Acknowledgements

The SIGMOD/PODS 2001 Joint Conference gratefully acknowledges the contributions of the following institutions.

Principal Sponsor:

ORACLE

Sponsor:

Microsoft

Contributors:

HEWLETT PACKARD

Informix

Supporting Organization:

Persistent Systems Pvt. Ltd.

In addition, the joint conference acknowledges the contributions of bags by

IBM

and T-shirts for the conference organizers by the
Department of Computer Science
University of California, Santa Barbara



Final Program for the Joint Conference of

2001 ACM SIGMOD International Conference on Management of Data

and the

Twentieth ACM SIGMOD-SIGACT-SIGART Symposium of Principles of Database Systems (PODS)

Fess Parker's Doubletree Resort
Santa Barbara, California
May 21-24, 2001



Note

Welcome Message from the Chairs

SIGMOD and PODS started as separate conferences with overlapping aims but different emphasis: SIGMOD (whose first conference took place in 1975) has always had a more applied focus on database technology, while PODS (initiated in 1982) has focused on the fundamental theory underlying the database field. Over the years, however, it became apparent that the two communities are highly symbiotic and could benefit from increased interaction. Reflecting this, and to accommodate the many people who wished to attend both conferences, PODS and SIGMOD have been held concurrently since 1991 as one "Joint Conference".

There is a single registration process for the joint conference, and all registrants are invited to attend the social events (receptions, lunches and banquet). In addition, although PODS and SIGMOD have separate program committees and separate proceedings, registrants may attend any of the sessions - PODS or SIGMOD - during the four-day joint conference. PODS and SIGMOD overlap for two days, but each has a separate third day without overlap. Registrants are also encouraged to attend the two affiliated workshops being held in association with the main conference.

The SIGMOD/PODS technical programs will continue to be exciting, and reflect the rapid evolution of database technology in response to the need for handling increasingly complex and massive information. We hope that the collaboration that will result from this joint conference will stimulate this process. We encourage you to join us in Santa Barbara and play a part in this exciting and enriching event.

Jianwen Su
SIGMOD General Chair

Serge Abiteboul
PODS General Chair

Timos Sellis
SIGMOD PC Chair

Peter Buneman
PODS PC Chair

Program Highlights

Common Events

- SIGMOD/PODS keynote address by *Dr. Gene Myers*, *Director of Informatics Research, Celera Genomics:* Whole Genome Assembly: Tactical and Strategic Implications.
- Keynote speech by *Steve Kirsch, Founder and Chief Executive Officer, Propel:* To the Internet and Beyond: Database Challenges for New Applications

PODS

- Presentation of PODS Awards
  - Best Paper Award: *Optimal Aggregation Algorithms for Middleware*, Ronald Fagin (IBM Almaden), Amnon Lotem (Univ. of Maryland), and Moni Naor (Weizmann Institute of Science)
  - Best Newcomer paper Award: *Relaxed Multi-Way Trees with Group Updates*, Kim S. Larsen (Univ. of Southern Denmark)
- 26 research papers
- Invited talk by Victor Vianu (UC San Diego): *A Web Odyssey: From Codd to XML*
- 2 invited tutorials
  - *The Parameterized Complexity of Database Queries*, Martin Grohe (University of Illinois at Chicago)
  - *The Challenges of Delivering Content on the Internet*, Tom Leighton (Akamai & MIT)

SIGMOD

- Presentation of SIGMOD awards including:
  - Best paper award: *Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases*, Eamonn Keogh, Kaushik Chakrabarti, Sharad Mehrotra, and Michael Pazzani (UC Irvine)
- 44 research papers divided in fifteen sessions
- 11 industrial papers divided into four sessions
- 21 prototype demonstrations divided into three sessions:
  - XML and Web
  - Schemas, Models and Processes
  - Content-based Retrieval, Web and Data
- 3 panel sessions:
  - XML Data Management: Go Native or Spruce up Relational Systems?
  - Will Database Researchers Have ANY Role in Data Security?
  - Application Servers: Born-Again TP Monitors for the Web?
- 5 tutorials:
  - *Online Query Processing*
    Peter Haas (IBM Almaden) and Joe Hellerstein (UC Berkeley)

Note

Note

Program Highlights

- *Semantic B2B Integration*
  Christoph Bussler (Oracle)
- *Time Series Similarity Measures*
  Dimitrios Gunopulos (UC Riverside) and Gautam Das (Microsoft Research)
- *Models and Languages for Describing and Discovering E-Services*
  Fabio Casati and Ming-Chien Shan (HP Labs.)
- *MPEG-7 Standard for Multimedia Databases*
  R. Smith (IBM T. J. Watson Research Center)

Companion Workshops

- Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2001)
- 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE '01)
- Fourth International Workshop on the Web and Databases (WebDB 2001)
- Workshop on Network-Related Data Management (NRDM 2001)

Sunday, May 20

9:00–18:00

2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE01)

The need to access data regardless of location, along with the proliferation of new personal information services and applications, has necessitated new approaches to the management and access of data by mobile or roaming users. Besides advances in communications and hardware, achieving pervasive mobile computing requires innovative theories, paradigms and applications in data management. MobiDE2001 will provide a single forum for researchers and technologists to discuss the state-of-the-art, present their contributions, and set future directions in data management for mobile and wireless access. To this end, the workshop is being organized in a manner designed to foster interaction and exchange of ideas among the participants.

Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2001)

Now five years old, the Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD) has fostered discussion and investigation of data mining research issues pertinent to large databases and data warehouses. In this time, data mining as a discipline has matured considerably. Particularly strong progress has been made in the design of scalable algorithms that transform oceans of bits into interpretable patterns and predictive models. This year's sixth DMKD workshop is aimed at discussing the next generation of data mining research, with the goal of bringing together researchers and experienced practitioners from academia and industry. The atmosphere of the workshop will be informal, fostering interaction through short presentations and open discussion. The focus of this year's program is on new applications of data mining, industrial experiences with data mining projects, limitations of current data mining tools, intersection of data mining with other disciplines, and research visions.

18:00–20:30

SIGMOD/PODS Early Registration (Santa Ynez Foyer)

19:00–20:30

SIGMOD/PODS Welcome Reception (Plaza Del Sol)

Local Information

Santa Barbara is a “get out in the sunshine and fresh air” place, with all the activities that implies. The average annual rainfall is about 15 inches, and the average rainfall for May is only 0.2 inches.

Santa Barbara has a unique geographic situation. For roughly 40 miles on either side of Santa Barbara, the Pacific coastline runs almost due East/West. This alignment moderates the ocean currents that flow past Santa Barbara’s front door. Towering mountains, incredibly close to the sea, shelter the coastal strip from inland heat and cold. The result is a comfortable year-round average temperature of 64 degrees, with an average maximum of 74 and an average minimum of 56 degrees. The highs and lows in May are around 70 and 50 degrees, respectively. For May 21, the record high is 78 degrees (1989) and the record low is 42 degrees (1942).

Points of interest:

- Santa Barbara County Courthouse. Timeless Spanish-Moorish masterpieces with colorful corridors, historic mural room, 85 ft. observation tower. Great views from the clock tower. 1100 Anacapa St.
- Zoological Gardens. Over 700 animals on the grounds of a former beachside estate. 500 Ninos Dr.
- El Presidio De Santa Barbara Historic Park. The Presidio includes buildings that were part of the last Spanish military outpost in the New World. The chapel contains restored 18th century decorations. 123 E. Canon Perdido St.
- ARLINGTON CENTER FOR THE PERFORMING ARTS. Originally built in 1875 the interior contains a mock Spanish village. Curved ceiling is "star studded" and creates an outdoor illusion. 1317 State St.
- Santa Barbara County is a world-class touring and wine tasting destination, renowned for distinguished wines and golden countryside dotted with quaint western towns and friendly locals. Today's wine boom centers in the Santa Ynez and Santa Maria Valleys, boasting 56 wineries. Acclaimed for its Chardonnay and Pinot Noir varietals, the area also produces Merlot, Cabernet Franc, and Rhone varietals such as Syrah.
- Santa Barbara Historical Museum. Charming adobe complex, built in 1817, houses one of the finest collections of regional history in California. 136 E. De La Guerra St.
- STEARNS WHARF & SEA CENTER. Oldest operating wharf on the West coast. Sea Center offers a glimpse into marine life that inhabits the Santa Barbara Channel. Nature Conservancy includes displays dedicated to the preservation of rare and endangered species and habitat types. At the foot of State St.
- SANTA BARBARA MUSEUM OF NATURAL HISTORY. Nationally renowned museum specializing in North American West Coast and California natural history. Special theme exhibits. 2559 Puesta del Sol Rd.

