

INFRASTRUCTURE 00:

NSF CISE/EIA RI and MII PI'S Workshop

00-11R

Snowbird, Utah

July 7-9, 2000

A workshop for the principal investigators of NSF/CISE infrastructure awards, sponsored by the CISE Research Infrastructure and CISE Minority Institutions Infrastructure Programs, Division of Experimental and Integrative Activities (EIA) in the Directorate Computer and Information Science and Engineering (CISE) of the National Science Foundation.

Infrastructure for Research and Training on High-Performance Heterogeneous Distributed Database Management

High Performance Database Research Center
School of Computer Science, Florida International University
Miami, FL 33199; rishen@fiu.edu; <http://hpdrc.cs.fiu.edu>

PI: Naphtali Rishe

Co-investigators: W. Sun, M. Chekmasov, N. Prabakar, S. Chen, D. Beryzoza, M. Chekmasova

Introduction

Florida International University (FIU) is one of the largest majority-minority doctoral-granting universities in the United States. Nearly 70% of our students are minorities. The University has the largest contingent of Hispanic students of any doctoral-granting university in the country, and graduates the most Hispanic engineering students in the Nation. The High Performance Database Research Center (HPDRC) was founded in 1994, and is associated with the School of Computer Science at Florida International University. HPDRC conducts research on database management systems and various applications, leading to the development of new types of database systems and refinement of existing database systems.

The general goals of the project, now in its third year, are to provide an infrastructure that will enable FIU's HPDRC to perform heterogeneous database research and to better recruit and retain minority students through their M.S. and Ph.D. degrees. Students participate in in-depth research and training in heterogeneous database integration.

FIU is an urban university whose surrounding community base is substantially comprised of under-represented minorities: 88% of the students of Miami-Dade County Public Schools are members of under-represented minorities, including 53% Hispanic, 33% Black non-Hispanic, and 2% Other. One goal of this project is to establish a regional outreach program to attract talented local minority students to FIU. Without the support of this project, those students would otherwise not be able to take advantage of the career and educational opportunities, or would have to attend an out-of-state university (a non-favorable choice of many of the local minority students).

The infrastructure being assembled will provide the students with a networked computing environment on which the research work will be conducted. The ultimate research goal is to develop a heterogeneous database management system, using semantic modeling to integrate and reconcile information from multiple, disparate data sources. Of particular interest are the methodologies to integrate geo-spatial and Web data sources. Geo-spatial data are vital to environmental research and studies (e.g. the Global Warming effect) but are often collected and stored in independently operated organizations. Web data generate new issues in data integration because, unlike traditional databases or data repositories, Web data are usually made available through form-filling interfaces, without divulging the data model behind the scenes. Specific research issues include: heterogeneous data model integration using semantic modeling, specification of Web data sources, geospatial data integration, reconciliation, and fusion (e.g. overlapping raster and vector data), rapid integration methodologies, query processing and optimization, and exploration of mobile agent technology.

HPDRC maintains a WWW page describing its projects and staff at <http://hpdrc.cs.fiu.edu>.

Third Year Accomplishments

Goals, Objectives, and Targeted Activities

The goals of our MII (Minority Institute Infrastructure) grant are to provide an infrastructure that will enable FIU's HPDRC to better recruit minority faculty members, better recruit and retain graduate students through the Ph.D., and to perform more in-depth research and training in database management.

Since the grant's inception in the Fall of 1997, we have been striving to achieve these goals. The activities we have been engaged in are described in the following sections.

Recruiting Minority Faculty: HPDRC is still seeking minority faculty members to hire. Maria Martinez, an FIU Computer Science PhD graduate and a visiting professor of Electrical and Computer Engineering at FIU, is an HPDRC-affiliated faculty member who continues to serve as a role model to our students. Minority SCS faculty members Dawn Holmes and Joslyn Smith are taking part in the NSF-sponsored HPDRC activities with students. Tenure-track offers have been made to two minority and female professors who are database researchers. They were not able to accept our offers.

