

SCALABLE IN-MEMORY DATA MANAGEMENT MODEL FOR ENTERPRISE APPLICATIONS

Anupama Piyumali Pathirage

(138223D)



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
Degree of Master of Science
www.lib.mrt.ac.lk

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

March 2015

SCALABLE IN-MEMORY DATA MANAGEMENT MODEL FOR ENTERPRISE APPLICATIONS

Anupama Piyumali Pathirage

(138223D)



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Thesis submitted in partial fulfilment of the requirements for the Master of Science in
Computer Science.

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

March 2015

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:.....

Name: Anupama Piyumali Pathirage



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

I certify that the declaration above by the candidate is true to the best of my knowledge and that this report is acceptable for evaluation for the Post Graduate Project.

Signature:

Date:.....

Name: Dr. Amal Shehan Perera

ABSTRACT

Project Title: Scalable In-Memory Data Management Model for Enterprise Applications

Authors: Pathirage A.P

Supervisor/s: Dr.Shehan Perera (Supervisor)

Dr.Malaka Walpola (Coordinator)

With the rapid advances in technology and data volume, having efficient and scalable data management system is essential for most of the enterprise applications. So In-Memory data management systems are becoming the highly used data management solution in most of the time critical enterprise solutions. Although In Memory Data Management Systems are widely used, still they are having problems such as scalability issues, concurrency problems etc. This project is an effort that aims to propose a scalable enterprise solution for in memory data management, identifying the bottlenecks in the current In-Memory Data management systems.

Although there are various benchmarks are available for Disk Resident Databases, lack of a fair metric for comparing the performance of different in-memory database systems has become a problem when selecting the appropriate data management system for enterprise applications. Currently there are various in-memory databases are available and when using them with the enterprise applications, developers have to put lot of effort as there is no standard API/Interfaces available for them.

This research project addresses these two problems by providing an unbiased performance benchmark for various in-memory databases and developing a data connector framework to access different data sources such as in-memory databases, disk resident databases, flat file data bases and in-memory data caches.

This report provides details about the problem background, existing system implementations and current research areas in this domain and how I'm going to achieve the objective.

Keywords: In-Memory Database, In-Memory Data Grid, Disk Resident Database, Data Access Layer, Database Benchmarking

ACKNOWLEDGEMENT

I would like to thank Dr. Shehan Perera, my supervisor, for his invaluable support, assistance and advices given throughout this project. His expertise and continuous guidance enabled me to complete my work successfully and his help in moderating the content was invaluable. I would also like to thank Dr. Malaka Walpola, the project co-ordinator, for his continuous support and feedback on the structure of the project which motivates me to do my best.

Further I would like to thank all my colleagues for their help on finding relevant research materials, sharing knowledge and experience and for their encouragement. My sincere appreciation goes to my husband and parents for the continuous support and motivation given to me to make this thesis a success.

Finally I would like to thank all my colleagues at DirectFN, who helped me to enhance my knowledge and for the support given to me to manage my MSc research work.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

TABLE OF CONTENT

Declaration	I
Abstract	II
Acknowledgement	III
Table of Content	IV
Table of Figures	VI
Table of Tables	VII
List of Abbreviations	VIII
1. Introduction	2
1.1 Problem Background	2
1.2 In-Memory Data Management Systems	4
1.2.1 In-Memory Databases (IMDB).....	5
1.2.2 In-Memory Data Grid (IMDG).....	6
1.3 Limitations of Existing Solutions	8
1.4 Motivation	8
1.5 Objectives.....	9
2. Literature Review	10
2.1 Disk Resident Databases vs. IMDS	10
2.2 In-Memory Database Architecture	11
2.2.1 Impact of Memory Residency on IMDB functionality	14
2.3 Application of Main Memory Databases	18
2.3.1 IMDB for Embedded Systems	18
2.3.2 IMDB for Enterprise Applications.....	19
2.4 Performance Benchmarks for In-Memory Database.....	21
2.4.1 Wisconsin Benchmark	21
2.4.2 TimesTen Performance Throughput Benchmark (TPTBM).....	23
2.4.3 Telecom Application Transaction Processing Benchmark(TATP).....	24
2.4.4 Transaction Processing Performance Council -C Benchmark(TPCC).....	26
2.5 Cloud based In-Memory Databases	27
3. Benchmarking Methodology	29
3.1 Analysis of Comparison and Evaluation Scenarios	29
3.1.1 Overview of Selected IMDB.....	30
3.1.2 Overview of Selected DRDB	34
3.1.3 Overview of In-Memory Data Caches	35

