A Data Structure for High Performance Object Oriented Databases

by

Mudasser Fraz Wyne

A thesis submitted to the Faculty of Science
of The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Computer Science
Faculty of Science
The University of Birmingham
United Kingdom
February 1996

information by using extra bits at the storage level with different bit combinations being set for various lock modes. These structures can form the components of a model of a high performance architecture for OODBS.

The work presented here is assessed by both analytical and empirical study of the mT-Tree and it concludes that mT-Tree occupies less space and accesses information faster. To support these theoretical results, the mT-Tree and (for comparison) hcC-Tree have been implemented and tested.

1.2 Scope of the Thesis

In this thesis we are presenting a data structure for high performance OODB. This structure can be used as a basis for a high performance architecture model (MOOD) for OODBMS. Many related issues of the MOOD architecture need to be studied and discussed in greater depth. The components of the architecture which need more work include checkpointing [88], recovery [16, 49, 69], query processing [59, 60, 86, 96] and garbage collection [8, 56]. We have discussed these issues in general but these are the subject of much more research and will complement the contributions made by the thesis.

1.3 Summary of the Thesis

This thesis consists of three sections. Chapter 2 and 3 review object oriented database management systems and their architecture components, and give an introduction to main memory database system architecture. The second section (chapters

4, 5, 6 and 7) covers the proposed high performance architecture and data structure for the object oriented database system. The third section includes the performance evaluation for the mT-Tree and finally a chapter on conclusions and future research.

Synopsis

Non-traditional applications such as engineering design, manufacturing, real time systems and software engineering demand high performance object management systems. A typical implementation of an object oriented database on a conventional architecture leads to a reduction in performance in terms of data storage, data access and transaction throughput.

Our aim is to meet the objectives of: storage of large and variable length objects; faster data access, and high transaction throughput. To achieve these aims we propose storing the entire database in main memory, leading to the notion of Main Memory Object Oriented Databases (MOOD). Main memory database systems are available but until now restricted to the relational database model. The key components of a MOOD system affecting the system performance are the indexing scheme, the storage structure, and the concurrency control mechanism. To gain high performance, the storage structure is made efficient both for objects and schema; data access and query processing is enhanced by the efficient access method; transaction rate is increased by improving concurrency control.

A new indexing structure (mT-Tree) for Main Memory OODBMS is designed that merges single class and class hierarchy indexing and is equally efficient for com-

posite class hierarchy. The problems associated with the existing storage structures lead to the design of a new storage structure which supports class storage as well as storage of class hierarchy and composite class hierarchy. A concurrency control implementation suited to faster main memory environment is also designed. The new locking structure supports both class hierarchy, and composite class hierarchy as well as single class. Locking information is attached to the data for faster transaction processing.

The data structure designed in this thesis is evaluated both analytically and empirically. The results show that a performance enhancement is achieved as is desired of such a data structure.

Acknowledgements

I am grateful to Islamic Development Bank, Saudi Arabia for providing the financial support to achieve one of my goals in life.

I would like to thank my research supervisor Prof. Jorge B. Bocca for bringing me to the light in the world of database technology. I am also indebted to him for allowing me to explore and develop my own ideas and interests, and for putting me back to the right path when I was losing the direction.

I would also like to thank my teacher and Director Dr. Inam-Ur-Rahman at Centre for Nuclear Studies, Pakistan, for his consistent support and help to make my dream come true.

I also acknowledge support of my friends in the Knowledge Base Group (KBG), Julie Wilks and Ian Wilcox at the University of Birmingham. I am specially obliged to Dr. Iain McLaren and Dr. Salem-Al-Naemi of the KBG, for their useful comments and numerous discussions over my research. Salem without you my stay at Birmingham would have not been so enjoyable.

I wish to thank my head of school, Dr. Peter Dodd, all his colleagues and my fellow research students for their help and support. I am specially grateful to the members of my thesis monitoring group, Prof. Aaron Sloman and Alan Sexton for

their questions and suggestion during my research.

A big thanks to my parents for their love, support and prayers throughout my life. Finally, I am grateful to my wife Durdana and sons Ahraaz and Ahraar, for their patience, understanding and encouragement in difficult times during my work: without their support and help this thesis would have not been possible.