Program at Glance

WEDNESDAY	9:00	PODS Research 7: *Aggregates*
		SIGMOD Research 7: *Histograms*
		SIGMOD Research 8: *Performance*
		SIGMOD Tutorial 3: *Semantic B2B Integration*
		SIGMOD Demo 3: *Content-based Retrieval, Web and Data Mining*
	11:00	PODS Invited Tutorial 2: *The Challenges of Delivering Content on the Internet*
		SIGMOD Research 9: *Compression*
		SIGMOD Research 10: *Materialized Views*
		SIGMOD Panel 2: *Will Database Researchers Have ANY Role in Data Security?*
		SIGMOD Industrial 2: *Catalog Integration*
	12:30	SIGMOD Business Lunch
	14:30	PODS Research 8: *Data Mining*
		SIGMOD Research 11: *Caching and Buffer Management*
		SIGMOD Tutorial 4: *Models and Languages for Describing and Discovering E-Services*
		SIGMOD Industrial 3: *Applications & Lessons Learned*
		SIGMOD Demo 4: Same as Demo 2
	16:30	PODS Research 9: *Indexing/Transaction*
		SIGMOD Research 12: *Multi-dimensional Data*
		SIGMOD Research 13: *XML*
		SIGMOD Industrial 4: *Database Internals*
		SIGMOD Demo 5: Same as Demo 1
	19:00	California Barbeque

THURSDAY	9:00	SIGMOD/PODS Keynote Address
	11:00	SIGMOD Research 14: *Sampling and Association Rules*
		SIGMOD Research 15: *Schemas and Queries*
		SIGMOD Tutorial 5: *MPEG-7 Standard for Multimedia Databases*
		SIGMOD Panel 3: *Application Servers: Born-Again TP Monitors for the Web?*
		SIGMOD Demo 6: Same as Demo 3

Monday, May 21

7:30–9:00

Continental breakfast (Santa Ynez Foyer)

9:00–10:00

PODS Invited Talk: A Web Odyssey: From Codd to XML (Santa Ynez/San Rafael)

Chair: Peter Buneman (Univ. of Pennsylvania)

The Web presents the database area with vast opportunities and commensurate challenges. This talk discusses some of the developments related to the Web from the viewpoint of database theory. As will be argued, the Web scenario requires revisiting some of the basic assumptions of the area. We are thus at an important juncture. Database theory could retain its classical focus and turn inwards. Or, it could attempt to take heads-on the challenge of the Web and contribute to an important part of its formal foundations. To do so, it will have to leave its familiar shores and reinvent itself. This exciting journey of rediscovery is already under way.

Victor Vianu is a Professor of Computer Science at the University of California at San Diego. He is the author of over seventy research articles in the database area, as well as co-author of a graduate textbook on database theory. He has served as the chair of the executive committee of PODS, as well as general chair and program chair. He was a member of the SIGMOD executive committee, and has served as general chair of the SIGMOD conference. Victor Vianu is area editor for database systems and theory of the Journal of the ACM, and also serves on the editorial boards of the ACM Transactions on Computational Logic and of Discrete Mathematics and Theoretical Computer Science.

10:00–10:30

Coffee break (Santa Ynez Foyer)

10:30–12:00

PODS Research Session 1: Querying Web Data (Santa Ynez/San Rafael)

Chair: Frank Neven (Limburgs Universitair Centrum)

Querying Websites Using Compact Skeletons, Anand Rajaraman (Cambrian Ventures) and Jeffrey D. Ullman (Stanford Univ.)

Dynamically Distributed Query Evaluation, Trevor Jim (AT&T Labs) and Dan Suciu (Univ. of Washington)

Flexible Queries Over Semistructured Data, Yaron Kanza and Yehoshua Sagiv (Hebrew Univ.)

12:00–13:30

Lunch (on your own)

13:30–15:00

PODS Research Session 2: Queries/Optimization (Santa Ynez/San Rafael)

Chair: Luc Segoufin (INRIA)

Multiobjective Query Optimization, Christos H. Papadimitriou (UC Berkeley) and Mihalis Yannakakis (Bell Labs)

Monday, May 21

Pipelining in Multi-Query Optimization, Nilesh N. Dalvi (Indian Inst. of Tech.), Sumit K. Sanghai (Indian Inst. of Tech.), Prasan Roy (Bell Labs), and S. Sudarshan (Indian Inst. of Tech.)

Optimization of Sequence Queries in Database Systems, Reza Sadri (UCLA), Carlo Zaniolo (UCLA), Amir Zarkesh (ZAIAS Tech.), and Jafar Adibi (ZAIAS Tech.)

15:00–15:15

Coffee Break (Santa Ynez Foyer)

15:15–16:45

PODS Invited Tutorial 1: The Parameterized Complexity of Database Queries (Santa Ynez/San Rafael)
Chair: Phokion Kolaitis (UC Santa Cruz)

Parameterized complexity theory, a relatively new branch of complexity theory, allows a refined analysis of the complexity of database query evaluation taking into account that typically one evaluates small queries against large databases. This analysis does not only give us a better understanding of the complexity of query evaluation, but also suggests new evaluation algorithms. Such algorithms, mostly based on different types of tree-decompositions of queries or database instances, have received a lot of attention in recent years.
In this tutorial, I will give a brief introduction into parameterized complexity theory aimed towards database applications. Then I will explain the new algorithmic techniques that lead to efficient evaluation algorithms for various classes of queries.

Martin Grohe is an Assistant Professor in the Department of Mathematics, Statistics. and Computer Science of the University of Illinois at Chicago. His research interests are in algorithms and complexity theory, logic in computer science, and database theory. He received his PhD from the University of Freiburg, Germany, in 1994.

17:00–18:00

PODS Research Session 3: Award Talks
(Santa Ynez/San Rafael)
Chair: Serge Abiteboul (INRIA)

Relaxed Multi-Way Trees with Group Updates, Kim S. Larsen (Univ. of Southern Denmark)
(Best newcomer award)

Optimal Aggregation Algorithms for Middleware, Ronald Fagin (IBM Almaden), Amnon Lotem (Univ. of Maryland), and Moni Naor (Weizmann Institute of Science)
(Best paper award)

21:00–22:30

PODS Business Meeting (Fiesta)

Program at Glance

MONDAY	9:00	PODS Invited Talk: *Victor Vianu*
	10:30	PODS Research 1: *Querying Web Data*
	13:30	PODS Research 2: *Query/Optimization*
	15:15	PODS Invited Tutorial 1: *The Parameterized Complexity of Database Queries*
	17:00	PODS Award Talks

TUESDAY	9:00	SIGMOD Opening and SIGMOD/PODS Plenary Address
	11:00	PODS Research 4: *XML*
		SIGMOD Research 1: *Cubes and Aggregates*
		SIGMOD Research 2: *Efficient Computation*
		SIGMOD Industrial 1: *Data Management in a Highly-Connected World*
		SIGMOD Tutorial 1: *Online Query Processing*
	12:30	SIGMOD Awards Lunch
	14:30	PODS Research 5: *Partial Information*
		SIGMOD Research 3: *Classification and Clustering*
		SIGMOD Research 4: *Distributed and Heterogeneous Databases*
		SIGMOD Tutorial 2: *Time Series Similarity Measures*
		SIGMOD Demo 1: *XML and Web*
	16:30	PODS Research 6: *Expressiveness/ Complexity*
		SIGMOD Research 5: *Indexing*
		SIGMOD Research 6: *Text Management*
		SIGMOD Panel 1: *XML Data Management: Go Native or Spruce up Relational Systems?*
		SIGMOD Demo 2: *Schemas, Models and Processes*
	19:00	Reception

SIGMOD Committee

Undergraduate Scholarship Program co-Chairs
Mario A. Nascimento (Univ. of Alberta)
Aidong Zhang (SUNY at Buffalo)

Far East Coordinator
Hong-va Leong (Hongkong Polytechnic Univ.)

