SCALABLE IN-MEMORY DATA MANAGEMENT MODEL FOR ENTERPRISE APPLICATIONS

Anupama Piyumali Pathirage

(138223D)

Degree of Master of Science

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)

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

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.

TABLE OF CONTENT

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ABBREVATIONS

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