Florida International University
High Performance Database Research Center

Naphtali Rishe, Director

Wei Sun (Assoc. Director), David Barton, Yi Deng, Cyril Orji,
Affiliated Professors
Michael Alexopoulos, Leonardo Loureiro, Carlos Ordonez, Mario Sanchez, Artyom Shaposhnikov,
Group Coordinators

School of Computer Science
FIU, University Park, Miami, FL 33199
Telephone: (305) 348-2025, 348-2744
FAX: (305)-348-3549; Internet: rishen@fiu.edu
WWW: http://www.cs.fiu.edu/HPDRC

The High-performance Database Research Center is a division of Florida International University, School of Computer Science. It conducts research on database management systems and various applications, leading to the development of new types of DBMS, new database techniques, and the refinement of existing ones. Our Center has a strong commitment to training graduate students and preparing them for their future roles as scholars and specialists in the industry. The Center is funded by various government agencies and industry; the largest benefactor is NASA with $3.8 million. Other sponsors include: National Science Foundation, U.S. Department of Defense (BMDO, ARO, USAF, and DISA), U.S. Department of the Interior, U.S. Information Agency, NATO, Florida Department of Commerce, Florida Department of Education, Baxter Corporation, and GeoNet Limited.

The Center presently employs thirty-one research assistants: R. Alentado, E. Alvarez, M. Alexopoulos, K. Beznosov, I. Bluvstein, S. Chen, W. Du, S. Fedorishin, S. Graham, M. Gutierez, S. Guo, S. Hong, Y. Ivanov, R. Kallem, S. Kolla, Y. Ling, L. Loureiro, J. Li, S. Lu, R. Martinez, R. Medina, N. Morisseau-Leroy, K. Naboulsi, C. Ordonez, B. Parenteau, V. Patil, Z. Rong, M. Sanchez, A. Shaposhnikov, M. Somasekhar, A. Vaschillo, M. Wang.

Following are the principal current projects of our Center.

1. High-performance Semantic DBMS

Our largest project is the development of algorithms and a prototype of a massively paralleled Semantic DBMS. Our system should be useful for most typical database applications, as well as for specialized domains such as Earth Sciences.

Many database applications, e.g. those for Earth Sciences, have three essential needs: (1) strong *semantics* embedded in the database — to handle the complexity of information; (2) storage of multi-dimensional spatial, image, scientific, and other *non-conventional data*; and (3) *very high performance* — to allow massive data flow. Abundant evidence demonstrates that semantic/object-oriented databases can satisfy the first two needs better than relational databases. We are currently developing a semantic/object-oriented approach that will also satisfy the high performance need of earth science applications.

Our research aims to significantly improve the usability and efficiency of highly parallel database computers and machine clusters (tightly networked groups of machines). Our prototype database management system will have substantial advantages over current database machines:

(1) *Usability*. Our object-oriented system is based on the Semantic Binary Model, unlike most current database systems, which are mainly based on the Relational Model. Inherent in the semantic model are superior logical properties like: friendlier and more intelligent generic user interfaces based on the stored meaning of the data, comprehensive enforcement of integrity constraints, greater flexibility, and substantially shorter application programs.

Semantic databases represent information as a collection of objects and relationships between these objects. The Semantic Binary Model of databases is a semantic model with object-oriented features [Rishe-92-DDS]. Data items related to objects can be of arbitrary size, multi-valued, or missing entirely. We have applied this approach to various types of data, including scientific and multi-media data. Flexibility is enhanced since objects are not required to be identified by keys.

(2) *Efficiency*. The algorithms implemented in our system, e.g. [Rishe-91-FS], makes it more efficient than conventional database machines. This is due, in part, to the system's understanding of the data's semantics and to the higher abstraction level. Our prototype system under development is highly efficient for both small and massive numbers of processors equipped with separate memories and storage devices. In particular, the use of semantics allows better exploitation of parallelism [Rishe&*al.*-91-PA], by providing a means of distributing data among different processors in a way which is transparent to both database programmers and database users [Rishe&*al.*-94-LB].

Following are some of the problems we are currently addressing in the framework of this research:

High Performance Data Storage Scheme: **1.** *Parallel storage structure.* **2.** *Efficient interconnection network.* **3.** *Massive I/O.* **4.** *Efficient retrieval from tertiary storage.* **5.** *Balanced load.*

Beyond Conventional Data: **6.** *Scientific data.* (In particular, our semantic implementation supports variable-length, unlimited precision numeric data [Rishe-92-IB].) **7.** *Spatial data.* (Efficient storage of temporo-spatial data combined with facts and scientific data.) **8.** *Multi-media.* **9.** *Compression.* (The data stored undergoes close to optimal compression, without any negative impact on query processing performance.)

