

Enterprise Data Warehouse Management & Monitoring System (EDWMM)

by

**Umer Sajid
Usman Iftikhar**



Supervised by

Mrs. Iram Jamshaid

A report is submitted to the department of Computer Science & Computer Engineering, Bahria Institute of Management & Computer Science, Islamabad.
In partial fulfillment of requirement of the degree of Software Engineering.

**Department of Computer Sciences & Computer Engineering
Bahria Institute of Management and Computer Science, Islamabad
Bahria University Islamabad**

Acknowledgements

Thanks to almighty Allah, who enlightened and blessed us with this opportunity to accomplish this goal.

We owe debt of gratitude to our supervisor Mrs. Iram Jamshaid, whose valuable guidance and expert supervision has always been a source of inspiration for us. We are very thankful for her continuing interest, advices and tolerance at each and every step. Without her keen interest and constructive criticism, it would have been impossible to complete this final project.

We do not find words adequate to express our deepest gratitude to all those who willingly participated with us in this project. We would cordially thank all fellows for their constant help, providing every possible suggestion and keeping alive the spirit of work. Our parents and siblings who really supported us and remained the source of encouragement and motivation throughout our project.

Usman Iftikhar
Umer Sajid

Abstract

The success of an organization depends on its ability to acquire accurate and timely data about its operation, to manage this data effectively, and to use it to analyze its own activities. The need for powerful and flexible data management has grown much in all fields of life. This Final year project is an industrial-based project assigned by a data warehousing company “TeraData”. In this project main objective was to investigate research and design an Extract Transform Load (ETL) architecture which would segregate ETL process into smaller and more manageable steps.

Main Objectives of the project is to:

- Design, engineer and development of an end to end ETL process for a Telecommunication industry as follows:
 - Using TeraData as RDBMS for loading data during ETL process.
 - Designing physical Data Model for optimal Business Intelligence queries & data loading.
 - Engineering a complete ETL process and load source data into a logical data model by following data warehousing principles used during ETL process.
 - Capturing changes coming from sources on daily basis and merge them into already existing data warehouse. (Change Data Capture, (CDC))
- Developing GUI based Enterprise Data Warehouse Management & Monitoring System (EDWMM).
 - EDWMM will monitor all EDW loading jobs which trigger as part of ETL process.
 - EDWMM will control the loading of data based on static metadata registered in metadata repository inside Teradata.
 - EDWMM will register all data errors in error tables and generate email & SMS alerts for Database Administrators to take action.
 - EDWMM will be able to schedule ETL jobs on a fixed time & recurring basis.

The primary reason behind developing this project is the mutual interest of both the group members in data Warehousing. Secondly inspiration by the reputation and performance of TeraData. So, all we wanted was to get a data Warehousing project from TeraData which was more like a dream come true.

List of Figures

Figure 4.1:	System Design	23
Figure 4.2:	System Interface Description	25
Figure 4.3:	High level ETL architecture	27
Figure 4.4:	ETL steps	29
Figure 4.5:	EDWMM integration	30
Figure 4.6:	Error handling	33
Figure 4.7:	ETL process flow	40
Figure 4.8:	States of EDWMM GUI	42

List of Tables

Table 4.1:	ETL Transformation steps	28
Table 4.2:	Example Table 1	35
Table 4.3:	Example Table 2	36
Table 4.4:	Standard ETL jobs	41
Table 5.1:	User Login Test Case	46
Table 5.2:	Change Password Test Case	47
Table 5.3:	Start ETL jobs automatically Test Case	48
Table 5.4:	User Login Test Case data	50
Table 5.5:	Auto Jobs Processing Test Case data.....	53

Table of Contents

Contents	Page #
Acknowledgements	i
Abstract	ii
List of Figures	iii
Table of contents	iv
Chapter 1 INTRODUCTION.....	1
1.1 Scope of the project.....	2
1.2 Project documentation overview	3
1.2.1 Introduction.....	3
1.2.2 Requirements specification.....	3
1.2.3 Design Considerations.....	3
1.2.4 System Design.....	4
1.2.5 System Testing.....	4
1.2.6 Conclusion.....	4
1.2.7 Appendices.....	4
1.3 Major features of the project	5
1.4 Project Milestones.....	7
1.4.1 Inception phase.....	7
1.4.2 Elaboration phase.....	7
1.4.3 Construction phase.....	8
1.4.4 Transition Phase.....	8

Chapter 2 Requirements specification	10
2.1 Existing system	10
2.2 Proposed system	11
2.2.1 Product Overview	11
2.3 User requirements.....	13
2.4 Functional requirements.....	14
2.5 Non-functional requirements.....	15
Chapter 3 DESIGN CONSIDERATIONS	18
3.1 Design Constraints	18
3.1.1 Hardware and software environment	18
3.1.2 End user characteristics.....	19
3.2 Architectural Strategies.....	20
3.2.1 Software Process Model.....	20
3.2.2 Reuse of existing software components.....	20
3.2.3 Future enhancements/plans	20
Chapter 4 SYSTEM DESIGN	22
4.1 System design overview.....	22
4.2 System Architectural Design.....	24
4.2.1 Chosen System Architecture.....	24
4.2.2 Discussion of alternative design.....	24
4.2.3 System Interface Description	25
4.2.4 Sub modules.....	26
4.3 ETL Architecture	27
4.4 ETL transformation steps	28
4.4.1 General principles	28
4.4.2 Extract Step.....	31
4.4.3 Staging step.....	32
4.4.4 Error handling step.....	33
4.4.5 Change data capture step.....	34
4.4.6 Key conversion step.....	34

4.4.7	Transformation step.....	35
4.4.8	Apply step.....	38
4.4.9	Purge step.....	39
4.5	ETL jobs.....	40
4.6	Detailed EDWMM GUI.....	42
Chapter 5 System Testing.....		44
5.1	System Testing.....	44
5.1.1	Objectives	44
5.2	Testing.....	44
5.2.1	Unit Testing	44
5.2.2	Integrated Testing	44
5.2.3	Validation Testing.....	44
5.3	Configuration Review	48
5.4	Test cases.....	46
	Test Cases	49
Chapter 6 Conclusion.....		59
6.1	Future Enhancements.....	59
Appendix A		62
Appendix B	SAMPLE CODE.....	64
Appendix C	References.....	76