European Coordinator
Francois Llirbat (INRIA)

Program Committee
Sibel Adali (Rensselaer Polytechnic Inst.)
Brad Adelberg (Northwestern Univ.)
Rakesh Agrawal (IBM Almaden)
Roberto J. Bayardo (IBM Almaden)
Alexandros Biliris (AT&T Labs)
José Blakeley (Microsoft)
Tiziana Catarci (Universita degli Studi di Roma "La Sapienza")
Christos Faloutsos (Carnegie Mellon Univ.)
Daniela Florescu (Propel)
Cesar Galindo-Legaria (Microsoft)
Minos Garofalakis (Bell Labs)
Johannes Gehrke (Cornell Univ.)
Dimitrios Gunopulos (UC Riverside)
Laura Haas (IBM Almaden)
Alon Halevy (Univ. of Washington)
Joe Hellerstein (UC Berkeley)
Yannis Ioannidis (Univ. of Athens)
H. V. Jagadish (Univ. of Michigan)
Matthias Jarke (GMD-FIT & RWTH Aachen)
Martin Kersten (CWI)
Laks V.S. Lakshmanan (Concordia Univ. & IIT)
David Maier (Oregon Graduate Institute)
Sharad Mehrotra (UC Irvine)
Bongki Moon (Univ. of Arizona)
Shojiro Nishio (Osaka Univ.)
Maria Orlowska (Univ. of Queensland)
Dimitris Papadias (Hong Kong Univ. of Sci. & Tech.)
Vishy Poosala (Bell Labs)
Krithi Ramamritham (Univ. of Mass. & IIT Bombay)
Arnon Rosenthal (MITRE)
Kenneth Salem (Univ. of Waterloo)
Hans-Jorg Schek (ETH Zurich)
Michel Scholl (CNAM & INRIA)
Eric Simon (INRIA)
Ramakrishnan Srikant (IBM Almaden)
VS Subrahmanian (Univ. of Maryland)
Dan Suciu (Univ. of Washington)
Yannis Theodoridis (Computer Technology Institute)
Jeff Ullman (Stanford Univ.)
Kyu-Young Whang (Korea Advanced Inst. of Science and Tech.)

Tuesday, May 22

7:30–9:00

Continental Breakfast (Santa Ynez Foyer)

9:00–10:30

SIGMOD Opening and SIGMOD/PODS Plenary Address (Santa Ynez/San Rafael)
Chair: Timos Sellis (National Technical Univ. of Athens)

Whole Genome Assembly: Tactical and Strategic Implications, Gene Myers (Celera Genomics)

The assembly of a shotgun data set of end-sequence reads was considered controversial at the time Celera proposed to apply the method to the Human Genome in mid-1998. Critics claimed that the computation would involve an impossible amount of computer time, that the size and repetitiveness of the genome would confound all attempts at assembly should sufficient computer efficiency be achieved, and that even if an assembly was produced it would be of an extremely poor quality and partial nature.

In 1999 the informatics research team at Celera produced an assembly of the Drosophila genome from a whole genome shotgun data set consisting of 3.2 million reads, 72% of which were paired-end reads from 2Kbp and 10Kbp inserts in a 1 to 1.32 mix. The assembly consisted of completely ordered and oriented contigs covering an estimated 97.2% of the genome with only 1630 gaps of average size 1,415bp.

Results of high quality have now been obtained for the human and mouse genomes and our reconstruction of the human genome has been reported upon in *Science*. We discuss our algorithmic approach, the strategic pros and cons of the method, and the implications for the future of bioinformatics.

Gene Myers is a Professor of Computer Science at the University of Arizona currently on leave at Celera Genomics where he is the Vice President of Informatics Research with the mission of determining the sequence of the entire human genome. He received his Ph.D. in Computer Science from the University of Colorado in 1981. His research interests include design of algorithms, pattern matching, computer graphics, and computational molecular biology. His most recent academic work has focused on algorithms for the central combinatorial problems involved in DNA sequencing, and on a wide range of sequence and pattern comparison problems. Among the tools he has developed are Blast—a widely used tool for protein similarity searches, FAKtory—a system to support DNA sequencing projects, Anrep—a pattern matching language for applications in molecular biology, and Mac- & PC-Molecule—a molecular visualization tool for Apple and Wintel computers. Dr. Myers is an associate editor for the Journal of Computational Biology, and on the editorial boards of BioInformatics and RAIRO Informatique Theorique et Applications.

10:30–11:00

Coffee break (Santa Ynez Foyer)

11:00–12:30

PODS Research Session 4: XML (Sierra Madre North)
Chair: Foto Afrati (National Tech. Univ. of Athens)

On XML integrity constraints in the presence of DTDs, Wenfei Fan (Bell Labs & Temple Univ.) and Leonid Libkin (Univ. of Toronto)

Tuesday, May 22

Extended Path Expressions for XML, Makoto Murata (IBM Tokyo)

XML with Data Values: Typechecking Revisited, Noga Alon (Tel Aviv Univ.), Tova Milo (Tel Aviv Univ.), Frank Neven (Limburgs Universitair Centrum), Dan Suciu (Univ. of Washington), and Victor Vianu (UC San Diego)

SIGMOD Research Session 1: Cubes and Aggregates (Santa Ynez)
Chair: Dimitrios Gunopulos (UC Riverside)

Efficient Computation of Iceberg Cubes with Complex Measures, Jiawei Han (Simon Fraser Univ.), Jian Pei (Simon Fraser Univ.), Guozhu Dong (Wright State Univ.), and Ke Wang (Simon Fraser Univ.)

On Computing Correlated Aggregates Over Continual Data Streams, Johannes Gehrke (Cornell Univ.), Flip Korn (AT&T Labs), and Divesh Srivastava (AT&T Labs)

Iceberg-Cube Computation with PC Clusters, Raymond T. Ng, Alan Wagner, and Yu Yin (Univ. of British Columbia)

SIGMOD Research Session 2: Efficient Computation (San Rafael)
Chair: Sharad Mehrotra (UC Irvine)

Outlier Detection for High Dimensional Data, Charu C. Aggarwal and Philip S. Yu (IBM T. J. Watson Research Center)

Bit-Sliced Index Arithmetic, Denis Rinfret (UMass/Boston), Patrick O'Neil (UMass/Boston & Microsoft), and Elizabeth O'Neil (UMass/Boston & Microsoft)

Space-Efficient Online Computation of Quantile Summaries, Michael Greenwald and Sanjeev Khanna (Univ. of Pennsylvania)

SIGMOD Industrial Session 1: Data Management in a Highly-Connected World (Fiesta)
Chair: Guy Lohman (IBM Almaden)

StorHouse Metanoia - New Applications for Database, Storage & Data Warehousing
Felipe Carino Jr., Pekka Kostamaa, Art Kaufmann, and John Burgess (FileTek)

Enabling Dynamic Content Caching for Database-Driven Web Sites, K. Selcuk Candan, Wen-Syan Li, Qiong Luo, Wang-Pin Hsiung, and Divyakant Agrawal (C&C Research Labs)

The Network is the Database: Data Management for Highly Distributed Systems, Julio C. Navas and Michael Wynblatt (Siemens Tech.)