3.1.4	Overview of Flat File Database.....	36
3.1.5	Feature Comparison of Selected Database.....	37
3.2	Analysis of Benchmark Criteria	37
3.2.1	Benchmark Design.....	39
3.2.1.1	System Configuration	39
3.2.1.2	Test Data	39
3.2.1.3	Benchmark Workload and Experimental Design.....	40
3.2.2	Benchmark Execution	41
3.2.3	Benchmark Analysis	42
3.3	Results	42
3.3.1	Results for Insert Operation	42
3.3.2	Results for Select Operation.....	45
3.3.3	Results for Update Operation.....	48
3.3.4	Results for Delete Operation.....	50
4.	Framework Implementation	53
4.1	Problem Background	53
4.2	Design of the Framework	54
4.3	Implementation Details.....	56
4.3.1	Implementation of Flat File based DB.....	56
4.3.2	Implementation of In-Memory Cache.....	58
4.3.3	Implementation of the Framework for Data.....	61
4.4	Performance Analysis of Framework	65
5.	Conclusion And Future Work	67
5.1	Conclusion.....	67
5.2	Future Work.....	69
6.	References.....	70

TABLE OF FIGURES

Figure 1 : Moore's Law for Disk Speed	3
Figure 2 : In Memory Data Management System.....	5
Figure 3 : IMDG Architecture	7
Figure 4 : Disk Resident Databases vs. IMDS.....	11
Figure 5 : IMDB Architecture.....	13
Figure 6 : Usage of IMDB	18
Figure 7 : Enterprise Performance In Memory Cycle.....	20
Figure 8 : Times Ten Benchmark Throughput update (100% Updates).....	24
Figure 9 : TATP benchmark on transaction processing time	25
Figure 10 : SQLite Architecture	30
Figure 11 : MemSQL Architectue	33
Figure 12 : Elements of Oracle	34
Figure 13 : Database System Benchmark Methodology.....	38
Figure 14: Example Insert Statement.....	42
Figure 15 : Insert Operation -Run Time Comparison.....	43
Figure 16 : Insert Operation - Transactions per Second Comparison.....	43
Figure 17 : Insert Operation - Concurrent Connections vs TPS	44
Figure 18 : Example Select Statement	45
Figure 19 : Select Operation - Run Time Comparison.....	46
Figure 20 : Select Operation - Transactions Per Second Comparison	46
Figure 21: Select Operation - Concurrent Connections vs TPS	47
Figure 22 : Select with Joins - TPS Comparison.....	47
Figure 23 : Example Update Statement	48
Figure 24 : Update Operation - Run Time Comparison	48
Figure 25 : Update Operation - Transactions Per Second Comparison.....	49
Figure 26 : Update Operation - Concurrent Connections vs TPS.....	49
Figure 27 : Example Delete Operation	50
Figure 28: Delete Operation -Run Time Comparison.....	50
Figure 29 : Delete Operation - Transactions Per Second Comparison	51
Figure 30 : Delete Operation - Concurrent Connections vs TPS.....	51
Figure 31 : Proposed Architecture for Database API	55
Figure 32: Database organization in Flat File DB	56
Figure 33 : Query Execution Method of Flat File DB	57
Figure 34 : Flat File DB - Table Data	58
Figure 35 : Example usage of In-Memory Cache.....	59
Figure 36: Class Diagram of In-Memory Cache.....	60
Figure 37 : Class diagram of Data Connection Framework	62
Figure 38 : ExecuteQuery Method for SQLite DB	64
Figure 39 : Example usage of Framework.....	64
Figure 40 : Insert Operation Performance of Framework - With Oracle.....	65
Figure 41 : Select Operation Performance of Framework - With SQLite	66
Figure 42 : Select Operation Performance of Framework	66

TABLE OF TABLES

Table 1 : CSQL Wisconsin Benchmark Results	23
Table 2 : Feature Comparison of Selected Databases.....	37
Table 3 : Benchmark System Configurations	39
Table 4 : Database Table Data	40
Table 5 : Performance Metrics.....	41
Table 6 : Benchmark Tool Implementation Details.....	41



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

LIST OF ABBREVIATIONS

Abbreviation	Description
ACID	Atomicity, Consistency, Isolation, Durability
ANSI	American National Standards Institute
API	Application Programming Interface
CDC	Change Data Capture
CPU	Central Processing Unit
CRUD	Create, Read, Update, and Delete
CSV	Comma Separated Values
DAL	Data Access Layer
DBA	Database Administrator
DML	Data Manipulation Language
DRDB	Disk Resident Database
IMDB	In-Memory Database
IMDG	In-Memory Data Grid
IMDS	In-Memory Data Management System
IPC	Inter Process Communication
JIT	Just In Time
JDBC	Java Database Connectivity
ODBC	Open Database Connectivity
MMDB	Main Memory Database
MVCC	Multi Version Concurrency Control
RAM	Random Access Memory
RDBMS	Relational Database Management System
RTOS	Real Time Operating System
SQL	Structured Query Language
STL	Standard Template Library
TPS	Transactions per Second



University of Moratuwa, Sri Lanka.
 Electronic Theses & Dissertations
 www.lib.mrt.ac.lk