Preface

Research is increasingly focusing on using multicomputers [Cul94] [ÖV91] [Tan95]. Multicomputers are built from mass-produced PCs and workstations, often integrated with high bandwidth networks. Such infrastructures have emerged rapidly in the last decade and many organizations have typically hundreds of networked machines with a large total amount of RAM, CPU and disk-storage resources.

Multicomputers provide a challenge and promise to cope with the ever-increasing amount of information using new distributed data structures. Scalable Distributed Data Structures (SDDSs) form a class of data structures, first proposed in [LNS93], that allows for scalable distributed files that are efficient at searching and insertions. The approach taken sets virtually no upper limit on the number of nodes (computers) that participate in the effort. Multiple autonomous clients access data on the server-nodes using a local image to calculate where the data is stored. Their images might be outdated, but the client is updated when it incurs an addressing error.

This thesis contributes to the field of distributed data storage in the context of very large database systems. Even though several commercial database systems handle large volumes of data (>1TBytes at year 2000), the distribution of this data is still limited to a cluster of local machines. Today, with growing networking environments, more information is available and need to be organized to facilitate fast access. Single computers' resources are commonly shared over the networks using centralized approaches, that eventually cannot handle the load. Alternatively, computers can work together on one problem abolishing centralized approaches. When computers jointly cooperate on storing/processing we refer to them as a multicomputer. Linear scale-up is important performance target for multicomputers and parallel machines to fully utilize the increased cost of the hardware and interconnects. The optimal objective for scale-up is that a doubling of the computer resources allows the double work(load) to be tackled in the same time. However, there are many orthogonal aspects that should scale-up, such as storage, availability, processing, retrieval, and filtering. The research field of Scalable Distributed Data Structures has presented a number of solutions for storage, etc., but integration or actual implementations have been scarce. We present a number of novel scalable data structures, including experiences.
from the first actual implementation of an SDDS on a parallel machine. Furthermore, we show the feasibility of readily integrating an SDDS into an extensible database system. The thesis emphasis is on hash-like structures and the practical experiences learned through simulated experiments and implementations on multi-computers. One common property of all the structures presented herein are that they employ bit-string addressing allowing for low cost storage and navigation in the data structures.

Observing the growing of the amounts of data being made available through networks we are convinced that scalable distributed data structures will have implications on storage as well as processing in the internet and database community at the beginning of the new millennium.

The focus of the thesis is on novel scalable distributed data structures (SDDSs) for handling large amounts of data. The work is proven practical and does not engage in aspects of the complexity theory. It deals with explorative research in the domain of scalable distributed data structures, but is not conclusive. The thesis is based on a number of publications, listed below together with the chapters in which they occur. It means that some parts may be duplicated, such as introduction to SDDSs, keeping the continuity of the paper. The papers on LH*LH [KLR96] and hQT* [Kar98] deals with SDDSs, [KK98] deals with an SDDS's integration into a DBMS. The Ω-storage [Kar00] is a novel main-memory structure, which can easily be extended to be an SDDS.

**Thesis Overview**

The thesis is organized as follows. Part I deals with data structures, scalable distributed data structures in particular. Part II describes usage scenarios where SDDSs are implemented and integrated into DBMSs.

In Part I, Chapter 1 we introduce an SDDS LH* and data structures in general, completing the chapter with a short overview of some SDDSs.

Chapter 2 presents the first SDDS implementation on a large parallel machine (SMD). The LH*LH structure developed is based on LH* and extends earlier work by employing LH internally for local bucket management. It identifies the importance of local buffer management for real-time aspects and discusses hard-won experiences using the Parsytec machine. The chapter is based on a publication at the EDBT'96 conference [KLR96] and the subsequent Licentiate Thesis [Kar97].

In Chapter 3, hQT* a novel fully designed spatial data structure is presented. It includes a effective local storage schema that readily integrates with the distributed schema minimizing implementation efforts.

Part I of the thesis is concluded in Chapter 4 by introducing a novel main-memory tree structure for very large data sets: Ω-storage a new multi-attribute self organizing data storage structure. It provides an improvement
over kd-tree and bit-organized tree-structures by being more resistant to skewed data input. It can easily be extended to an SDDS scenario by utilizing splitting and distribution methods from the hQT* structure.

Part II: Chapter 5 starts with an overview of high performance database systems, their requirements and implications on data structures.

In Chapter 6, we outline the features of one system that integrate an SDDS into an extensible DBMS, using the DBMSs native extension interface. By implementation of one such system we show the viability of the concept and implicate possible extensions [KK98]. A vision of a new concept Live Optimization is introduced, which we believe is a promising way to simplify distributed query processing.

Chapter 7 concludes the thesis by summarizing contributions and by enumerating problems that need further attention.

Financial Support

The first part of my PhD work was supported by NUTEK (The Swedish National Board for Industrial and Technical Development), and CENIIT (The Center for Industrial Information Technology) and was performed during the years 1994–1996 at the Computer Science Department (IDA), at Linköpings University, Sweden.

Since, 1996 the project was partially funded by the HPCN/IMPACT project at CWI, Amsterdam, The Netherlands.