

List of Panels

Contributed by Pablo Guerrero
Wednesday, 23 January 2008
Last Updated Wednesday, 13 February 2008

- Declarative, Domain-Specific Languages -
Elegant Simplicity or a Hammer in Search of a Nail?

Sam Madden, MIT; Alan Demers, Cornell University;
Michael Carey, BEA; Boon Thau Loo, University of
Pennsylvania; John Whaley, moka5

- Scientific Data Management: An Orphan in
the Database Community?

Randal Burns (moderator), Johns Hopkins University; Susan B.
Davidson, Cornell University; Yannis Ioannidis, University
of Athens; Miron Livny, University of Wisconsin-Madison;
Jignesh M. Patel, University of Michigan

- Cloud Computing-Was Thomas Watson Right
After All?

Raghu Ramakrishnan (moderator), Yahoo!; Eric Baldeschwieler, Yahoo!;
James Hamilton, Microsoft; Miron Livny, University
Wisconsin-Madison; Yossi Matias, Google; Hamid Pirahesh, IBM

Panel #1:

Declarative, Domain-Specific Languages - Elegant Simplicity or a
Hammer in Search of a Nail?

Sam Madden (moderator), MIT; Alan Demers, Cornell
University; Michael Carey, BEA; Boon Thau Loo, University
of Pennsylvania; John Whaley, moka5

Duration: 1.5 hours

Tuesday Apr 8th, 14.00-15.30

Location: Coba

Recent years have seen a proliferation in proposals for database-inspired declarative languages for a number of non-database domains, including sensor networks, network protocol specification, event monitoring, program analysis, computer games, and the design of three-tier applications. The typical argument put forth in these proposals is that a declarative language is simple and hides many of the ugly details that are inherent in whatever domain is being addressed. For example, in sensor networks, a declarative language might allow users to specify what data they want from a collection of sensors without worrying about low-level details like power management or network organization. For network protocols,

a set of declarative rules can be used to specify how routing decisions should be made, allowing the protocol designer to avoid worrying about low-level decisions regarding how to maintain and propagate changes to distributed state like routing tables.

Given that these languages are usually built around a simple, limited core, a common criticism is that they are very restrictive in terms of the class of programs that can be expressed. In response, researchers inevitably glue additional language constructs and 'features' onto the basic systems to allow them to address a broader class of applications, increase extensibility, or to satisfy competing concerns. As a result, the final languages that get built are complex pseudo-declarative Franken-languages, which are arguably less easy to use than the more full-featured, conventional approaches they were designed to replace.

In this panel, panelists will be given five minutes to review their own work in the area or to highlight some of their favourite features or the most notable shortcomings of their co-panelists work. This will be followed by a moderated question and answer session designed to explore the future of this line of research and provoke interesting discussions.

Questions we will address include:

- Is this database research?
- What do we have to offer in other domains?
- Are those domains listening to us? How do we have impact?
- What is the advantage compared to other approaches?

Alan Demers is a senior research associate at Cornell University. Alan's current research focuses on database techniques for non-database applications including computer games and the design of data-driven web applications.

Michael Carey is a Senior Engineering Director at BEA Systems, Inc. and chief architect for the BEA AquaLogic Data Services Platform. Mike spent a dozen years on the University of Wisconsin-Madison faculty.

Sam Madden (panel moderator) is on the faculty at MIT. Sam's research focuses on data warehousing and query processing over streams, sensors, and other networked information sources.

Boon Thau Loo is on the faculty at the University of Pennsylvania. His research focuses on distributed data management, Internet-scale query processing, and the application of database technologies to networked systems.

John Whaley is one of the founders of moka5. His virtual machine and compiler infrastructure 'Joeq' and BDD and Datalog-based program analysis engine 'bdbddb' are widely used for program analysis research and education.

Panel #2:

Scientific Data Management: An Orphan in the Database Community?

Randal Burns (moderator), Johns Hopkins University; Susan B. Davidson, Cornell University; Yannis Ioannidis, University of Athens; Miron Livny, University of Wisconsin-Madison; Jignesh M. Patel, University of Michigan

Duration: 1.5 hours

Wednesday Apr 9th, 14.00-15.30

Location: Uxmal

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) 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 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 is currently a Professor at the Department of Informatics and 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 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 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.

Panel #3:

Cloud Computing-Was Thomas Watson Right After All?
Raghu Ramakrishnan (moderator), Yahoo!; Eric Baldeschwieler, Yahoo!; James Hamilton, Microsoft; Miron Livny, University Wisconsin-Madison; Yossi Matias, Google; Hamid Pirahesh, IBM

Duration: 1.5 hours

Friday Apr 11th, 11.00-12.30

Location: Coba

The infamous prediction "I think there is a world market for maybe five computers," was allegedly made by Thomas Watson, Sr. Perhaps he was merely a little ahead of his time, if he was referring to farms of computers accessed over networks, i.e., "computers in the clouds." Mainframes once dominated the computing landscape, drawing programmers from afar to submit their applications, along with burnt offerings (especially if bugs got into the vacuum tubes). The old order changed, however, and mainframes gave way to a myriad diaspora that took computing power to programmers' desktops.

As the cost of maintaining and securing computing environments surges,

are we witnessing another turn of the wheel, one in which computing power will again be concentrated in a handful of computing clusters, and applications worldwide are invoked over the Internet and intranets?

Grid computing is arguably the first sign of this evolution, and has established itself over the past decade as a cost-effective approach to high-throughput CPU-intensive computing. In recent years, the Map-Reduce approach to scalable aggregation of very large datasets on clusters of machines-Group By on steroids, for the SQL junkies among you-has grown in popularity. Many companies have long used massive clusters to support a get-put interface that virtualizes disk accesses, and provides lightweight, scalable data stores for high throughput mixtures of updates and queries.

This interface is now being supported as a web service, and richer support for such data-intensive "transactional" workloads are on the horizon.

These three approaches all underscore a common theme, the concentration of computing power where it is economical to administer it, in the new incarnation of a mainframe, the "computing cloud." In the meantime, the business models of hosted applications and hosted infrastructures are becoming widely accepted, showing a clear pathway to monetizing cloud computing.

In this panel, we will discuss these trends, agree vehemently on occasion, disagree vehemently more often (if only to entertain you), and in general will strive to make your life more interesting and fulfilled.

Raghu Ramakrishnan (moderator), is Chief Scientist for Audience and Research Fellow at Yahoo!, where he is working on the Sherpa hosted data services platform. He is on leave from the Computer Sciences Department at the University of Wisconsin-Madison.

Eric Baldeschwieler is Director of the Yahoo! Hadoop Team, which is the largest contributor to the Hadoop project, and designs, builds and supports very large computation grids based on Hadoop and other open source tools.

James Hamilton is Architect for Microsoft Live Platform Services, working on multi-tenant hosted systems and massively parallel data management. He was previously General Manager of the Microsoft Exchange Hosted Services team, and SQL Server Architect.

Miron Livny is Professor of Computer Sciences at University of Wisconsin-Madison, and heads the Condor Grid

Computing project, which supports high-throughput computing on large collections of distributively owned computing resources.

Yossi Matias is Director of Google's Tel Aviv R&D Center, and is on leave from the School of Computer Science at Tel Aviv University. He is on the founding steering committee of the Israeli Academic Grid and was a member of a national committee evaluating strategies for Grid technologies.

Hamid Pirahesh is an IBM Fellow and the manager of the DataBase Technology Institute at IBM Almaden. He has direct responsibilities in various aspects of IBM's DB2 product, and his research interests include Web services and information integration in Web based federated and distributed systems.