Database Research at the Data-Intensive Systems Center

David Maier, Lois Delcambre, Calton Pu, Jon Walpole
Oregon Graduate Institute

Goetz Graefe, Len Shapiro
Portland State University

DISC
c/o Dept. of Computer Science & Engineering
Oregon Graduate Institute
20000 NW Walker Road
P.O. Box 91000
Portland, OR 97291-1000

{maier, lmd, calton, walpole}@cse.ogi.edu *{graefe, len}@cs.pdx.edu*

1 Introduction

This report briefly describes the research activities of the newly formed Data-Intensive Systems Center (DISC) in Portland, Oregon. DISC includes faculty from the Oregon Graduate Institute of Science and Technology and Portland State University. By data-intensive applications we mean applications with high data volume, high data complexity, and/or high data processing demands. Such applications place unique demands on the application software, database management systems, operating systems, and network communication facility. Data-intensive applications often require distributed, heterogeneous, and parallel databases.

DISC's activities include direct interaction with industrial and other users of data-intensive applications, with a goal of conducting research in the context of real-world problems.

One component of DISC is the Technology Exploratorium. The exploratorium will house state-of-the-art hardware and software specifically intended to support data-intensive applications. The exploratorium will provide hands-on, "kick the tires" exposure to a variety of database management systems (including traditional and object-oriented products), operating systems, design tools, applications, and datasets with trained staff and working examples. The goal is to establish the exploratorium to support technology awareness, exposure, training and transfer for local, regional and perhaps, national industrial firms and government labs and agencies. The initial focus will be guided by the technology interests of the industrial members of DISC. The exploratorium provides a mechanism to strengthen industrial competitiveness in the Information Age and also to directly support research and education.

The focus of this report is on DISC's funded research projects, summarized below. Interested readers should contact the investigators for further information and reprints. Many of the reports are available by anonymous ftp to ftp.cs.pdx.edu under /pub/faculty and cse.ogi.edu under /pub/tech-reports.

2 Rowena: A Storage System for Continuous Media Data

Faculty Participants: Jonathan Walpole, David Maier

The primary objectives of the Rowena project are to design and build a prototype storage system that provides access to large capacity storage with guaranteed latency and throughput. Rowena is designed specifically for continuous media data types, such as digital audio and video, and is based on the concepts of quality of service (QOS) based interfaces, admission testing, resource reservation and prefetching. There are two components to the application programming interface. First, applications present quality of service specifications to the storage system, indicating, among other things, their throughput, latency, and synchronization requirements. Rowena then determines the resources required to meet the specification. If sufficient resources are available they are reserved for the application. Otherwise the application's request is said to fail the admission test. Admission testing provides an appropriate way of reducing service in the presence of system overload. The second component of the application programming interface consists of guaranteed read and write calls that provide access to data with negligible latency. Rowena attempts to hide the latency of storage accesses by prefetching data into the resources that have been reserved for the application. Prefetching is driven by information passed in the initial quality of service request.

We are working closely with researchers at Tektronix to study the application of this storage architecture to a digital, network-based TV production studio.

This project is funded in part by Tektronix, the Oregon Advanced Computing Institute, and NSF.

D. Maier, J. Walpole, R. Staehli, "Storage System Architectures for Continuous Media Data," to appear in proceedings of FODO 1993.

R. Staehli, J. Walpole, "Constrained-Latency Storage Access" IEEE Computer, volume 26, number 3, pages 44-53, March 1993.

3 Query Optimization

DISC researchers are engaged in two projects involving query optimization. Although the projects cover distinct data models, there is close interaction between them and with other researchers throughout the world.

3.1 Extensible Query Optimization

Faculty Participants: Goetz Graefe, David Maier, Leo Fegaras

The Volcano project is divided into query optimization and query execution. The optimization efforts have resulted in the development of a second-generation optimizer generator, which is far superior to our earlier EXODUS prototype, particularly in search efficiency. Validation studies of that optimizer generator have been conducted for both object-oriented and scientific database systems. Moreover, we have completed a prototype optimizer that generates dynamic query evaluation plans for queries with incomplete bindings at compile-time, e.g., predicate constants in embedded queries and run-time availability. We are currently working towards a third optimizer generator that will permit more concise specification of an algebraic optimizer.

This project is funded by NSF, Texas Instruments, and an Equipment grant from Digital Equipment Corp. to the University of Colorado at Boulder.

G. Graefe and W. J. McKenna, "The Volcano Optimizer Generator: Extensibility and Efficient Search", Proc. IEEE Int'l. Conf. on Data Eng., Vienna, Austria, April 1993, 209.

