UNIVERSITI PUTRA MALAYSIA

DATA REPLICATION WITH 2D MESH PROTOCOL FOR DATA GRID

ROHAYA BINTI HJ LATIP

FSKTM 2009 1

DATA REPLICATION WITH 2D MESH PROTOCOL FOR DATA GRID

By

ROHAYA BINTI HJ LATIP

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

July 2009

DEDICATIONS

To my beloved ones,

Abah and mak

My hubby, Anuar

My kids, Aisyah and Aiman

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

DATA REPLICATION WITH 2D MESH PROTOCOL FOR DATA GRID

By

ROHAYA BINTI LATIP

May 2009

Chairman: Associate Professor Hamidah Ibrahim, PhD.

Faculty: Computer Science and Information Technology

Data replication is one of the widely approach to achieve high data availability and fault tolerant of a system. Data replication in a large scale distributed and dynamic network such as grid has effects the efficiency of data accessing and data consistency. Therefore a mechanism that can maintain the consistency of the data and provide high data availability is needed. This thesis discusses protocols and strategies of replicating data in distributed database and grid environment where network and users are dynamic. There are few protocols that have been implemented in distributed database and grid computing which is discussed such as Read One-Write All (ROWA), Voting (VT), Tree Quorum (TQ), Grid Configuration (GC), Three Dimensional Grid Structure (TDGS), Diagonal Replication in Grid (DRG) and Neighbor Replication in Grid (NRG).

In this thesis, we introduce an enhanced replica control protocol, named Enhance Diagonal Replication 2D Mesh (EDR2M) protocol for grid environment and compares its

result of availability, and communication cost with the latest protocol TDGS (2001) and NRG (2007). EDR2M proves data consistency by fulfilling the Quorum Intersection Properties. Evaluations that is suitable and applicability for EDR2M protocol solutions via analytical models and simulations.

A simulation of EDR2M protocol is developed and the performance metrics evaluated are data availability, and communication cost. By getting the sufficient number of quorum, number of nodes in each quorum, and selecting the middle node of the diagonal sites to have the copy of the data file have improved the availability and communication cost for read and write operation compared to the latest protocol, TDGS (2001) and NRG (2007). Thus, the experiment has showed scientifically that EDR2M is the adequate protocol to achieve high data availability in a low communication cost by providing replica control protocol for a dynamic network such as grid environment.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

REPLIKASI DATA DENGAN PROTOCOL JEJARING 2D UNTUK DATA GRID

Oleh

ROHAYA BINTI LATIP

Mei 2009

Pengerusi: **Profesor Madya Hamidah Ibrahim, PhD.**

Fakulti: **Sains Komputer dan Teknologi Maklumat**

Replikasi data adalah satu daripada pendekatan yang digunakan secara meluas bagi mencapai tahap ketersediaan data yang tinggi dan mengekalkan tahap toleransi kesalahan bagi sesebuah sistem. Replikasi data di dalam rangkaian teragih berskalar besar dan dinamik seperti grid telah memberi kesan ke atas kecekapan capaian data dan konsistensi data. Oleh yang demikian, satu mekanisma yang mampu mengekalkan konsistensi dan menyediakan ketersediaan data yang tinggi adalah diperlukan. Tesis ini membincangkan protokol and strategi replikasi data di dalam pangkalan data teragih dan persekitaran grid di mana rangkaian dan penggunanya yang dinamik. Terdapat beberapa protokol yang telah dilaksanakan di dalam pangkalan data teragih dan perkomputeran grid yang dibincangkan seperti Read-One Write-All (ROWA), Voting (VT), Tree Quorum (TQ), Grid Configuration (GC), Three Dimensional Grid Structure (TDGS), Diagonal Replication in Grid (DRG), dan Neighbor Replication in Grid (NRG).

Di dalam tesis ini, kami memperkenalkan satu protokol kawalan replica yang baru yang dinamakan protokol *Enhance Diagonal Replication 2D Mesh* (EDR2M) untuk persekitaran grid dan hasilnya dibandingkan terhadap ketersediaan data, dan kos komunikasi dengan protokol terkini TDGS (2001) dan NRG (2007). EDR2M protocol menyediakan pengukuhan data dengan memenuhi Properti Persilangan Korum. Penilaian yang bersesuaian dan berkaitan untuk penyelesaian protokol EDR2M melalui model analitikal dan simulasi.