SIGMOD Committee

General Chair
Jianwen Su (UC Santa Barbara)

Program Chair
Timos Sellis (National Technical Univ. of Athens)

Associate Program Committee Chair
Divyakant Agrawal (UC Santa Barbara)

Proceedings Chair
Sharad Mehrotra (UC Irvine)

Electronic Proceedings Chair
Walid Aref (Purdue Univ.)

Registration Chair
Mei-Ling Liu (California Polytechnic State Univ.)

Finance Chair
Sunil Prabhakar (Purdue Univ.)

Local Arrangement Chair
Edward Chang (UC Santa Barbara)

Corporate Sponsor's Chair
Ambuj Singh (UC Santa Barbara)

Publicity Chair
Dimitrios Gunopulos (UC Riverside)

Web-site Chair
Xiaoyang Sean Wang (George Mason Univ.)

Local Demo Chair
James Frew (UC Santa Barbara)

Demonstrations Chair
Michael Franklin (UC Berkeley)

Demonstrations Committee
Swarup Acharya (Lucent Technologies)
Paul Aoki (Xerox PARC)
Ling Liu (Georgia Inst. of Tech.)
Mary Tork Roth (IBM Almaden)

Industrial Program Chair
Guy Lohman (IBM Almaden)

Industrial Program Committee
Surajit Chaudhuri (Microsoft)
Christoph Freytag (Humboldt Univ.)
C. Mohan (IBM Almaden)
Glenn Paulley (Sybase)
Patrick Valduriez (INRIA)

Panels Chair
Philip A. Bernstein (Microsoft)

Tutorials Chair
Gerhard Weikum (Univ. of the Saarland)

Tutorials Committee
Pamela Drew (Boeing)
Christian Jensen (Aalborg Univ.)
Betty Salzberg (Northeastern Univ.)

PODS Committee

Executive Committee:
Serge Abiteboul (INRIA)
Peter Buneman (Univ. of Pennsylvania)
Georg Gottlob (TU Wien)
Alberto Mendelzon (Univ. of Toronto)
Christos H. Papadimitriou (UC Berkeley)
Victor Vianu (UC San Diego)

General Chair
Serge Abiteboul (INRIA)

Program Chair
Peter Buneman (Univ. of Pennsylvania)

Proceedings Chair
Wenfei Fan (Bell Labs & Temple University)

Program Committee
Foto Afrati (National Tech. Univ. of Athens)
Gustavo Alonso (ETH Zurich)
Paolo Atzeni (Università Roma Tre)
Luca Cardelli (Microsoft Research)
Amr El Abbadi (UC Santa Barbara)
Michael Franklin (UC Berkeley)
Gosta Grahne (Concordia Univ.)
Richard Hull (Bell Lab.)
Neil Immerman (Univ. of Mass. at Amherst)
Phokion G. Kolaitis (UC Santa Cruz)
Heikki Mannila (Nokia Research Center & Helsinki Univ. of Tech.)
Rajeev Motwani (Stanford Univ.)
Frank Neven (Limburgs Universitair Centrum)
Luc Segoufin (INRIA)
Jianwen Su (UC Santa Barbara)
Jennifer Widom (Stanford Univ.)

Local Arrangement Chair
Edward Chang (UC Santa Barbara)

Finance Chair
Sunil Prabhakar (Purdue Univ.)

Publicity Chair
Dimitrios Gunopulos (UC Riverside)

Web-site Chair
Xiaoyang Sean Wang (George Mason Univ.)

Tuesday, May 22

SIGMOD Tutorial 1: Online Query Processing
(Sierra Madre South)
Chair: Gerhard Weikum University of the Saarland)

Techniques for online query processing provide the user with incrementally refining results for long-running queries, and permit "on-the-fly'" control. User studies show that data analysis is an iterative process; online processing techniques are well suited to multi-step analyses because users can dynamically trade off the runtime of each step against the quality of the results seen at that step. Technological pressures also drive the need for online processing: data appetites are growing faster than Moore's Law, and hence computing big queries to completion will remain time-consuming for the foreseeable future. Online processing techniques also are useful for data cleansing and for processing continuous or long-running queries against data sources with unpredictable performance, such as "deep web" sources. In this tutorial, we survey work on online query processing, including algorithms, statistical methods, system design issues and human-computer interactions. We also discuss the relation of online query processing to related fields such as data mining and OLAP, and discuss our experience in developing online systems.

Peter J. Haas received a Ph.D. in Operations Research from Stanford University in 1986 and has been a Research Staff Member at IBM Almaden Research Center since 1987. Much of his research focuses on the application of probabilistic and statistical techniques to the analysis and design of database management systems. He developed the fundamental statistical methodology underlying online aggregation processing, and led an effort to develop an online aggregation prototype for DB2. A number of his methods for estimating both query processing costs and database catalog statistics have been incorporated into the DB2 query optimizer, and he is currently developing techniques to support sampling and estimation in general DB2 SQL queries. He is an Associate Editor of the journal "Operations Research" and an occasional instructor at Stanford University.

Joseph M. Hellerstein is an Associate Professor of Computer Science at the University of California, Berkeley. His re-search focuses on database and information systems, including federated systems, adaptive dataflow, interactive data analysis and transformation, query optimization and execution, and indexing of complex data and queries. He is an Alfred P. Sloan Research Fellow, and a recipient of the NSF CAREER and NASA New Investigator awards; MIT's Technology Review recently named him one of the top 100 young technology innovators (TR100). Most recently, he was selected as a member of the first occasional SIGMOD/ PODS jazz trio. He is a member of the advisory boards of ACM SIGMOD and Ars Digita University. Hellerstein is the Chief Scientist and co-founder of Cohera Corporation, and serves on the technical advisory boards of a number of software and technology companies. He received his Ph.D. at the University of Wisconsin, Madison.

12:30–14:30

SIGMOD Awards Lunch (Plaza Del Sol)

Tuesday, May 22

14:30–16:00

PODS Session 5: Partial Information
(Sierra Madre North)
Chair: Gosta Grahne (Concordia University)

Representing and Querying XML with Incomplete Information, Serge Abiteboul (INRIA), Luc Segoufin (INRIA), and Victor Vianu (UC San Diego)

Querying Partially Sound and Complete Data Sources, Alberto O. Mendelzon (Univ. of Toronto) and George A. Mihaila (IBM T. J. Watson Research Center)

On Computing Functions with Uncertainty, Sanjeev Khanna and Wang-Chiew Tan (Univ. of Pennsylvania)

SIGMOD Research Session 3: Classification and Clustering (Santa Ynez)
Chair: Johannes Gehrke (Cornell Univ.)

Probe, Count, and Classify: Categorizing Hidden-Web Databases, Panagiotis G. Ipeirotis (Columbia Univ.), Luis Gravano (Columbia Univ.), and Mehran Sahami (E.piphany)

Data Bubbles: Quality Preserving Performance Boosting for Hierarchical Clustering, Markus M. Breunig, Hans-Peter Kriegel, Peer Kröger, and Jörg Sander (Univ. of Munich)

Mining Needle in a Haystack: Classifying Rare Classes via Two-phase Rule Induction, Mahesh V. Joshi (IBM Watson Research Center & Univ. of Minnesota), Ramesh C. Agarwal (IBM T. J. Watson Research Center), and Vipin Kumar (Univ. of Minnesota)

SIGMOD Research Session 4: Distributed and Heterogeneous Databases (San Rafael)
Chair: Laura Haas (IBM Almaden)

Efficient Evaluation of XML Middle-ware Queries, Mary Fernandez (AT&T Labs), Atsuyuki Morishima (Univ. of Tsukuba), and Dan Suciu (Univ. of Washington)

Filtering Algorithms and Implementation for Very Fast Publish/Subscribe Systems, Françoise Fabret (INRIA), H. Arno Jacobsen (Univ. of Toronto), François Llirbat (INRIA), Joao Pereira (INRIA), Kenneth A. Ross (Columbia Univ.), and Dennis Shasha (New York Univ.)