Retaining Graduate Students Through the Ph.D.: Rosany Rodriguez, a minority PhD student, is nearing the completion of her dissertation. Melinda Whiley, a graduate of Florida A&M University, joined FIU's HPDRC as a graduate research assistant in March 2000 as a result of our recruiting efforts. We continue to recruit promising students to take advantage of the funds provided by our MII grant and continue to fund current graduate students via MII as new students are identified. Professors Jai Navlakha (FIU School of Computer Science) and Luis Martinez-Perez (FIU College of Education) continue to research methods to enhance the retention of graduate students through the PhD. We have stepped up our recruiting efforts by sending out a letter containing information on FIU's PhD program in Computer Science to current graduate students. We have also sent information on FIU's graduate programs in Computer Science to 800 graduates of FIU's CS undergraduate program who have qualifying GPAs. In the Fall of 1999, Dr. Luis Martinez-Perez and Dr. Richard Campbell (FIU's Dean of Graduate Studies) visited four campuses of InterAmerican University and the University of Puerto Rico. As a result of this exploratory visit, a recruiting trip was scheduled and completed in late February 2000. The recruitment team consisted of Dr. Martinez-Perez, Dr. Sushil Gupta with the College of Engineering, Nancy Colon with the Office of Graduate Studies, and Saul Sztam, Director of Academic Support services with the College of Journalism and Mass Communications. Students who expressed interest in FIU are being contacted through followup correspondence. We are continuing recruitment efforts directly with Florida A&M University, Florida Memorial College, Miami-Dade Community College, and the Miami-Dade County Public Schools. We have involved high school students in our research through Miami-Dade County Public School's Advanced Academic Internship Program. Their involvement provides us a pipeline of researchers from high school through the PhD.

Affinity Groups: Four Affinity Groups modeled after those at the University of Texas, El Paso are in place at HPDRC. The Affinity Groups are made up of faculty members, postdoctoral associates, and graduate and undergraduate students. The following groups continue to pursue research tracks at HPDRC: *Semantic Database Engine Group* – devoted to designing and developing semantic database technology; *Applications Group* – devoted to investigating spatial data technology and applications and GIS; *Heterogeneous Database Group* – devoted to deepening research in distributed heterogeneous databases; and *Semantic-Relational Systems Group* – devoted to making the semantic database technology available to all database users.

Grants Awarded: FIU HPDRC has been awarded an additional two years of funding for our NASA Institutional Research Award. FIU has been awarded a wide-ranging grant from the Office of Naval Research entitled "Educational Innovations in Science, Engineering, and Mathematics, with a Comprehensive Student Pipeline: From pre-College to Graduate Studies, Promoting Recruitment, Retention and Academic Excellence;" Naphtali Rische is one of the Co-PI's of this award. FIU HPDRC has been awarded additional funds to continue its work in developing better database applications for researchers at Everglades National Park. FIU HPDRC continues to receive industrial support. This year Leighland Bridge Holdings awarded a grant to HPDRC to perform research leading to new methods to extract data from the WWW.

Outreach Program to Schools: During the Summer 1999 semester, FIU's College of Education held two five-day intensive in-service training workshops entitled "Databases for School, Life and Space" for

teachers. The workshops were organized as part of the College of Education's commitment to developing workshops for math, science, and technology middle school and high school teachers in the Miami-Dade County Public School System. The main objective of these workshops was to impact the pipeline for potential computer science students by creating awareness among teachers about the FIU School of Computer Science and the HPDRC. The participants were lead through a intensive training sessions on Microsoft Access. Access was chosen because, as a database program, it provides the basic foundation needed for teachers to begin to understand how databases are used in the HPDRC. Furthermore, by teaching teachers how to use databases, it is hoped that they will show their students how to use databases. The increase of knowledge related to databases is another way this program is attempting to "prime the pump" and create database and computer applications awareness throughout the FIU local feeder pattern. Workshop participants visited the HPDRC for one afternoon of each workshop. Dr. Naphtali Rische welcomed the participants, explained how the HPDRC has been involved in the advancement of database technology, and provided an overview of ongoing HPDRC projects, including the Everglades Research Project and TerraFly. Debra Davis, a computer science graduate student performing research at HPDRC, provided a presentation on TerraFly. She demonstrated the program and explained how database technology is being used for similar applications. She described how teachers could use the technology and provided teachers with the CD version of the program, then answered questions about database technology and TerraFly.

