse federation. Retrieved March, 22, 2010, from <http://wirehouseapplicatiojn.pdf>.
- Nichols, J. & Poe, S. (1999). Quality assurance, practical management, and outcomes of point-of-care testing Laboratory perspectives, Part 1. *Clin Lab Manage Rev*.
- Perez, D., Somodevilla, M. & Pineda, H. (2007). Fuzzy spatial data warehouse: A multidimensional model. Eighth Mexican International Conference on Current Trends in Computer Science.
- Rousset, M. (2002). Knowledge representation for information integration. In XIIIth Int. Symp. on Methodologies for Intelligent Systems (ISMIS 2002), 2366, pages 1-3. Springer Verlag.

- Reijers, H. & Mansar, S. (2005). Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. *Omega – The International Journal of Management Science* 33 (4), pp. 283-306.
- Rupnik, R. & Jaklic, J. (2009). The Deployment of Data Mining into Operational Business Processes. In: Ponce, J., Karahoca, A. (eds.) *Data Mining and Knowledge Discovery in Real Life Applications*, I-Tech.
- Schepperle, H. & Merkel, A. (2004). Data warehouse. *Internationals Symposium: Data-Warehouse-System und Knowledge-Discovery*.
- Sen, A. & Sinha, A. (2005). A Comparison of Datawarehousing Methodologies, *Communication of the ACM*, 48(3), 79-84.
- Sharma, S. & Osei-Bryson, K. (2009). Framework for formal implementation of the business understanding phase of data mining projects. *Expert Systems with Applications* 36 (2).
- Tagaris, A., Konnis, G., Benetou, X., Dimakopoulos, T., Kassis, K., Athanasiadis, N., R uping, S., Grosskreutz, H. & Koutsouris, D. (2009). Integrated Web Services Platform for the facilitation of fraud detection in health care e-government services. In: *Proc. ITAB 2009*.
- Tony S. & Peter C. (2007). *Data Warehouse Architecture for Clinical Data Warehousing*. Australian Computer Society. *Conferences in Research and Practice in Information Technology*, 68.
- Wierschem, D., McMillen, J. & McBroom, R. (2003). What Academia Can Gain from Building a data Warehouse,” *Educause Quarterly*, 1, pp. 41-46.

- Fang, R. & Tuladhar, S. (2006). Teaching Data Warehousing and Data Mining in a Graduate Program in Information Technology,” *Journal of Computing Sciences in Colleges*, 21(5), pp. 137-144.
- Wen, Y. & Chin, A. (2000). *An Object-Relational Modeling for Clinical Data Warehouse*. Nioa Song, Kaohsiung county, Taiwan.
- Werro, N. (2008). *Fuzzy Classification of Online Customers*. PhD thesis, University of Fribourg.
- Witten, I. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. 2nd ed., Morgan Kaufmann, San Francisco.
- Wegener, D., Sengstag, T. & Sfakianakis, S. (2009). GridR: An R-based tool for scientific data analysis in grid environments. *Future Generation Computer Systems* 25 (4), pp. 481-488.
- Zimmerman, H. (1991). *Fuzzy Set Theory and its applications*. Kluwer Academic Publishers.
- Zhao, J. & Ma, H. (2004). Quality-assured design of on-line analytical processing systems using abstract state machines. In: Ehrich, H.-D. und Schewe, K.-D. (Eds.), *Proceedings of the Fourth International Conference on Quality Software (QSIC 2004)*. pp. 224–231. Braunschweig, Germany. IEEE Computer Society Press.