

APPROVAL

The CSE597 Seminar topics of Masters of Science course project Report title **“STUDY AND ANALYSIS OF HETEROGENIOUS WEB DATA INTEGRATION”** submitted by Rozana Alam, ID No:093 0365 050, to the Department of Electrical Engineering & computer Science of North South University, Dhaka , has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of masters of Science in Computer Science and Engineering approved as to its style and contents.

Board of Examiners

- | | |
|-----------------------|--------------|
| 1. Dr. Shazzad Hosain | (Supervisor) |
| 2. | (Supervisor) |
| 3. | (Supervisor) |
| 4. | (Supervisor) |

.....
Prof. Dr. M Abdul Awal

Professor & chairman

Electrical Engineering & computer Science Department

North South University, Dhaka-1229, Bangladesh.

LETTER FROM CANDIDATE

To

Prof. Dr. M Abdul Awal

Professor & chairman

Electrical Engineering & computer Science Department

North South University, Dhaka-1229, Bangladesh.

Dear Sir,

I am completely aware of the fact that plagiarism is a discipline act and Department of Electrical engineering and Computer Science has stern policies for plagiarism. I hereby announce that the theories ,model, results, figures tables hence all the works presented in this conjecture , are done by me and I have cited in a proper way whenever I have referred to or made use of other's works.

.....

Rozana Alam

ID:093 036 5050

DECLARATION

I, hereby, declare that the work presented in this Project is the outcomes of the investigation performed by me under the supervision of Dr. Shazzad Hosain , Assistant Professor, North South University , Dhaka-1229, Bangladesh . I also declare that this Thesis or no part thereof has been or is being submitted elsewhere for the award of any degree or diploma.

Countersigned

Signature

.....

(Dr. Shazzad Hosain)

Supervisor

.....

(Rozana Alam)

Candidate

ABSTRACT

From the beginning of data processing systems through the middle of the 80's decade, integration between systems was always a custom, platform-specific affair. A continual reliance on information systems with key databases combined with an exponential growth of data. Heterogeneous Data is data from any number of sources, largely unknown and unlimited, and in many varying formats. Data integration can be defined as "the use of common field definitions and codes across different parts of an organization". Data integration is a set of processes used to extract or capture, restructure, move, and load or publish data, in either operational or analytic data stores, in either real time or in batch mode. Because of the operational and analytic nature of integrating data, the frequency and throughput of the data have developed into different types of data integration architectural patterns and technologies. Therefore, an investigation of the architectural types or "patterns" of data integration have been shown in this report. Data integration can be classified in two types direct access and data access via mediator. In this paper I tried to show different kinds of data integration where we use an example database software "USHR system" for show up the both kinds of data integration i.e direct access and mediator based data integration. Here data integrations in the database software "USHR system" is explained with flowchart and architectural diagram . Two different database is studentdb(db1) and bracias(db2) where we applied the direct access method and data integration via mediator method for integrate data individually. Then we get the resulted form "Faculty Salary Form after integrating db1 and db2" by using the direct access method and "data migrated form" for data integration via mediator method.

ACKNOWLEDGMENTS

First of all I would like to thank Almighty Allah. Today I am successful in completing my Project work with such ease because He gave me the necessary ability, chance and a cooperating supervisor.

It is my pleasure to take this opportunity to thank a few people who assisted, encouraged, directed and supported me throughout my internship program.

Firstly, I would like to thank my supervisor, Dr. Shazzad Hosain, without whose patient guidance and persistent support, I could not have finished my project work. He gave me both the motivations for starting new research topics and the freedom of working on my research interests. Not only has he taught me the systematic way to create database, but also he has given numerous suggestions to improve my English writing skills. All that I learnt from him definitely will benefit me in my future career.

Special thanks are given to Professor Dr. Abdul Awal, Chairman of Department of Electrical and Computer Science (NSU) and one of my honorable teachers Mr. Ahsan Raja Choudhury, Assistant professor, University of Dhaka, who were the supervisors of my Bachelor course project for their support and continuous encouragement.

