

Scientific Data Management: An Orphan in the Database Community?

Increasingly, scientific discovery relies on querying vast amount of information for correlations and comparisons. Scientists in biology, astronomy, medicine, etc. are assembling databases that are commonly hundreds of terabytes. Petascale databases will become the norm in the next ten years for disciplines as disparate as astronomy, biology, environmental engineering, geophysics, hydrology, oceanography, and medicine (just to name a few!). At the same time, interdisciplinary research between Computer Science and other science and engineering disciplines lies at the forefront of the National research agenda, as evidenced by the National Academies Keck Futures Initiative and by the creation of the National Science Foundation's Office of Cyberinfrastructure.

Despite these factors, the database community has been slow to extend its research mission to include scientific database applications. Specifically, there is little support for interdisciplinary work in program committees, journals, and in hiring. Scientific discovery relies critically on database technologies, such as data mining, indexing and data organization, query processing, schema representation and ontology, and stream processing. Yet, scientists often find themselves building their own solutions without the involvement of DB researchers.

The panel will address both the positive and negative aspects for researchers who choose to work in scientific database applications, which include: the relative ease of acquiring federal grants for interdisciplinary scientific database work; the difficulty of getting publications about scientific database applications into top-tier conferences; the barriers to inter-disciplinary research with a focus on the extra demands it places on graduate student training and on the reduction in volume of research output; and the limited availability of tenure-track faculty positions for scientific-database researchers in Computer Science departments.

The panel assembles leading experts in the application of databases to scientific computing problems. They will discuss the trends and opportunities in scientific database applications and will debate what actions the database community should take in response.



Randal Burns (Moderator)

Randal Burns is an Assistant Professor in the Department of Computer Science at the Johns Hopkins University. He earned his PhD in 2000 from the University of California Santa Cruz. His research interests include data organization and indexing, distributed query processing, storage security, and data protection. Randal received the

NSF CAREER award in 2003 and the DoE Early Career Principal Investigator award in 2002. He is the holder of 12 US patents.



Susan B. Davidson

Susan B. Davidson received the B.A. degree in Mathematics from Cornell University, Ithaca, NY, in 1978, and the M.A. and Ph.D. degrees in Electrical Engineering and Computer Science from Princeton University, Princeton NJ, in 1980 and 1982. Dr. Davidson joined the University of Pennsylvania in 1982, and is now the Department Chair and Weiss

Professor of Computer and Information Science. She is an ACM Fellow, a Fulbright scholar, founding co-Director of the Center for Bioinformatics at UPenn (PCBI), and recently stepped down as Deputy Dean of the School of Engineering and Applied Science (SEAS). Dr. Davidson's research interests include database systems, database modeling, distributed systems, and bioinformatics. Within bioinformatics she is best known for her work in data integration, XML query technologies, and provenance in workflow systems.



Yannis Ioannidis

Yannis Ioannidis is currently a Professor at the Department of Informatics & Telecommunications of the University of Athens. He received his Diploma in Electrical Engineering from the National Technical University of Athens in 1982, his MSc in Applied Mathematics from Harvard University, and his Ph.D. degree in Computer

Science from the University of California at Berkeley in 1986. Immediately after that he joined the faculty of the Computer Sciences Department of the University of Wisconsin at Madison, where he became a Professor before finally leaving in 1999. His research interests include database and information systems, digital libraries, personalization, scientific systems and workflows, eHealth systems, and human-computer interaction. Yannis is an ACM Fellow, and recipient of several awards including the VLDB "10-Year Best Paper Award", the NSF PYI award, and several teaching awards. He currently serves as the ACM SIGMOD Vice-Chair and is a member of the Scientific Advisory Board of the Max Planck Institute for Informatics. Between July 2002 and March 2004 he served as the Information Technology advisor to the Minister of Health of Greece.



Miron Livny

Miron Livny received a B.Sc. degree in Physics and Mathematics in 1975 from the Hebrew University and M.Sc. and Ph.D. degrees in Computer Science from the Weizmann Institute of Science in 1978 and 1984, respectively. Since 1983 he has been on the Computer Sciences Department faculty at the University of

Wisconsin-Madison, where he is currently a Professor of Computer Sciences, the director of the Center for High Throughput Computing and is leading the Condor project. Dr. Livny's research focuses on distributed processing and data management systems and data visualization environments. His recent work includes the Condor distributed resource management system, the DEVise data visualization and exploration environment and the BMRB repository for data from NMR spectroscopy



Jignesh M. Patel

Jignesh M. Patel is an Associate Professor at the University of Michigan. He received his PhD from the University of Wisconsin in 1998. Since 1999 he has been a faculty member in the EECS department at the University of Michigan, where a major focus of his research has been on developing database management techniques for life sciences.

He is the recipient of an NSF Career Award, multiple Microsoft eScience Awards, and multiple IBM Faculty Awards.