Adaptable Query Optimization and Evaluation in Temporal Middleware, Giedrius Slivinskas (Aalborg Univ.), Christian S. Jensen (Aalborg Univ.), and Richard T. Snodgrass (Univ. of Arizona)

Thursday, May 24

SIGMOD Panel 3: Application Servers: Born-Again TP Monitors for the Web? (Sierra Madre South)
Chair: C. Mohan (IBM)
Panelists:
Larry Cable (iPlanet), Matthieu Devin (Oracle), Scott Dietzen (BEA), Pat Helland (Microsoft), and Dan Wolfson (IBM)

SIGMOD Demonstrations Session 6: Content-based Retrieval, Web and Data Mining
(Santa Cruz/Anacapa)

Same demos as in *Demonstrations Session 3*

12:30

End of SIGMOD

14:00–18:00

Fourth International Workshop on the Web and Databases (WebDB'2001)

The Web is becoming the standard computing platform for the development of new-generation information systems. This is forcing to revisit some of the most traditional database concepts (from data models to languages, storage structures, optimization techniques, etc).
WebDB 2001–the fourth edition of a successful series–will feature 18 contributions selected by the Program Committee out of 65 submissions. The main goal of this fourth edition is to preserve the informal spirit of the previous editions, to promote novel research directions, and to provide a forum for researchers, theoreticians and practitioners alike, to disseminate and discuss new ideas related to data management for the World Wide Web.

Friday, May 25

9:00–18:00

Fourth International Workshop on the Web and Databases (WebDB'2001)

Workshop on Network-Related Data Management (NRDM 2001)

A significant amount of data is being steadily captured at various levels on the Internet. This includes both trace data and configuration data, and is being used increasingly to discern the topology of the Internet, to understand routing, and to characterize Internet traffic patterns. The NRDM 2001 workshop is intended to bring together researchers and practitioners in the database and the networking communities, to understand the challenges of managing and analyzing network related data.

Thursday, May 24

SIGMOD Research Session 15: Schemas and Queries (San Rafael)

Chair: Arnon Rosenthal (MITRE)

Data-Driven Understanding and Refinement of Schema Mappings, Ling Ling Yan (IBM Almaden), Renée J. Miller (Univ. of Toronto), Laura M. Haas (IBM Almaden), and Ronald Fagin (IBM Almaden)

Minimization of Tree Pattern Queries, Sihem Amer-Yahia (AT&T Labs), SungRan Cho (Stevens Inst. of Tech.), Laks V. S. Lakshmanan (Concordis Univ. & IIT Bambay), and Divesh Srivastava (AT&T Labs)

Reconciling Schemas of Disparate Data Sources: A Machine-Learning Approach, AnHai Doan, Pedro Domingos, and Alon Halevy (Univ. of Washington)

SIGMOD Tutorial 5: MPEG-7 Standard for Multimedia Databases (Sierra Madre North)

Chair: Betty Salzberg (Northeastern Univ.)

The Moving Picture Experts Group (MPEG) is developing a new standard called the "Multimedia Content Description Interface," also known as MPEG-7. The goal of MPEG-7 is to enable fast, effective, and interoperable searching, indexing, filtering, and retrieval of multimedia data. The effort is being driven by requirements taken from a large number of applications related to multimedia databases, inter-active media services (music, TV programs), video libraries, and so forth. MPEG-7 is achieving this goal by providing an XML-Schema based standard for describing features of multimedia content. In particular, MPEG-7 provides tools for describing multimedia content at the perceptual feature, syntactic, and semantic levels, and for describing models that bridge the gap between low-level features and high-level semantics. MPEG-7 presents a number of significant technical challenges for multimedia databases. Since MPEG-7 standardizes only the metadata structures them-selves (Descriptors and Description Schemes) and the Description Definition Language (DDL), it is left open for industry competition and future innovation to produce technologies for extracting, searching, and filtering MPEG-7 descriptions. For one, new query methods are needed for similarity matching with fuzzy constraints across the different levels of multimedia content description, including features, structure, and semantics. In addition, specialized database index structures are needed for MPEG-7 Descriptors because of their high-dimensionality and requisite specialized metrics. New methods are also needed for automatically classifying multimedia content at the semantic level based on automatically extracted features. In this tutorial, we study the emerging MPEG-7 standard and examine the new challenges for MPEG-7 multimedia databases.

John R. Smith received his M. Phil and PhD. degrees in Electrical Engineering from Columbia University in 1994 and 1997, resp. He is currently Manager of the Pervasive Media Management Group at IBM T. J. Watson Research Center. He is also Chair of the MPEG Multimedia Description Schemes (MDS) group and Technical Chair of MPEG-7 in NCITS. Dr. Smith's research interests include multimedia databases, content analysis, compression, indexing, and retrieval. At Columbia, he developed the WebSEEk and VisualSEEk image and video search engines, where he received the Eliahu I. Jury award for outstanding achievement as a graduate student in the areas of systems communication or signal processing. Dr. Smith is an Adjunct Professor at Columbia University and a member of IEEE.

Tuesday, May 22

SIGMOD Tutorial 2: Time Series Similarity Measures (Sierra Madre South)

Chair: Bongki Moon (University of Arizona)

Time series datasets come up in a variety of domains, including financial analysis, environmental or medical research, and modeling network, telecommunications, or web usage. The pervasiveness and importance of time series data has sparked a lot of research work on the topic. In the tutorial we examine the different time series similarity models that have been proposed, in terms of efficiency and accuracy. We also examine the indexing techniques that can be used for different models, and the dimensionality reduction techniques that have been proposed to improve indexing performance. The solutions encompass techniques from a wide variety of disciplines, such as databases, signal processing, speech recognition, pattern matching, combinatorics and statistics.

Dimitrios Gunopulos is an Assistant Professor at the Dept. of Computer Science and Engineering, in the Univ. of California, Riverside. His research is in the areas of Data Mining, Databases and Algorithms. His research interests include efficient indexing techniques for moving points, approximating range queries, similarity searching in sequence databases, finding frequent sequences or sets, approximate set indexing, local adaptive classification techniques. Dr. Gunopulos has held positions at the IBM Almaden Research Center and at the Max-Planck-Institut for Informatics. He completed his undergraduate studies at the Univ. of Patras, Greece and graduated with M.A. and Ph.D. degrees from Princeton Univ. His research is supported by NSF, DoD, and ATT.

Gautam Das is a researcher in the Data Management, Exploration and Mining Group at Microsoft Research. His research interests are in the design and analysis of algorithms that enable search and retrieval from information stores. He is investigating problems such as approximate answering of aggregate queries, integrating searching and browsing capabilities over heterogeneous data sources, and similarity search and retrieval problems in categorical databases and time series databases. Prior to Microsoft, Dr. Das held positions at Compaq (Tandem) Austin and the University of Memphis. He graduated with a B.Tech in computer science from IIT Kanpur, India, and with a Ph.D. in computer science from the University of Wisconsin, Madison.

SIGMOD Demonstrations Session 1: XML and Web (Santa Cruz/Anacapa)

Dissemination of Dynamic Data, Pavan Deolasee (IIT Bombay), Amol Katkar (IIT Bombay), Ankur Panchbudhe (IIT Bombay), Krithi Ramamritham (IIT Bombay & UMass-Amherst), and Prashant Shenoy (UMass-Amherst)

COSIMA - Your Smart, Speaking E-Salesperson, Werner Kießling, Stefan Holland, Stefan Fischer, and Thorsten Ehm (Univ. of Augsburg)

REVIEW: A Real Time Virtual Walkthrough System, L. Shou, C. H. Chionh, Z. Huang,Y. Ruan, and K.-L. Tan (National Univ. of Singapore)

Kweelt: More than just "yet another framework to query XML!", Arnaud Sahuguet (Univ. of Pennsylvania)

Tuesday, May 22

VQBD: Exploring Semistructured Data, Sudarshan S. Chawathe, Thomas Baby, and Jihwang Yeo (Univ. of Maryland)

OminiSearch: A Method for Searching Dynamic Content on the Web, David Buttler, Ling Liu, Calton Pu, Henrique Paques, Wei Han, and Wei Tang (Georgia Inst. of Tech.)