Finally, I would like to thank my family members for inspiration, support and encouragement. I would like to take this opportunity to thank all of my good friends and well wishers for their encouragement.

Dedicated to my newly born nephew
Md. Joheb Alam.

TABLE OF CONTENTS

APROVAL	i
LETTER FROM CANDIDATE	ii
DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDEGEMENTS	v
TABLE OF CONTENTS	vii
LIST OF FIGURES	x
LIST OF TABLE	xi

Chapter I:	Introduction.....	1
	1.1 Overview.....	1
	1.2 What is Data Integration?.....	2
	1.3 Benefits of integrating heterogeneous data sources.....	3
	1.4 Integrating diverse data source paradigms.....	3
	1.5 Different kinds Data integration.....	4
	1.5.1 Redundancy-free solutions.....	5
	1.5.1.1 Direct Access.....	5
	1.5.1.2 Data Access via Mediator.....	6

	1.6 Overview of the book.....	7
Chapter II:	Literature review	8
	2.1 Data integration by using workflow	8
	2.2 Data integration via mediator	9
Chapter III:	Data Integration by Using Direct Access.....	13
	3.1 Workflow data integration.....	13
	3.2 Architectural diagram of data integration for USHR system	14
	3.3 Flowchart for “USHR system”.....	16
	3.4	17
	Activities.....	
	3.5 Activity inputs, outputs and blocks	18
	3.6	19
	Activities and resources.....	
	3.7	20
	Workflow execution	
	3.8	21
	Workflow execution types.....	
	3.9	22
	Key features of workflow.....	
	3.10 UML model for “USHR system”.....	22
	3.11.1 “USHR system” UI forms from dbstudent(db1) database.....	25
	3.11.2 “USHR system”, forms from HR database form(db2)	28
	3.11.3 Data integration form for db1 and db2.....	31
Chapter IV:	Data Integration via Mediator.....	32
	4.1 Mediator System.....	32

4.2	Source Query Condition.....	33
3.4	The Domain Predicate.....	33
4.4	Architectural diagram for the Mediator System for db1 and db2.....	34
4.5	Result for the source query mapping algorithm.....	35
Chapter V:	Conclusion.....	37
5.1	Required software.....	37
5.2	Future work	37
5.3	Conclusion.....	37

BIBLIOGRAPHY

LIST OF FIGURES

1.1	Types of data integration.....	5
3.1	Data integration using workflow in USHR system.....	15
3.2	(i) and (ii) Data integration using single workflow and server in USHRsystem.....	17
3.3	Activity inputs and outputs.....	18
3.4	Activity inputs, input literals and outputs.....	19
3.5	Activities and resources.....	20
3.6	Executing a workflow.....	21
3.7	UML model for dbstudent (db1).....	23
3.8	UML model for HR department(bracias database)	23
3.9	The database for “USHR system” with integrating db1 and db2.....	24
3.10	University Database form(db1).....	25
3.11	Add Student information form(db1).....	26
3.12	Student information search form(db1).....	26
3.13	Student’s course add form(db1).....	27
3.14	Faculty details form add form(db1).....	27
3.15	Student’s result add form(db1).....	28
3.16	Student’s result (db1).....	28
3.17	login form from db2.....	29
3.18	ABC integrated accounting information system from db2.....	29
3.19	Employee Entry Form from db2.....	30
3.20	Salary Detail Entry Form from db2.....	30
3.21	Faculty Salary Form after integrating db1 and db2.....	31
4.1	Architectural diagram for Data integration via mediator	34
4.2	Data integration via Mediator form.....	36
4.3	Data migrated form.....	36

LIST OF TABLE

Table1.1	Usage scenarios, advantages and disadvantages of the direct data integration pattern.	6
Table1.2	Usage scenarios, advantages and disadvantages of data integration via mediator.	7