Satu simulasi protokol EDR2M dibangunkan dan matrik-matrik prestasi yang dinilai adalah ketersediaan data, dan kos komunikasi. Dengan memperolehi bilangan korum dan bilangan nod yang sesuai di dalam setiap korum, serta memilih nod di tengah-tengah lokasi pepenjuru untuk mempunyai fail data sebagai salinan telah berjaya mempertingkatkan tahap ketersediaan data dan mengurangkan kos komunikasi terhadap operasi membaca dan menulis data banding dengan protokol-protokol yang lalu, TDGS (2001) dan NRG (2007). Oleh yang demikian, kajian ini secara saintifiknya telah menunjukkan EDR2M adalah protokol yang telah berjaya memperolehi tahap ketersediaan data yang tinggi di samping kos komunikasi yang rendah dengan menyediakan protocol kawalan replika untuk rangkaian yang dinamik seperti persekitaran grid.

ACKNOWLEDGEMENTS

Miracles are great, but they are so unpredictable..

therefore I am very thankful to Allah for giving me these miracles of strength, inspirations and patience to complete this thesis.

My deepest thank also goes to Assoc. Prof. Dr. Hamidah Ibrahim for her efficient guidance and trust as I floundered my way through this process. I am also grateful to the supervisory committee, Assoc. Prof. Dr. Mohamed Othman, Assoc. Prof. Dr Md Nasir Sulaiman and Mr. Azizol Abdullah for their fruitful discussions and valuable suggestions. Thanks to all those friends for all their generous input, constructive criticisms and laughter are in here somewhere.

My final words go to my family. In this type of work the relatives are always mistreated. I must therefore thanks my husband, Hairul Anuar and my kids, Aisyah and Aiman for putting up with my late hours, my spoiled weekends but above all for putting up with me and surviving the ordeal. Thanks to my parents, Hj Latip Bin Hj Sharif and Hajah Hamidah Binti Alimun. You started all this: now I have to finish it. Thank you for your unending *doa*, willingness to accept and eagerness to love.

A great thanks to all…

APPROVAL

I certify that an Examination Committee has met on 18 May 2009 to conduct the final examination of Rohaya Binti Latip on her Doctor of Philosophy thesis entitles " Data Replication with 2D Mesh Protocol for Data Grid" in accordance with Universities and University Colleges Act 1971 and Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the candidate be awarded the Doctor of Philosophy.

Members of the Examination Committee are as follows:

Mohd Hasan Selamat
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Abdul Azim Abd Ghani, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Shamala a/p K. Subramaniam, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Jemal Abawajy, PhD
Professor
Faculty of Science and Technology
School of Engineering and Information Technology
Deakin University
3072 Geelong, Australia
(External Examiner)

_____

BUJANG KIM HUAT, PhD
Professor / Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 13 July 2009

This thesis is submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Hamidah Ibrahim, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Mohamed Othman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Md. Nasir Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Azizol Abdullah
Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

_____
HASANAH MOHD. GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 17 July 2009

DECLARATION

I declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

_____

ROHAYA BINTI HJ LATIP

Date: 17 July 2009

TABLE OF CONTENTS

		Page
DEDICATION		ii
ABSTRACT		iii
ABSTRAK		v
ACKNOWLEDGEMENTS		VII
APPROVAL		viii
DECLARATION		x
LIST OF TABLES		xiii
LIST OF FIGURES		xiv
LIST OF ABBREVIATIONS		xvii
CHAPTER		
1	**INTRODUCTION**	
	1.1 Background	1
	1.2 Problem Statement	3
	1.3 Objectives	6
	1.4 Scope	6
	1.5 Organization of Thesis	7
2	**LITERATURE REVIEW**	
	2.1 Grid Computing	9
	2.2 Communication Cost	13
	2.3 Data Availability	13
	2.4 Data Consistency	14
	2.5 Quorum	15
	2.6 Replica Control Protocol	15
	2.6.1 Read One-Write All (ROWA)	16
	2.6.2 Voting Protocol (VT)	18
	2.6.3 Tree Quorum (TQ)	25
	2.6.4 Grid Structure (GS)	