Securing XML Documents: the Author-X Project Demonstration, Elisa Bertino, Silvana Castano, and Elena Ferrari (DSI Univ. di Milano)

16:00–16:30

Coffee break (Santa Ynez Foyer)

16:30–18:00

PODS Research Session 6: Expressiveness/ Complexity (Sierra Madre North)
Chair: Neil Immerman (Univ. of Mass. at Amherst)

String Operations in Query Languages, Michael Benedikt (Bell Labs), Leonid Libkin (Univ. of Toronto), Thomas Schwentick (Univ. of Jena), and Luc Segoufin (INRIA)

Robbers, Marshals, and Guards: Game Theoretic and Logical Characterizations of Hypertree Width, Georg Gottlob (Technische Universität Wien), Nicola Leone (Univ. of Calabria), and Francesco Scarcello (Univ. of Calabria)

On the Complexity of Join Predicates, Jin-Yi Cai, Venkatesan T. Chakaravarthy, Raghav Kaushik, and Jeffrey F. Naughton (Univ. of Wisconsin)

SIGMOD Research Session 5: Indexing (Santa Ynez)
Chair: Walid Aref (Purdue Univ.)

Optimizing Multidimensional Index Trees for Main Memory Access, Kihong Kim, Sang K. Cha, and Keunjoo Kwon (Seoul National Univ.)

Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases, Eamonn Keogh, Kaushik Chakrabarti, Sharad Mehrotra, and Michael Pazzani (UC Irvine)

Main-Memory Index Structures with Fixed-Size Partial Keys, Philip Bohannon, Peter McIlroy, and Rajeev Rastogi (Bell Labs)

SIGMOD Research Session 6: Text Management (San Rafael)
Chair: Kyu-Young Whang (KAIST)

Automatic Segmentation of Tree Text into Structured Records, Vinayak R. Borkar, Kaustubh Deshmukh, and Sunita Sarawagi (IIT Bombay)

Thursday, May 24

7:30–9:00

Continental breakfast (Santa Ynez Foyer)

9:00–10:30

SIGMOD/PODS Keynote Speech (Santa Ynez/San Rafael)
Chair: Divyakant Agrawal (UC Santa Barbara)

To the Internet and Beyond: Database Challenges for New Applications, Steve Kirsch (Propel)

Steve Kirsch was chairman and founder of Infoseek Corporation, a leading Internet navigation service. After Disney's acquisition of Infoseek he started a new company, Propel, focused on infrastructure for the expanding Internet e-commerce market. Prior to Infoseek, Steve founded Frame Technology in 1986 and Mouse Systems in 1982. He has been involved with the Internet since 1972 when he worked with the father of the Internet Vint Cerf and others at the first ARPANET node as a systems programmer while attending high school. He has BS and MS degrees in Electrical Engineering and Computer Science from MIT.

Steve and his wife Michele have a $75M charitable fund and have donated over $10M to charitable causes over the past 5 years. Recently, in response to an $11M shortfall at the United Way of Santa Clara County, the Kirsch Foundation donated $1M to establish an Emergency Response Fund at the Community Foundation of Silicon Valley to benefit the charities affected by the shortfall. In 1999, they were named "Outstanding Philanthropists of the Year" by the Silicon Valley chapter of the National Society of Fund Raising Executives. The time Steve doesn't spend at work or with family is usually spent on philanthropy.

Steve was the keynote speaker at the SIGIR (Special Interest Group on Information Retrieval) conference in Australia (Aug 24-28). He has given keynote presentations at Internet World, Web World, Inc. magazine CEO conference, AEA Innovation Now, COSMOS '97 (Japan), and Infostor. He has spoken at Comdex, Agenda, Seybold, PC Forum, SCOUG, CalTech Enterprise Forum, the Software Publisher's Association, and several other forums.

10:30–11:00

Coffee break (Santa Ynez Foyer)

11:00–12:30

SIGMOD Research Session 14: Sampling and Association Rules (Santa Ynez)
Chair: Alexandros Biliris (AT&T Labs)

Using the Golden Rule of Sampling for Query Estimation, Yi-Leh Wu, Divyakant Agrawal, and Amr El Abbadi (UC Santa Barbara)

Selectivity Estimation using Probabilistic Models, Lise Getoor, Benjamin Taskar, and Daphne Koller (Stanford Univ.)

Communication Efficient Distributed Mining of Association Rules, Assaf Schuster and Ran Wolff (TECHNION)

Wednesday, May 23

SIGMOD Research Session 13: XML (San Rafael)
Chair: H. V. Jagadish (Univ. of Michigan)

Updating XML, Igor Tatarinov, Zachary G. Ives, Alon Y. Halevy, and Daniel S. Weld (Univ. of Washington)

On Supporting Containment Queries in Relational Database Management Systems, Chun Zhang (Univ. of Wisconsin), Jeffrey Naughton (Univ. of Wisconsin), David DeWitt (Univ. of Wisconsin), Qiong Luo (Univ. of Wisconsin), and Guy Lohman (IBM Almaden)

Monitoring XML Data on the Web, Benjamin Nguyen (INRIA), Serge Abiteboul (INRIA), Gregory Cobena (INRIA), and Mihai Preda (Xyleme S.A.)

SIGMOD Industrial Session 4: Database Internals (Sierra Madre South)
Chair: Michael Franklin (UC Berkeley)

Orthogonal Optimization of Subqueries and Aggregation, César Galindo-Legaria and Milind Joshi (Microsoft)

Exploiting Constraint-Like Data Characterizations in Query Optimization, Parke Godfrey (York Univ.), Jarek Gryz (York Univ.), and Calisto Zuzarte (IBM Canada)

Fast-Start: Quick Fault Recovery in Oracle, Tirthankar Lahiri (Oracle), Amit Ganesh (Oracle), Ron Weiss (Oracle), and Ashok Joshi (NuGenesis Tech.)

SIGMOD Demonstrations Session 5: XML and Web (Santa Cruz/Anacapa)

Same demos as in *Demonstrations Session 1*

19:00–21:30

California Barbeque (beach)

Tuesday, May 22

Efficient and Effective Metasearch for Text Databases Incorporating Linkages among Documents, Clement Yu (UI Chicago), Weiyi Meng (SUNY Binghamton), Wensheng Wu (UIUC), and King-Lup Liu (DePaul Univ.)

SIGMOD Panel 1: XML Data Management: Go Native or Spruce up Relational Systems? (Sierra Madre South)
Chair: Per-Åke Larson (Microsoft)
Panelists:
Dana Florescu (Propel Software), Goetz Graefe (Microsoft), Guido Moerkotte (Univ. Mannheim), Hamid Pirahesh (IBM), and Harald Schöning (Software AG)

SIGMOD Demonstrations Session 2: Schemas, Models and Processes (Santa Cruz/Anacapa)

Model Management - A Solution to Support Multiple Data Models, their Mappings and Maintenance, Kajal T. Claypool (Worcester Polytechnic Inst.), Elke A. Rundensteiner (Worcester Polytechnic Inst.), Xin Zhang (HP), Su Hong (HP), Harumi Kuno (HP), Wang-Chien Lee (Verizon Comm.) and Gail Mitchell (Verizon Comm.)