Beyond Conventional Query Processing: **10.** *Guaranteed optimality for basic queries.* Every simple query is normally answerable from information under the control of just one processor, thus reducing communication overhead and allowing most processors to work on different queries simultaneously. (For any basic query, its originating processor deduces the identity of the information-possessing processor by matching the query with a copy of the data partitioning table.) Moreover, the answer to such a simple query would be available either from the latter processor's memory or from just one access to the processor's storage unit (in the current technology, it is a retrieval of one block from a disk). (The processor deduces the identity of the data block containing the answer to the query by matching it with memory-held data structures.) An algebra of "basic queries," which are the building blocks for complex queries, is defined in [Rishe-91-FS]. The basic queries include the range queries and many other "bulk" queries. **11.** *Efficiency of complex queries.* **12.** *Retrieval of spatial data by content.* Content-based or feature-based retrieval is particularly important to earth science applications that handle a massive amount of data. **13.** *High-level query languages* [Rishe&*al.*-91-DT], [Rishe-95-UM]. **14.** *Linear relation of throughput vs. cost.* One of our design goals is to attain a close-to-linear correlation between the number of processors (with storage devices) in the system and the system's throughput in terms of the number of simple queries and transactions per unit of time. This close-to-linear correlation should be valid even for massively paralleled systems involving thousands of processors.

Concurrency Control: 15. *Efficient and semantically-safe concurrency control.* Efficient algorithms to control concurrency of transactions are required. Utilizing the semantic knowledge of the database, the algorithms must assure not only the avoidance of interference at a physical storage level (e.g. concurrent transactions do not spoil the data structure; nor do the users see the database in a state partially updated by a transaction), but also non-interference at the semantic level (e.g., the information on which the program relies during the computation of a transaction does not change until the transaction is complete). This is based on the principle of "information relied upon in a transaction." We employ fast optimistic concurrency control with some optimizing compromises toward pessimistic control. Granularity at the level of elementary logical facts is attained. However, there is no overhead in the physical data structure.

2. Database Design Tool

We have developed a tool [Rishe&Sun-94-PC] for the design of relational databases, including schemas, integrity constraints, reports, and data entry forms, with the use of semantic binary schemas. The tool is based on our top-down database design methodology [Rishe-93-MT], [Rishe-92-DDS]. In this methodology, a conceptual description of the enterprise is designed using a semantic binary model. This description is then converted into the relational database design. Our tool automates virtually all the busy work of design. The tool allows for intelligent design decisions to be taken by the database designer, or it can default to "rules-of-thumb" principles guided by the system's knowledge of the database's semantics. It creates a turn-key database application and its documentation with graphically-illustrated design reports, manuals, application glossaries, and data dictionaries, as well as an application-customized report generator. Changes in the semantic description or designer's instructions are propagated into the final application. We have used this tool for the design of very large applications. For example, for one client, the Everglades National Park, we have developed an ORACLE database containing about 2,000 attributes in about 200 tables of environmental observations, both archival and real-time.

3. Heterogeneous Distributed Database Management Over Public Networks

Recent advances in public network technology and emerging database interface standards has enabled the implementation of a new type of distributed database systems. This in turn, has created demand for improved query distribution and query optimization algorithms.

Current technology offers the possibility of connecting local computer systems with remote servers via high speed telephone lines. In addition, direct access to Intelligent Networks (SS7 messaging service of the telephone companies) allows for centralized management of communication. A network coordinator node may communicate with every node in the network, using the

Intelligent Network to serialize and optimize data flow over data lines, e.g. ISDN. Moreover, this coordinator node is capable of establishing the necessary telephone connections between all the nodes and managing system resources such as data line allocation.

These new technological possibilities are of particular interest for distributed database system implementation. High bandwidth data exchange is possible, and the central coordinator node can take responsibility for enforcing the required consistency constraints present in database environments [Rishe-94-MN]. An example of related research that we are conducting is cost minimization for ISDN usage.

4. Spatial Data Applications

We are developing several spatial database study-applications in order to demonstrate and evaluate our semantic high-performance DBMS. The data sets selected for this development process include ocean temperature data (supplied by the University of Miami), TOMS2 (Total Ozone Mapping Spectrometer deployed by NASA) and SeaWiFS (SEA-viewing WIde Field-of-view Sensor to be deployed by NASA 1995; simulation data is currently used). When working with these data sets, the methodology in place ranges in scope from developing of the respective metadata, to producing geodetic displays of the interpretive values of the data. This has been done while still adhering to aggressive performance criteria. The goal of these processes is to develop a semantic binary schema and render the data in hyperquadrant form for implementation into the semantic database engine that will be the means to sustain a temporal GIS.