		29
	2.6.5 Three Dimensional Grid Structure (TDGS)	32
	2.6.6 Diagonal Replication on Grid (DRG)	36
	2.6.7 Neighbor Replication on Grid (NRG)	38
	2.7 Simulation Modeling and Analysis	41
	2.8 Summary	45
3	**RESEARCH METHODOLOGY**	
	3.1 Discrete Event Simulation	48
	3.2 Define study	49
	3.2.1 Assumptions	49
	3.2.2. Input Variables	50
	3.2.3 Output Variables	51
	3.2.4 Model Topology	51
	3.2.5 System Analysis	52
	3.3 Initialization of Variables	53
	3.4 Development of EDR2M Protocol	54
	3.4.1 Dynamic Number of Replicas	58
	3.4.2 Updating in EDR2M Protocol	60
	3.5 Components of the Simulation	61
	3.5.1 Initialization Routine	62
	3.5.2 Event Scheduler Routine	62
	3.5.3 Event Routines	63
	3.5.4 Report Generator Routine	64
	3.6 Interface of the Simulation	64
	3.7 Simulation Validation	76
	3.8 Evaluate, Interpret, and Experiment	77
	3.9 Summary	78
4	**REPLICATION PROTOCOL IN 2D MESH**	79
	4.1 Quorum Intersection Property	79
	4.2 Framework of the Protocol	80
	4.2.1 Diagonal Replication in 2D Mesh (DR2M)	82
	4.2.2 Enhanced Diagonal Replication in 2D Mesh Protocol (EDR2M)	85
	4.3 Dynamic Network	92
	4.3.1 New Members of Grid Network	92
	4.3.2. Disconnection in the Network	97
	4.4 Updating the Primary Databases	97
	4.5 Enhance Quorum Algorithm by Creating Groups	100
	4.6 Balancing the Number of Nodes in the Group	102
	4.7 Shortest Path	104
	4.8 Summary	107
5	**RESULTS AND DISCUSSIONS**	

5.1 Communication Cost 108
5.2 Data Availability 110
5.3 Summary 114

6 CONCLUSION AND FUTURE WORKS
6.1 Conclusions 115
6.2 Contributions 116
6.3 Recommendations for Future Works 118

REFERENCES 120

BIODATA OF STUDENT 126

LIST OF PUBLICATIONS/AWARDS 127

LIST OF TABLES

Table		Page
2.1	Summarized Existing Data Replica Control Protocol	46
3.1	Simulation Parameters	54
3.2	System Requirements	54
3.3	Parameters Initialization	62
3.4	Events Definition	63
4.1	Updating Quorums	99

LIST OF FIGURES

Figure		Page
2.1	Optimization Service Goal	16
2.2	Tree Quorum Organization of Nine Copies of Data	26
2.3	An Example of Write Quorum in TQ	27
2.4	A Tree Organization of 13 Copies of a Data Object	28
2.5	A Grid Structure of $n = 25$	30
2.6	Eight Copies of a Data Object	32
2.7	Example of TDGS Organization of 24 Nodes	35
2.8	25 Copies of a Data Object in DRG	36
2.9	Diagonal Sets of $D_2(s)$ and $D_3(s)$	37
2.10	A Grid Organization of Nine Copies of a Data	39
2.11	Performance Analysis Techniques	43
3.1	Discrete Event Simulation Lifecycle	49
3.2	A Grid Organization of Nine Nodes	50
3.3	Example of Quorums and Groups Organized in 2D Mesh	52
3.4	Flow Chart of the Simulator	56
3.5	Empty Nodes are Located at Nodes 2352 and 2401 Logically	58
3.6	New Virtual Columns and Rows Added in the Simulation	59
3.7	Procedure for Adding New Virtual Column and Row	60
3.8	Procedure of Write and Read Operations	61