G. Graefe and K. Ward, "Dynamic Query Evaluation Plans", Proc. ACM SIGMOD Conf., Portland, OR, May-June 1989, 358.

R. H. Wolniewicz and G. Graefe, "Algebraic Optimization of Computations over Scientific Databases", Proc. Int'l. Conf. on Very Large Data Bases, Dublin, Ireland, August 1993.

3.2 Query Optimization in Object-Oriented Databases

Faculty Participants: David Maier, Goetz Graefe, Leo Fegaras

Object-oriented database systems with behavioral encapsulation support powerful data abstractions, but lack the query-processing performance of relational systems. The Revelation project combines the power of abstract data types with database optimization technology into a semantically powerful and efficient query processing system.

If behavior is encapsulated, a query processor cannot reason about semantics and costs over an entire query. Optimizing queries over encapsulated types requires revealing such information to a trusted system component. The project includes the definition of an object-oriented data language, a query algebra, a revealing mechanism, an extensible optimizer, and an appropriate query evaluation mechanism.

The Revelation group participates in the EREQ

(Encore-Revelation-Exodus Query) project to develop an architecture for query processing in persistent object bases. The main results of this work will be a reference internal architecture, an object algebra, a physical plan language and prototype query optimizers that conform to these interfaces.

This project is funded by NSF grants (Revelation) and by ARPA funds (EREQ).

J. A. Blakeley, W. J. McKenna and G. Graefe, "Experiences Building the Open OODB Query Optimizer", Proc. ACM SIGMOD Conf., Washington, DC, May 1993, 287.

S. Daniels, G. Graefe, T. Keller, D. Maier, D. Schmidt and B. Vance, "Query Optimization in Revelation, an Overview", IEEE Database Eng. 14, 2 (June 1991), 58.

T. Keller, G. Graefe and D. Maier, "Efficient Assembly of Complex Objects", Proc. ACM SIGMOD Conf., Denver, CO, May 1991, 148.

D. Maier, "Specifying a Database System to Itself", in Specifications of Database Systems, D. J. Harper and M. C. Norrie (eds.), Springer Verlag, 1992.

D. Maier, S. Daniels, T. Keller, B. Vance, G. Graefe and W. McKenna, "Challenges for Query Processing in Object-Oriented Databases", in Query Processing for Advanced Database Applications, J. C. Freytag, G. Vossen and D. Maier (ed.), Morgan-Kaufman, San Mateo, CA, 1994.

D. Maier, G. Graefe, L. Shapiro, S. Daniels, T. Keller and B. Vance, "Issues in Distributed Complex Object Assembly", Proc. Workshop on Distr. Object Management, Edmonton, BC, Canada, August 1992.

4 Query Execution and Algorithms

Faculty Participants: Goetz Graefe, Len Shapiro

Volcano research into query execution techniques has recently culminated in a detailed survey paper. Earlier work can be divided into mechanisms for parallel query execution, performance enhancement through data compression, and general algorithms, particularly for relational division, sorting, and hash-based query processing. Results of research into mechanisms for parallelism are currently finding increased industry use because they combine clean software engineering with generality regarding the underlying hardware architecture. Our work comparing sorting and hashing has led to an analysis of their dualities and their comparative strength and weaknesses, a focused effort to improve traditional hash-based algorithms in their areas of relative weakness, and a polemic summarizing these and further issues and trying to rekindle the debate of sort- vs. hash-based query processing.

This project is funded by NSF, Sequent Computer Systems, Pacific Power and Light, Texas Instruments, an Equipment grant from Digital Equipment Corp. to the University of Colorado at Boulder, and an ARPA Research Assistantship in Parallel Processing administered by the Institute for Advanced Computer Studies, University of Maryland, and the Oregon Advanced Computing Institute.

G. Graefe, "Relational Division: Four Algorithms and Their Performance", Proc. IEEE Int'l. Conf. on Data Eng., Los Angeles, CA, February 1989, 94.

G. Graefe, "Encapsulation of Parallelism in the Volcano Query Processing System", Proc. ACM SIGMOD Conf., Atlantic City, NJ, May 1990, 102.

G. Graefe, "Query Evaluation Techniques for Large Databases", ACM Computing Surveys 25, 2 (June 1993), 73-170.

G. Graefe, "Sort-Merge-Join: An Idea Whose Time Has(h) Passed?", to appear in Proc. IEEE Int'l. Conf. on Data Eng., February 1994.

G. Graefe and L. D. Shapiro, "Data Compression and Database Performance, Proc". ACM/IEEE-CS Symp. on Applied Computing, Kansas City, MO, April 1991.