We have continued to develop an outreach program that will ultimately consist of both visits to FIU and a traveling 'show' that includes a presentation geared to the appropriate audience at schools. The presentation is followed by a hands-on demonstration of interesting database projects to which the students can relate, such as advanced 'virtual reality' demonstrations and the like. One aspect of this show is viewing a South Florida Landsat image through which it is possible to 'fly' by updating the image in real-time from the semantic database in which the Landsat data is stored. During the Summer 1999 semester, FIU's College of Education and the HPDRC worked closely to develop a half-day seminar, the first of its kind for FIU, for minority students interested in computer applications. Seventy students attending Miami Central High School (Miami-Dade County's magnet school for computer science) were invited to FIU for this "Day of Databases." The students were welcomed and provided with practical information by FIU staff members from the College of Education, the Office of Admissions, the School of Computer Science, and the HPDRC. Staff members from the FIU Office of Admissions spoke to the students and provided campus tours. Admissions staff members engaged the students in an information-packed orientation session during which catalogues were handed out and many of the students' questions were answered. More than 30 students submitted cards requesting more information about FIU. After the tour, Dr. William Kraynek, Associate Director for the School of Computer Science, promoted the possibilities and benefits of studying at the School of Computer Science. Several students expressed an interest in studying computer science and asked questions ranging from the amount of math required for computer science students to the differences between computer engineering and computer science. After Dr. Kraynek's presentation, Dr. Maxim Chekmasov (HPDRC's Acting Manager), spoke about various uses for databases and the research being conducted at HPDRC. Debra Davis, a computer science graduate student performing research at HPDRC, then conducted a presentation on TerraFly, demonstrating how it uses database information to create visual images of geography. The HPDRC session ended with the introduction of four teenaged HPDRC student researchers. Several of these students began working with the HPDRC while they were in high school. The students discussed the projects they are working on and encouraged the participants to pursue their dreams of attending college.

FIU's High Performance Database Center hosted seven students from the Miami-Dade Public Schools under the school system's Advanced Academic Internship Program (AAIP) during the 1998-1999 school year as well as seven additional students during the 1999-2000 school year. These students worked alongside researchers at HPDRC and contributed to the research and development goals of the Center. After serving as an intern, Roy Duque de Estrada graduated from Hialeah-Miami Lakes High and has enrolled as a freshman at FIU. He is presently volunteering at HPDRC to stay involved in cutting-

edge computer science research; we hope to hire him to a paying position in the near future. The involvement of these students provides us a pipeline of researchers from high school through the PhD.

Course Development: The courses proposed in our response to the site visit have been approved as experimental by the School of Computer Science's curriculum committee. Dr. Shu-Ching Chen, an Assistant Professor hired to the School of Computer Science in Fall 1999 and an affiliate of HPDRC has had a new course relevant to the research mentioned above approved by the Computer Science curriculum committee. "Distributed Multimedia Database Systems and Information Systems," will be a 6000 level course offered as an elective to MS and PhD students. Dr. Chen and Dr. Prabakar (formerly Dr. Prabhakaran; also an HPDRC affiliate) will take part in developing a course on Computer Networks to be taught at the 5000 level. This course should prove useful to students performing distributed database research.

Components and Materials Required and Indications of Success

Infrastructure Additions: The infrastructure acquired is being used every day by the student and faculty researchers. Additional acquisitions are planned before the end of the current award period.

REU Supplement: HPDRC requested an REU supplement to our MII grant. The students supported, in part, by the REU supplement are detailed in the Immediate Impact section.

Students Supported directly by MII: Twenty-two graduate and undergraduate students have been directly supported, in part, by our MII grant during the past year; forty-seven have been supported since the grant's inception. The majority of these have been members of under-represented groups. Students supported by our MII grant are detailed in the Immediate Impact section.

Publications: During the past year, we published 21 items under the support of this grant. Additionally, a patent covering parts of our Semantic Database technology has been awarded: Patent Number 5920957, awarded July 6, 1999, "Efficient Optimistic Concurrency Control and Lazy Queries for Databases and B-Trees."

Evaluation

Degree of Success

Toward the goal of better recruiting and retaining under-represented minority students in our graduate programs, we have successfully increased supported student enrollment as a result of the support from our MII grant. Several Affinity Groups and an Outreach Program are in place to enhance our teaching environment and graduate recruitment. We have worked progressively towards the design and analysis of the heterogeneous database system. The effort led to the publication of several technical papers. Appropriate facilities have been acquired and are being constantly enhanced to support students and the research. These activities and achievements evidence a great success in fulfilling the grant's goals and a continuing improvement over the first two years' results.

Outcome

The semantic object-oriented database (Sem-ODB) under development at Florida International University's High Performance Database Research Center (FIU HPDRC) is a general-purpose DBMS that supports a wide spectrum of applications ranging from transaction-oriented to decision-support applications. A Multi-user Semantic Database Engine is operational. A main goal of our work has been to achieve the quality that would make the SDB server viable as a commercial product. Additional features that have to be available in every commercial database system (integrity constraint checking, backup-recovery features, administrative tools, performance monitors, etc.) have been implemented. The theory necessary to support database versions has been developed and implemented. This allows each client to be fully isolated from all other clients and for each client to operate against a stable and consistent database. This feature is especially useful and efficient when used in Data Warehouse applications.