3.9	Output Screen for 25 Nodes	66
3.10	Output Screen for 25 Nodes	67
3.11	Output Screen for 49 Nodes	68
3.12	Output Screen for 49 Nodes	69
3.13	Output Screen for 225 Nodes	70
3.14	Output Screen for 225 Nodes	71
3.15	Output Screen for 225 Nodes	72
3.16	Output Screen for 729 Nodes	73
3.17	Output Screen for 729 Nodes	74
4.1	Framework of the Protocol	81
4.2	Four Quorums Obtain in a 9 * 9 Network Size	83
4.3	Flow Chart of DR2M	85
4.4	Procedure for Finding the Number of Quorums in the Network	87
4.5	Venn Diagram of Read Quorum, q_R and Write Quorum, q_W	88
4.6	A Grid Organization with 81 Nodes	91
4.7	Condition to Add New Column and Row	93
4.8	Number of Nodes at Time t_1	93
4.9	Adding 2 Virtual Columns and Rows	94
4.10	Flowchart for Dynamic Network Members	96
4.11	Conditions of Disconnected Nodes	97
4.12	Graph Showing the Optimal Number of Nodes for a Quorum	100

4.13	Groups of Quorums Formed	101
4.14	Restructuring of Quorum for Balancing the Number of Nodes	103
4.15	Algorithm for Balancing the Nodes in Each Quorum	104
4.16	Links for Node Five	106
4.17	Shortest Path to the Primary Database	106
5.1	Communication Cost for Read Operation	109
5.2	Communication Cost for Write Operation	110
5.3	Selected Primary Database	111
5.4	Read Availability for 81 Nodes	112
5.5	Write Availability for 81 Nodes	113
5.6	Availability Results for Number of Nodes 49, 81, and 121Nodes	114

LIST OF ABBREVIATIONS

API	Application Programming Interface
CPU	Computer Processor Unit
DBMS	Database Management System
NRG	Neighbor Replication On Grid
ROWA	Read One Write All
TDGS	Three Dimensional Grid Structure
TQ	Tree Quorum
VT	Voting Technique

CHAPTER 1

INTRODUCTION

This chapter forms the introduction to the thesis. Discussions begin with emphasis on data management and problems of data replication in dynamic grid environment that consists of large number of nodes.

1.1 Background

The emerging technologies of computer network and database make distributed database ~~important~~ easily shared and accessed the ~~therefore~~ data across the network ~~can be shared and easily accessed~~. Grid serves the facilities by emerging computational and networking infrastructure as well as is specifically designed to support pervasive and reliable access to data and computational resources over wide area network and across organizational domains (Lamehamedi et al., 2002).

To allow data to be more accessible especially in large network like grid, ~~a lot of techniques is implemented, among which is data replication. D~~ data replication is an effective measure to access data in a geographically distributed environment (Venugopal et al., 2005) because identical copies of data are generated and stored at various globally distributed sites. Significantly this has reduced the data access latencies (Guy et al., 2002). Data replication is also used in distributed system to increase data availability and to achieve fault tolerance (Lamehamedi et al., 2002).

Formatted: Not Different first page header

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Data replication is performed by copying data to and from sites in order to ~~improve local service response time and~~ increase availability as well as to optimize data access. However, dealing with replicas of files adds a number of issues as opposed to a single file instance. One, replicas must be kept consistent and up to date; two, their location must be catalogued; three, their lifetime needs to be managed. Replica catalogues, however, only provide users with information about the location of the replica file, but not in selecting the replica with minimum access time.

Formatted: Font color: Auto

Data replications are frequently employed as part of backup and recovery strategy. To increase the availability, strategies of replicating data should include considerations to improve the performance of data grid. In principle, there are mainly two different replication approaches, which are synchronous and asynchronous replication. Synchronous replication aims at keeping all the replications permanently in sync, whereas asynchronous replication allows for a certain delay in updating replicas. Based on the relative slow performance of writing operations in synchronously replicated environment (Gray et al., 1996), the database research community is searching for effective protocols for asynchronous replication by tolerating low consistency (Stockinger et al., 2001). However, complex and expensive synchronization mechanisms are needed to maintain data consistency and data integrities (Agrawal and El Abbadi, 1991; Gifford, 1979; Stonebraker, 1979; Bernstein and Goodman, 1984; Davĉev and Burkhard, 1985; El Abbadi et al., 1985; Paris and Long, 1988; Jajodia and Mutchler, 1990).