5. Query Optimization

Efficient and Accurate Size Estimation in a DBMS: We have proposed novel strategies for estimating the size of the resulting relation after a join, selection and/or a projection using the regression model and a neural learning model [Sun&*al.*-93-IA], [Ling&*al.*-95-CE]. The proposed methods provide an instant and accurate size estimation with little run-time overheads. Our methods are comparable in accuracy to that of the 20% sampling, which is believed to be an accurate estimation and incurs run-time cost that adds to the response time of processing a query. Our approach has also been shown to be applicable to various practical situations, such as very skewed data, correlated data, complex selections, projections and joins.

Semantic Query Optimization and Solving Satisfiability/Implication Problems: A novel concept of the minimal knowledge base has been introduced, which eliminates redundancy and inconsistency of constraints. A method to maintain the knowledge base and a systematic approach to optimize queries using constraints have been proposed. We have proved that the general semantic query optimization is NP-hard and provided efficient algorithm for several restricted types of queries [Sun&Yu-94-SQ]. Semantic query optimization in OODB in centralized and distributed environment is also studied, where many distinct object-oriented features are incorporated. Further, our studies [Sun&Weiss-94-IA] and [Guo&*al.*-95-SS] have provided a comprehensive solution to the satisfiability of a conjunctive formula and implication and equivalence between two conjunctive formulae involving inequalities (equi-joins, theta joins, selections, etc.) in a DBMS. These problems originated from semantic query optimization. However, our results are fundamental to the design and implementation of relational, object-oriented, and deductive database systems.

6. I/O Problems in DBMS

Recent trends in the relative improvements in speeds between the Central Processing Unit (CPU) and Input/Output (I/O) system components suggest that the I/O system could become a bottleneck in computer systems. If this happens, many applications, such as database systems, could become so I/O-limited that increases in CPU speeds would no longer be helpful. High performance I/O systems are needed to address this imbalance. Our research has focused on investigating a number of issues related to I/O performance in databases. We are developing and verifying algorithms that improve the efficiency of random small writes. This type of workload dominates many on-line transaction processing (OLTP) applications. We are also characterizing I/O performance in very large databases and developing a data management model for improved I/O performance. Applications such as brokerage houses, retail chains and mail order houses are increasingly making use of very large databases for advanced OLTP applications. Because many of these advanced applications make use of multi-media information systems, our research has also focused on techniques for efficient storage and retrieval of multimedia data.

Non-volatile (safe) disk buffers improve performance in DBMSs. Writes to the disk buffer are durable without incurring the cost of the physical disk writes. Moreover, physical writes at the disk can be performed at low cost using write piggybacking and multiblock purges. However, safe disk buffers are expensive. As part of our efforts in improving random small writes in database systems, we have developed efficient algorithms using unsafe buffers to simulate safe ones. This is done by writing each block twice to disk. The first write is performed immediately, anywhere. The second write is deferred, but written at a fixed location. These two writes together cost less than a single random write [Orji&Solworth-94-WT].

7. Software Engineering and Distributed Systems Research

The main theme of research of our software engineering and distributed systems group, led by Yi Deng, is to use software engineering approaches and methods to address the design of distributed information systems. One project undertaken by this group is to use an operational approach for modeling, decomposition, and analysis of object-oriented (OO) distributed information systems. We have developed an executable formal method called G-Nets [Deng&*al.*-94-ES], by integrating the theory of Petri nets and the OO structuring mechanism. The purpose of G-Nets is to precisely specify the design of distributed OO systems. This includes not only the behavior, but also system structure or architecture. A goal of the project is to provide a tool to support the control design of our high-performance DBMS. Currently, this project proceeds from three inter-related aspects. First, we are working to extend the formal representation of G-Nets to the real-time domain. The new model, called real-time G-Nets (RTG-Nets), is aimed at architectural modeling of OO real-time systems. This is an emerging research area for OO development. Second, we are implementing a distributed system environment for the execution, simulation and testing of G-Net specifications. The goal is to provide a way, in addition to formal analysis, to analyze and verify a system design at an early stage of the development. Third, we are investigating an architectural decomposition method based on the G-Net representation to support the refinement and transformation of distributed OO designs. This is done by gradually decomposing large-grained system components to fine-grained objects, and by progressively introducing non-functional properties.