Clio: A Semi-Automatic Tool For Schema Mapping Mauricio A. Hernández (IBM Almaden), Renée J. Miller (Univ. of Toronto), Laura M. Haas (IBM Almaden)

Materialized View and Index Selection Tool for Microsoft SQL Server 2000, Sanjay Agrawal, Surajit Chaudhuri, and Vivek Narasayya (Microsoft Research)

The Prototype of the DARE system, Tiziana Catarci and Giuseppe Santucci (Univ. di Roma)

Monitoring Business Processes through Event Correlation based on Dependency Mode, Asaf Adi, David Botzer, Opher Etzion, and Tali Yatzkar-Haham (IBM Haifa)

Fault-tolerant, Load-balancing Queries in Telegraph, Mehul A. Shah and Sirish Chandrasekaran (UC Berkeley)

Snowball: A Prototype System for Extracting Relations from Large Text Collections, Eugene Agichtein, Luis Gravano, Jeff Pavel, Viktoriya Sokolova, and Aleksandr Voskoboynik (Columbia Univ.)

19:00–20:30

Reception (Plaza Del Sol)

7:30–9:00

Continental Breakfast (Santa Ynez Foyer)

9:00–10:30

PODS Research Session 7: Aggregates
(Sierra Madre North)
Chair: Paolo Atzeni (Università Roma Tre)

Equivalences among Aggregate Queries with Negation, Sara Cohen (Hebrew Univ.), Werner Nutt (Heriot-Watt Univ.) and Yehoshua Sagiv (Hebrew Univ.)

Optimal and Approximate Computation of Summary Statistics for Range Aggregates, Anna C. Gilbert, Yannis Kotidis, S. Muthukrishnan, and Martin J. Strauss (AT&T Labs)

Efficient Computation of Temporal Aggregates with Range Predicates, Donghui Zhang (UC Riverside), Alexander Markowetz (Philipps Universität Marburg), Vassilis Tsotras (UC Riverside), Dimitrios Gunopulos (UC Riverside), and Bernhard Seeger (Philipps Universität Marburg)

SIGMOD Research Session 7: Histograms (Santa Ynez)
Chair: Jose Blakeley (Microsoft Research)

Independence is Good: Dependency-Based Histogram Synopses for High-Dimensional Data, Amol Deshpande (UC Berkeley), Minos Garofalakis (Bell Labs), and Rajeev Rastogi (Bell Labs)

STHoles: A Multidimensional Workload-Aware Histogram, Nicolas Bruno (Columbia Univ.), Surajit Chaudhuri (Microsoft Research), and Luis Gravano (Columbia Univ.)

Global Optimization of Histograms, H. V. Jagadish (Univ. of Michigan), Hui Jin (National Univ. of Singapore), Beng Chin Ooi (National Univ. of Singapore), and Kian-Lee Tan (National Univ. of Singapore)

SIGMOD Research Session 8: Performance (San Rafael)
Chair: Krithi Ramamritham (Univ. of Mass. & IIT Bombay)

Improving Index Performance through Prefetching, Shimin Chen (Carnegie Mellon Univ.), Phillip B. Gibbons (Bell Labs), and Todd C. Mowry (Carnegie Mellon Univ.)

Efficient and Tunable Similar Set Retrieval, Aristides Gionis (Stanford Univ.), Dimitrios Gunopulos (UC Riverside), and Nikos Koudas (AT&T Labs)

PREFER: A System for the Efficient Execution of Multi-parametric Ranked Queries, Vagelis Hristidis (UC San Diego), Nick Koudas (AT&T Labs), and Yannis Papakonstantinou (UC San Diego)

Wednesday, May 23

computer science from University of California, Berkeley in 1980. He has published more than 50 technical papers and been granted 15 software patents.

SIGMOD Industrial Session 3: Applications & Lessons Learned (San Rafael)
Chair: Cesar Galindo-Legaria (Microsoft Research)

Experiences in Mining Aviation Safety Data, Zohreh Nazeri, Eric Bloedorn, and Paul Ostwald (MITRE)

The Nimble Integration Engine, Denise Draper, Alon Y. Halevy, and Daniel Weld (Nimble Tech.)

Lasting Impact of the Wild, Wild, Web, Reed M. Meseck (Consultant)

SIGMOD Demonstrations Session 4: Schemas, Models and Processes (Santa Cruz/Anacapa)

Same demos as in *Demonstrations Session 2*

16:00–16:30

Coffee Break (Santa Ynez Foyer)

16:30–18:00

PODS Research Session 9: Indexing/Transactions
(Sierra Madre North)
Chair: Amr El Abbadi (UC Santa Barbara)

Database-Friendly Random Projections, Dimitris Achlioptas (Microsoft Research)

Two-dimensional Substring Indexing, Paolo Ferragina (Univ. of Pisa), Nick Koudas (AT&T Labs), S. Muthukrishnan (AT&T Labs), and Divesh Srivastava (AT&T Labs)

Process Locking: A Protocol based on Ordered Shared Locks for the Execution of Transactional Processes, Heiko Schuldt (Swiss Federal Inst. of Tech.)

SIGMOD Research Session 12: Multi-dimensional Data (Santa Ynez)
Chair: Sunil Prabhakar (Purdue Univ.)

Epsilon Grid Order: An Algorithm for the Similarity Join on Massive High-Dimensional Data, Christian Böhm, Bernhard Braunmüller, Florian Krebs, and Hans-Peter Kriegel (Univ. of Munich)

Modeling High-Dimensional Index Structures using Sampling, Christian A. Lang and Ambuj K. Singh (UC Santa Barbara)

Progressive Approximate Aggregate Queries with a Multi-Resolution Tree Structure, Iosif Lazaridis and Sharad Mehrotra (UC Irvine)

On the Effects of Dimensionality Reduction on High Dimensional Similarity Search, Charu C. Aggarwal (IBM T. J. Watson Research Center)

A Condensed Representation to Find Frequent Patterns, Artur Bykowski and Christophe Rigotti (Laboratoire d'Ingénierie des Systèmes d'Information)

SIGMOD Research Session 11: Caching and Buffer Management (Santa Ynez)
Chair: Minos Garofalakis (Bell Labs)

Dynamic Buffer Allocation in Video-on-Demand Systems, Sang-Ho Lee (Korea Advanced Inst. of Science and Tech.), Kyu-Young Whang (Korea Advanced Inst. of Science and Tech.), Yang-Sae Moon (Korea Advanced Inst. of Science and Tech.), and Il-Yeol Song (Drexel Univ.)

Adaptive Precision Setting for Cached Approximate Values, Chris Olston, Boon Thau Loo, and Jennifer Widom (Stanford Univ.)

Proxy-Server Architectures for OLAP, Panos Kalnis and Dimitris Papadias (Hongkong Univ. of Science and Tech.)

SIGMOD Tutorial 4: Models and Languages for Describing and Discovering E-Services (Sierra Madre South)
Chair: Christian Jensen (Aalborg Univ.)

E-services are business functions made available via the Internet by service providers, and accessible by clients that could be human users or software applications. The main benefit of the e-services environment is that clients are able to dynamically discover the available e-service that best meets their needs, to examine its properties and capabilities, and to determine if and how to access it. However, in order to deliver e-services to clients, service providers are faced with several challenges. In particular, they need to describe e-services in a way that is accessible and understandable by the clients and to advertise them in web directories, so that they can be discovered by brokers as well as by end-users. In this tutorial we discuss the main requirements for models and languages for service description and discovery, and we present relevant approaches proposed by vendors and by standardization consortia.

Fabio Casati is a researcher at HP Labs, Palo Alto. He got his PhD from Politecnico di Milano (Italy) in 1999. His research interests include workflow management, e-services, mobile environments, and business process intelligence. He is author of more than 30 papers in international conferences and journals, and has served as organizer and program committee member in several conferences. He is also a lecturer of "Technologies for e-business" at San Jose State University.