Scientific data sometimes needs to be distributed across the network. Each site might maintain an individual database with high autonomy. Nevertheless, data from several component databases may need to be combined in order to fulfill the requirements of an application. It is unavoidable that heterogeneity, which is partially caused by different data models or different query interfaces, will exist among component databases. Even if component databases implement the same data model, there is usually a need for us to deal with schematic or semantic conflicts. Such conflicts include structural conflicts, naming conflicts, abstraction conflicts, domain conflicts, etc. In order to meet all these requirements, we are building a heterogeneous database system (HDB) that extends our semantic binary database. Users can access a number of databases with the extended system as if there were only one semantic database because we use the Semantic Object-oriented Data Model to construct a global data schema. The query interfaces of our HDB are exactly the same as those provided by the Semantic Object-Oriented Database that is under development at HPDRC. The query processor and optimizer in our HDB are responsible for decomposing the global queries into one or more sub-queries. The sub-queries are then submitted to the corresponding component databases via ODBC. When the sub-queries are completely processed, their answers are re-assembled and reconciled at the destination site.

Impact

The exploding growth and use of the Internet and World Wide Web have enabled users to access huge volumes of data with unprecedented convenience and speed. However, the data sources often diverge in their data model (how the data are organized) and their retrieval interface (how the data can be queried). The deployment of a heterogeneous database will greatly benefit the users in translating isolated, multi-sourced data into integrative information. Focusing on reconciliation of text as well as geospatial data, our project will have a great impact on better facilitating earth scientists in collecting and integrating environmental data (images, maps, and texts) for analysis.

Immediate Impact

Students: The following undergraduates have been supported, in part, by our MII grant: Enrique Almendral, Abraham Anzardo, Michael Armentano, Jorge Besada, Joel Delgado, Jorge Du Quesne*, Julie Fernandez, Alejandro Gonzalez, Freddy Haayen, Robert Hazbun, Alexander Hernandez, Jose Iglesias, Alexander Jelinek, Sheldon Jones, Ying Liu, Daniel Mendez, Reinaldo Morejon, Luis Pachas, Wilbis Padron, Guido Pozo, Patrick Quinlivan, Dario Rivera*, Gianina Rocha, Julio Ruano*, and Roberto Valenti. All but one of these students are from under-represented groups; those marked with an * received their B.S. degrees during the past year. The following graduate students have been supported, in part, by our MII grant: Elma Alvarez, Debra Davis-Chu, Guillermo Fernandez, Dario Gonzalez, Scott Graham, Martha Gutierrez, Tin Ho*, Guangyi Li, Rebecca Miro, Khaled Naboulsi, Philippe Pardo, Joseph Pontillo, Steve Rios, and Rosany Rodriguez. All but four of these students are from under-represented groups; those marked with an * received their Master's degrees during the past year. The following undergraduates have been supported, in part, by the REU supplement to our MII grant: Abraham Anzardo, Juan Carlos Carrillo, Luis Espinal, Luis Llanes, Daniel Mendez, Michael Olivero, Jose Obando, Sebastian Ojanguren, Wilbis Padron, Oscar Parrales*, and Rob Valenti. All of these students are from under-represented groups; those marked with an * received their Bachelor's degrees during the past year.

Publications: 21 publications this year acknowledge the support of our MII award, including:

- N. Rische, K. Naboulsi, O. Wolfson, and B. Ehlmann. "An Efficient Web-based Semantic SQL Query Generator." The 19th IEEE International Conference on Distributed Computing Systems Workshop. Austin, TX, May 31 - June 5, 1999, pp. 23-30.
- N. Rische, A. Vaschillo, D. Vasilevsky, A. Shaposhnikov, S. Chen. "The Architecture for Semantic Data Access to Heterogeneous Information Sources". CATA-2000 ISCA "15th International Conference on Computers and Their Applications", New Orleans, Louisiana. pp. 134-139.
- N. Rische, J. Yuan, R. Athauda, X. Lu, X. Ma. "SemWrap: A Semantic Wrapper Over Relational Databases, With Substantial Size Reduction of User's SQL Queries". EDBT 2000 "Seventh International Conference on Extending Database Technology", Konstanz, Germany. pp. 13-14.