Future directions

The Center is expanding. Effective with the Fall 1995 semester we will be joined by a new faculty member (Dr. Chungmin Chen), two visiting professors, and five RAs. Substantial additional growth is planned for 1996 and 1997. The Center intends to further expand its government support base by offering solutions to fundamental database research problems. We are also doing much applied research for industry and are open to new challenges.

References

[Deng&*al.*-94-ES] Y. Deng, S.K. Chang and X. Lin, ‘‘Executable Specification and Analysis for the Design of Concurrent Object-Oriented Systems’’, *International Journal of Software Engineering and Knowledge Engineering*, Vol.4, No. 4, 1994, 427-450.

[Guo&*al.*-95-SS] S. Guo, W. Sun, and M. Weiss, ‘‘Solving Satisfiability, Implication, and Equivalence Problems Involving Conjunctive theta-Joins and Selections in Database Systems’’, to appear in *IEEE Trans. on Knowledge and Data Engg.*

[Ling&*al.*-95-CE] Y. Ling and W. Sun, ‘‘A Comprehensive Evaluation of Sampling-Based Size Estimation Methods in Database Systems’’, Proceedings of the IEEE 11th Int’l Conf. on Data Engg., Taiwan, March 1995.

[Orji&Solworth-94-WT] C. Orji and J. Solworth. ‘‘Write-Twice Disk Buffering’’. Proceedings of the Third International Conference on Parallel and Distributed Information Systems, Austin, Texas, 1994. (pp. 27-34)

[Rishe-91-FS] N. Rishe. ‘‘A File Structure for Semantic Databases.’’ *Information Systems*, 16, *4* (1991), pp. 375-385.

[Rishe-92-DDS] N. Rishe. *Database Design: The Semantic Modeling Approach*. McGraw-Hill, 1992, 528 pp.

[Rishe-92-IB] N. Rishe. ‘‘Interval-based approach to lexicographic representation and compression of numeric data.’’ *Data and Knowledge Engineering,* 8, *4* (1992), pp. 339-351.

[Rishe-93-MT] N. Rishe. ‘‘A Methodology and Tool for Top-down Relational Database Design.’’ *Data and Knowledge Engineering*, 10 (1993) 259-291.

[Rishe-94-MN] N. Rishe. ‘‘Managing Network Resources for Efficient, Reliable Information Systems,’’ panel position paper, *Proceedings of the Third International Conference on Parallel and Distributed Information Systems* (Austin, Texas, September 28-30, 1994), IEEE Computer Society Press, 1994, pp. 223-226.

[Rishe-95-UM] N. Rishe, ‘‘A Universal Model for Non-procedural Database Languages,’’ *Fundamenta Informaticae*, in press (1995).

[Rishe&*al.*-91-DT] N. Rishe, S. Navathe and D. Tal (eds.) *Databases: Theory, Design and Applications.* IEEE Computer Society Press, 1991, 296 pp.

[Rishe&*al.*-91-PA] N. Rishe, S. Navathe and D. Tal (eds.) *Parallel Architectures.* IEEE Computer Society Press, 1991, 306 pp.

[Rishe&*al.*-94-LB] N. Rishe, A. Shaposhnikov, and W. Sun. ‘‘Load Balancing Policy in Massively Parallel Semantic Databases’’ Proceedings of the First International Conference on Massively Parallel Computing Systems, IEEE Computer Society Press, 1994, pp. 328-333.

[Rishe&Sun-94-PC] N.Rishe and W.Sun. ‘‘A pipeline CASE tool for database design.’’ Proceedings of SEKE’94: Sixth International Conference on Software Engineering and Knowledge Engineering, pp. 336-343. KSI, 1994.

[Sun&*al.*-93-IA] W. Sun, Y. Ling, N. Rishe and Y. Deng. ‘‘An Instant and Accurate Size Estimation Method for Joins and Selections in a Retrieval-Intensive Environment’’, Proc. of SIGMOD 1993, May 1993, Washington, D.C., pp. 79-88.

[Sun&Weiss-94-IA] W. Sun and M. Weiss, ‘‘An Improved Algorithm for Implication Testing Involving Arithmetic Inequalities’’, *IEEE Trans. on Knowledge and Data Engg.,* Vol. 6, No. 6, 1994, pp. 997-1001.

[Sun&Yu-94-SQ] W. Sun and C. Yu.‘‘Semantic Query Optimization for Tree and Chain Queries’’, *IEEE Trans. on Knowledge and Data Engg.* Vol. 6, No. 1, 1994, pp. 136-151.