Ming-Chien Shan is a research program manager in the Hewlett Packard Laboratories, Palo Alto, California. He joined IBM DB2 team in 1978 working on query optimization, data definition manager and distributed DBMS. He then joined HP in 1985 and managed various research projects, including object-oriented DBMS, heterogeneous DBMS, workflow and telecom service provisioning. Currently, he is the manager of e-business solutions program. Ming-Chien received his PhD degree in

Wednesday, May 23

SIGMOD Tutorial 3: Semantic B2B Integration (Sierra Madre South)
Chair: Dan Suciu (Univ. of Washington)

The tutorial "Semantic B2B Integration" will give an introduction to the field of business-to-business (B2B) integration from a technical viewpoint with the focus on semantic integration aspects. The set of B2B integration concepts is introduced as well as their implementation in form of a technical semantic B2B integration architecture. A mix of examples is taken illustrating the problems that need to be solved in semantic B2B integration projects. The tutorial enables the audience to identify semantic B2B integration problems as well as to determine the benefits and deficiencies of various technical integration architecture approaches or B2B integration technologies.

Christoph Bussler is Member of Oracle's Integration Platform Architecture Group based in Redwood Shores, CA. He is responsible for the architecture of Oracle's next generation integration product. Prior to joining Oracle he was at Jamcracker, Cupertino, CA, responsible for defining Jam-cracker's ASP aggregation architecture, Netfish Technologies, Santa Clara, CA, responsible for Netfish's B2B integration server, The Boeing Company, Seattle, WA, leading Boeing's workflow research and Digital Equipment, Mountain View, CA, defining the policy resolution component of Digital's workflow product. He has a Ph.D. in computer science from the University of Erlangen, Germany, and a Master in computer science from the University of Munich, Germany.

SIGMOD Demonstrations Session 3: Content-based Retrieval, Web and Data Mining (Santa Cruz/Anacapa)

PBIR - Perception-Based Image Retrieval, Edward Chang (UC Santa Barbara), Kwang-Ting (Tim) Cheng (UC Santa Barbara), and Lihyuarn L. Chang (Morpho Software)

Spatial Data Management for Computer Aided Design, Hans-Peter Kriegel, Andreas Müller, Marco Pötke, and Thomas Seidl (Univ. of Munich)

RETINA: A REal-time TraffIc NAvigation System, Kam-Yiu Lam (City Univ. of Hong Kong), Edward Chan (City Univ. of Hong Kong), Tei-Wei Kuo (National Taiwan Univ.), S.W. Ng (City Univ. of Hong Kong), Dick Hung (City Univ. of Hong Kong)

Dynamic Content Acceleration: A Caching Solution to Enable Scalable Dynamic Web Page Generation, Anindya Datta (Chutney Tech.), Kaushik Dutta (Chutney Tech.), Krithi Ramamritham (IIT Bombay & UMass Amherst), Helen Thomas (Chutney Tech.), and Debra VanderMeer (Chutney Tech.)

Lots o' Ticks: Real-time High Performance Time Series Queries on Billions of Trades and Quotes, Arthur Whitney (KX Systems) and Dennis Shasha (New York Univ.)

DNA-Miner: A System Prototype for Mining DNA Sequences, Jiawei Han (Simon Fraser Univ.), Hasan Jamil (Miss. State Univ.), Ying Lu (Simon Fraser Univ.), Liangyou Chen (Miss. State Univ.), Yaqin Liao (Simon Fraser Univ.), and Jian Pei (Simon Fraser Univ.)

DyDa: Data Warehouse Maintenance in Fully Concurrent Environments, Jun Chen, Xin Zhang, Songting Chen, Andreas Koeller, and Elke A. Rundensteiner (Worcester Polytechnic Inst.)

10:30–11:00

Coffee break (Santa Ynez Foyer)

11:00–12:30

PODS Invited Tutorial 2: The Challenges of Delivering Content on the Internet (Sierra Madre North)

Chair: Richard Hull (Bell Labs)

In this talk, we will give an overview of how content is distributed on the internet, with an emphasis on the approach being used by Akamai. We will describe some of the technical challenges involved in operating a network of thousands of content servers across multiple geographies on behalf of thousands of customers. The talk will be introductory in nature and should be accessible to a broad audience.

Tom Leighton is a Professor of Applied Mathematics at the Massachusetts Institute of Technology. He is the head of the Algorithms Group at the MIT Laboratory for Computer Science. He received the B.S.E. degree in Electrical Engineering and Computer Science from Princeton University in 1978, and the PhD degree in Applied Mathematics from MIT in 1981. He was a Bantrell Postdoctoral Research Fellow at the MIT Laboratory for Computer Science (LCS) from 1981 to 1983, and he joined the MIT faculty as an Assistant Professor of Applied Mathematics in 1982. He is currently Head of the MIT LCS Algorithms Group. He is the author of over 100 research papers in the areas of parallel algorithms and architectures, communication protocols for networks, combinatorial optimization, probabilistic methods, VLSI computation and design, sequential algorithms, and graph theory. He is also the author of two books, including Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hybercubes. He was acting chair for the 31st ACM Symposium on Theory of Computing, and is a member of the Editorial Boards of SIAM Journal on Computing, Algorithmica, and Journal of Algorithms, Combinatorica, and Journal of Graph Theory.

He is is currently on leave and serving as Chief Scientist for Akamai Technologies, Inc. Akamai Technologies is a leader in the field of content distribution for major websites.

SIGMOD Research Session 9: Compression (Santa Ynez)

Chair: Eric Simon (INRIA)

Query Optimization In Compressed Database Systems, Zhiyuan Chen (Cornell Univ.), Johannes Gehrke (Cornell Univ.), and Flip Korn (AT&T Labs)

Wednesday, May 23

SPARTAN: A Model-Based Semantic Compression System for Massive Data Tables, Shivnath Babu (Stanford Univ.), Minos Garofalakis (Bell Labs), and Rajeev Rastogi (Bell Labs)

A Robust, Optimization-Based Approach for Approximate Answering of Aggregate Querie, Surajit Chaudhuri, Gautam Das, and Vivek Narasayya (Microsoft Research)

SIGMOD Research Session 10: Materialized Views (San Rafael)

Chair: Tiziana Catarci (Universita degli Studi di Roma "La Sapienza")

Materialized View Selection and Maintenance Using Multi-Query Optimization, Hoshi Mistry (IIT-Bombay), Prasan Roy (Bell Labs.), S. Sudarshan (IIT-Bombay), and Krithi Ramamritham (IIT-Bombay & UMass-Amherst)

Generating Efficient Plans for Queries Using Views, Foto N. Afrati (National Technical Univ. of Athens), Chen Li (Stanford Univ.), and Jeffrey D. Ullman (Stanford Univ.)

Optimizing Queries Using Materialized Views: A Practical, Scalable Solution, Jonathan Goldstein and Per-Åke Larson (Microsoft Research)

SIGMOD Panel 2: Will Database Researchers Have ANY Role in Data Security? (Sierra Madre South)

Chair: Arnie Rosenthal (MITRE)

Panelists:

Klaus Dittrich (Univ. of Zurich), Jim Donahue (InterTrust), Bill Maimone (Oracle), and Elias Manousos (Securant)

SIGMOD Industrial Session 2: Catalog Integration (Fiesta)

Chair: Rakesh Agrawal (IBM Almaden)

Content Integration for E-Business, Michael Stonebraker and Joseph M. Hellerstein (Cohera)

Catalog Management in Websphere Commerce Suite, Thomas Maguire

12:30–14:30

SIGMOD Business Lunch (Plaza Del Sol)

14:30–16:00

PODS Research Session 8: Data Mining (Sierra Madre North)

Chair: Jianwen Su (UC Santa Barbara)

On the Design and Quantification of Privacy Preserving Data Mining Algorithms, Dakshi Agrawal and Charu C. Aggarwal (IBM T. J. Watson Research Center)