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

Florida International University
High Performance Database Research Center

Tevfik Rishe, Director
Wei Sun (Assoc. Director), David Barton, Ir Deng, Cyril Orji,
Affiliated Professors
Michael Alexopoulos, Leonardo Lubrano, Carlos Ordonez, Mario Sanchez, Artem 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

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-PB].

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

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

Florida International University

High Performance Database Research Center

Naphtali Rishe, Director

Wei Sun (Assoc. Director), David Barton, Yi Deng, Cyril Orji,
Affiliated Professors

Carlos Ordonez
Group Coordinators

School of Computer Science
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 DBMSs, 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. Athauda, E. Alvarez, M. Alexopoulos, K. Beznosov, I. Blaustein, S. Chen, W. Du, S. Fedorishin, S. Graham, M. Gutierrez, S. Guo, S. Hong, Y. Ivanov, R. Kallem, D. Kolla, Y. Liang, L. Loureiro, J. Luo, R. Martinez, R. Medina, N. Morisseau-Leroy, K. Nabutski, C. Ordonez, B. Parenteau, Y. 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 parallel 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 dataflow. 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

Florida International University
High Performance Database Research Center

Naphtali Rishe, ...
Wei Sun (Assoc. Director), David Barton, Yi Deng, Cyril Orji,
... Affiliated Professors
... Carlos Ordonez
Group Coordinator
School of Computer Science
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

Florida International University
High Performance Database Research Center

Telephone: (305) 348-2025, 348-2744
FAX: (305) 348-3549; Internet: rishen@fiu.edu
WWW: http://www.cs.fiu.edu/HPDRC