G. Graefe and S. S. Thakkar, "Tuning a Parallel Database Algorithm on a Shared-Memory Multiprocessor", Software - Practice and Experience 22, 7 (July 1992), 495.

5 Scientific Databases

DISC researchers are engaged in three projects involving database support for scientific computing. The projects are characterized by close interaction with domain scientists and an emphasis on building tools and prototypes.

5.1 Datatype Support for Scientific Computing

Faculty Particpants: David Maier, Jonathan Walpole and Michael Wolfe, Computer Science and Engineering; James Stanley, Materials Science and Engineering.

Scientific applications and databases rarely interoperate easily: scientific researchers expend significant time and effort writing special procedures to use their program with someone else's data, or their data with someone else's programs. Distribution and hardware heterogeneity further exacerbate the problems in connecting programs and data. This project is developing a Hybrid Data Manager architecture that uses an object-oriented database to provide a uniform interface between diverse programs and data sources.

For the domain of computational chemistry, we have developed a "computational proxy" mechanism to connect existing scientific applications to an object-oriented database. A computational proxy represents a run of an application, and organizes inputs, invocation, monitoring and output capture for the run. A complementary project in materials science explores providing a common interface to a variety of separately published datasets. An object database serves as the mediator, providing a common object model to applications by using the behavioral capabilities of the OODB to mask differences in data location and format. We are particularly interested in providing efficient query over "unmanaged data": data that resides in flat files rather than a database management system.

This project is funded by: NSF, Oregon Advanced Computing Institute and Pacific Northwest Labs, plus in-kind grants from Sequent Computer Systems, Servio Corp., Object Design, Inc., Versant and O2 Technologies.

J. B. Cushing, D. Maier, M. Rao, D. M. DeVaney and D. Feller, "Object-Oriented Database Support for Computational Chemistry", Proc. Sixth Intl. Conference on Statistical and Scientifc Database Management, Anscona, Switzerland, June 1992.

J. B. Cushing, D. Hansen, D. Maier and C. Pu, "Connecting Scientific Programs and Data Using Object Databases", IEEE Data Engineering Bulletin 16:1, March 1993.

D. Hansen, D. Maier, J. Stanley and J. Walpole, "Object-Oriented Heterogeneous Database for Materials Science", Scientific Programming 1:2, Winter 1992.

5.2 An Object-Oriented Toolbox for the Protein Data Bank

Faculty participant: Calton Pu

We are designing and implementing a new software structure and toolbox for the Protein Data Bank. The toolbox contains a graphical user interface as the front end, an interoperable programmer interface as the database back end, and a set of scientific tools that connect to both ends. We have implemented a demo program in C++, called PDBtool, which demonstrates the feasibility of our architecture. PDBtool is being extended for experimental use as a molecule verification toolbox. Currently, we are evaluating different database system managers for the backend storage and designing a similar system for the management of weather prediction data. New research issues in biology and computer science raised by this work will be discussed.

Currently, we are working on a robust version of the PDBtool and PDBlib, a class library for users of PDB, for beta test and subsequent release for general use by crystallographers and biologists. In addition, we are developing a new set of tools to be distributed with the new standard CIF format for molecular structure description.

This project is funded by NSF.

Pu, Calton et al, "A Prototype Object-Oriented Toolkit for Protein Structure Verification", Technical Report CUCS-048-92, Department of Computer Science, Columbia University, 1992.

5.3 Algebraic Optimization and Parallel Execution of Computations over Scientific Databases

Faculty Participant: Goetz Graefe

Since many scientific applications manipulate massive amounts of data, database systems are being considered to replace the file systems currently in wide use for scientific applications. In order to counteract the performance penalty of additional software layers (i.e., the database management system), we are investigating the use of traditional database techniques to enhance the performance of computations over scientific databases. Our two focus areas are automatic optimization and parallelization of processing plans that include both numeric and database operations.

An integrated algebra including both relational and numeric operations is the crucial point in our research. The important advantages of using a single algebra are that (a) the scope of query optimization, which had been limited to the database system's retrieval and pattern matching operations, has been extended to cover the entire computation, (b) the traditional two-level approach (retrieval vs. computation) has been overcome, permitting preliminary computations such as sampling to be performed before complex database operations such as matching of large database sets (joins, intersections, etc.), and (c) successful optimization techniques can be transferred easily from the database systems domain to scientific computations.

This project is funded by NSF.

Wolniewicz, R. H. and G. Graefe, "Algebraic Optimization of Computations over Scientific Databases," Proc. Int'l. Conf. on Very Large Data Bases, Dublin, Ireland, August 1993.