~~Recently, s~~Several researchers have proposed to impose a logical structure on the set of copies in database, and to use structural information to create intersecting quorums. Protocols that use a logical structure, e.g., the ROWA protocol (Ozsu and Valduriez, 1996; ~~Jain, 1991; ~~Ozsu, 1999), Voting protocol (VT) (Mat Deris, 2001), Tree Quorum protocol (TQ) (Agrawal and El Abbadi, 1990; Chung, 1993; Maekawa, 1992; Mat Deris et al., 2004), Grid Configuration (GC) protocol (Mat Deris et al., 2004; Agrawal and El Abbadi, 1996), Three Dimensional Grid Structure (TDGS) protocol (Mat Deris et al., 2004), and Neighbor Replication on Grid (NRG) protocol (Ahmad et al., 2007), execute operations with low communication cost while still providing fault-tolerance for both read and write operations. However, as the approaches improve performance for read availability results, they also degrade the write availability results especially when failures occur.

Since data access is one of the ~~big~~ issues, ~~most of the~~ researches~~'~~ such as Foster and Kesselman (2004); Lamehamedi and Szymanski (2007); Ranganathan and Foster (2004) implement protocol for~~ation was for~~ read only data ~~(Foster and Kesselman, 2004; Lamehamedi and Szymanski, 2007; Ranganathan and Foster, 2004).~~ and they do not focus on data consistency during updates. ~~Following this review~~Therefore, this thesis focuses on data availability and data consistency in a low communication cost by proposing the extension of the current protocols to improve the fault-tolerance of write operations through the notions of read and write quorums with respect to the 2D Mesh logical structure.

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

1.2 Problem Statement

A data grid is mainly focused to provide users with infrastructure that enables and facilitates reliable access and sharing of data management resources. Data access and transfer service should also be able to scale across the widely distributed locations and reach wide area networks in order to break administrative and geographical barrier (Lamehamedi and Szymanski, 2007; Ranganathan and Foster, 2001). By having multiple copies of data located at different sites has affected the consistency of data. ~~This new generation of universal cooperation and collaborative applications~~ ~~Therefore require a replica control new protocol is required to ensure efficient access to consistent data and sufficient replica control protocol in a low communication cost with high data availability.~~

Formatted: Font color: Black

~~In grid environment, data are being replicated at different sites and the connections of grid members are dynamic, whereby they keep on connecting and disconnecting to the grid network (Akhbarinia and Martins, 2007). This has increase the performance of response time because the nearest primary database must be thoroughly searched once the data needs to be replicated.~~

~~Another challenge in replicated environment is d~~Data consistency~~, whereby~~ maintains~~ing~~ data integrity and consistency in a replicated environment is of

prime importance. High precision applications may require strict consistency updates for every transactions. This implies that the performance has been severely affected every time a new replica is created. Furthermore if a file being replicated at more than one site, it occupies larger storage space, thus has to be administered and has caused overheads in storing multiple files (Tang et al., 2004).

Another challenge is the low write performance in applications that required high updates in replicated environment since transactions may need to be updated in multiple copies (Venugopal et al., 2005).

Formatted: Font color: Black

Formatted: Font color: Red

~~Currently TDGS (Mat Deris et al., 2008) has overcome the data availability and data consistency for distributed database with a low communication cost but the performance of data availability and communication cost could be improved for both read and write operations. However, w~~When the network size grows larger such as in grid environment, ~~The~~the primary database to be access will be further apart based on the previous protocol. Therefore this has decrease the performance of data availability ~~has decreased~~ and increase communication cost ~~has increase when the network size grows larger such as in grid environment.~~

.

~~NRG (Ahmad et al., 2007; Ahmad et al., 2008) also has focused on enhancing the performance of data availability in a low communication cost for grid environment but~~ The performance of communication cost has

increase because of ~~the performance of data availability is low and communication cost is still~~the increment number of quorum in the network ~~high~~ such as having replicated data at all nodes because all nodes is a member of its neighbor. By having replicated data at each node which is a member to another node will increase the communication cost. ~~The protocol could be enhanced to get better results.~~