Graefe, G., R. L. Cole, D. L. Davison, W. J. McKenna, and R. H. Wolniewicz, "Extensible Query Optimization and Parallel Execution in Volcano," in Query Processing for Advanced Database Applications, ed. J. C. Freytag, G. Vossen and D. Maier, Morgan-Kaufman, San Mateo, CA, 1994.

Graefe, G. and D. L. Davison, "Encapsulation of Parallelism and Architecture-Independence in Extensible Database Query Processing," IEEE Trans. on Softw. Eng., vol. 19, no. 7, July 1993.

6 Extended Transaction Processing using Epsilon Serializability

Faculty Participant: Calton Pu

We have introduced the notion of Epsilon Serializability (ESR) to relax the limitations imposed on concurrency, availability, and autonomy by classic serializability. For On-Line Transaction Processing (OLTP), ESR alleviates data contention in a more precise way than the level 2 consistency of DB2. For extended transaction models beyond OLTP, ESR offers controlled relaxation of consistency and atomicity, properties that can be combined with other methods to relax isolation and durability.

ESR gives application designers a fine-grain control in the trade-off between the amount of inconsistency tolerated by each transaction and improved performance. We will describe the recent results in the research based on ESR, including a formal characterization of ESR based on the ACTA framework, the distributed divergence control algorithms based on our previous work on centralized divergence control methods, the design of consistency restoration algorithms, and some other related works.

This project is funded by Oki Electric Ind.

Recent Publications:

Ramamrithan, K. and Pu, C., "A Formal Characterization of Epsilon Serializability", IEEE Transactions on Knowledge and Data Engineering, to appear 1993.

Pu, C. and Hseush, W.W. and Kaiser, G.E. and Yu, P. S. and Wu, K.L., "Distributed Divergence Control Algorithms for Epsilon Serializability, in Proceedings of the Thirteenth International Conference on Distributed Computing Systems", Pittsburgh, May, 1993

Wu, K.L. and Yu, P. S. and Pu, C., "Divergence Control for Epsilon-Serializability", in Proceedings of Eighth International Conference on Data Engineering", February 1992, pg 506-515.

Pu, C. and Leff, A., "Replica Control in Distributed Systems: An Asynchronous Approach", in Proceedings of the 1991 ACM SIGMOD International Conference on Management of Data, May 1991, Pg 377-386.

7 Application Modeling

Faculty Participants: David Maier, Lois Delcambre, and Leonard Shapiro

The goal of this research is to provide a single, integrated model for the structural, behavioral, and active aspects of an application. Object-oriented databases provide the starting point for this research, but we are adding an active mechanism to the rich structural and behavioral constructs that they provide. The major focus is the definition and successful integration of the active component with the behavioral and structural components. Research goals include providing:

1) An active invocation mechanism that is more expressive than single message passing. In a number of applications, e.g. manufacturing and design, processing must be invoked based on the presence of multiple objects or the occurrence of one or more events.

2) Events and conditions (much like triggers in active databases) to match the way things happen in the application.

3) Manipulation of multiple objects at once while preserving the behavioral abstraction provided by object-oriented databases. The idea is to invoke methods within the context of input and output objects for each processing step.

4) Formally-defined semantics for the active invocation mechanism to serve as the basis for implementation, facilitate the algorithmic processing and translation of models, and support the articulation and enforcement of constraints for well-formed specifications within the model.

The research is investigating the use of the Object Flow Model as the basis for database application modeling. The Object Flow Model is a conceptual modeling language that also serves as the basis for discrete event simulation. The Object Flow Model has been defined to include a rich structural description of objects, the specification of method signatures, and an active component called the Object Flow Diagram (OFD). An OFD is a data- (and event-) driven model, inspired by the dataflow model of computation, that derives its formal semantics from a deductive database rule language defined, in turn, using predicate transition networks.

This project is funded by Pacific Power and Light, and the Oregon Advanced Computing Institute.

Delcambre, L., Narayanswamy, J., Pollacia, L. "Simulation of the Object Flow Model: A Conceptual Modeling Language for Object- Driven Applications", Proceedings of the 26th Annual Simulation Symposium, Arlington, VA, April 1993.

Delcambre, L. and Pollacia, L.F. "The Object Flow Model for Data-Based Simulation", Proceedings of the 1993 Winter Simulation Conference (invited paper), Los Angeles, CA, December 1993.

Pollacia, L.F. and Delcambre, L. "The Object-Flow Model: A Formal Framework for Describing the Dynamic Construction, Destruction and Interaction of Complex Objects", Proceedings of the Twelfth International Conference on Entity-Relationship Approach, Dallas, TX, Dec. 1993.