

# Proceedings

## 30<sup>th</sup> INTERNATIONAL CONFERENCE on

### Very Large Data Bases

#### Editors:

Mario A. Nascimento

M. Tamer Özsu

Donald Kossmann

Renée J. Miller

José A. Blakeley

Berni Schiefer

August 31 - September 3, 2004

Toronto, Canada

Proceedings  
of the  
Thirtieth  
International Conference  
on  
Very Large Data Bases  
Toronto, Canada  
August 31 - September 3, 2004

Editors:

Mario A. Nascimento

M. Tamer Özsu

Donald Kossmann

Renée J. Miller

José A. Blakeley

Berni Schiefer

## Ordering Information

Morgan Kaufmann Publishers is the exclusive worldwide distributor for the VLDB proceedings volumes listed below:

	ISBN
2004 Toronto, Canada	0-12-088469-0
2003 Berlin, Germany	0-12-722442-4
2002 Hong Kong, China	1-55860-869-9
1990 Brisbane, Australia	1-55860-149-X
1989 Amsterdam, The Netherlands	1-55860-101-5
1988 Los Angeles, USA	0-934613-75-3
1985 Stockholm, Sweden	0-934613-17-6
1984 Singapore	0-934613-16-8
1983 Florence, Italy	0-934613-15-X

Prices are \$50.00 per copy for the 2000 - 2004 volumes, \$40.00 per copy for all other volumes.

Shipping is free from Morgan Kaufmann within the U.S. on prepaid orders. International shipping costs are \$7 per volume via DHL/regular mail combination, or \$20.00 per volume via international overnight courier. Morgan Kaufmann accepts credit card payments: the buyer should provide card number, expiration date, and name as it appears on the card for Visa, MasterCard, or American Express credit cards. Morgan Kaufmann also accepts cheque payments in U.S. dollars only; cheques must be drawn on a U.S. bank.

Order from Morgan Kaufmann Publishers

By Mail: Morgan Kaufmann Publishers/Elsevier Science  
Attention: Order Fulfillment Department  
11830 Westline Industrial Drive  
St. Louis, MO 63146-9938

By Phone: 1-800-545-2522 (from within US & Canada) and 1-314-453-7010 (International)  
By Fax: 1-800-535-9935 or 1-314-453-7095 (International)  
By Email: [custserv.mkp@elsevier.com](mailto:custserv.mkp@elsevier.com)  
By Web: <http://www.mkp.com>

VLDB 2004 ISBN 0-12-722442-4  
ISSN 0730-9317

Copyright © 2004 VLDB Endowment

# Conference Officers

## **General Chair**

John Mylopoulos, University of Toronto, Canada

## **VLDB Endowment Liaison**

Kyu-Young Whang, KAIST, Korea

## **Area Coordinator, North America**

Philip Bernstein, Microsoft Research, USA

## **Area Coordinator, South America**

Alberto Laender, Federal University of Minas Gerais, Brazil

## **Area Coordinator, Europe, Mideast & Africa**

Avigdor Gal, Technion, Israel

## **Area Coordinator, Far East & Australia**

Hongjun Lu, Hong Kong University of Science and Technology, China

## **Publicity and Publications Chair**

Mariano Consens, University of Toronto, Canada

## **Web Coordinator**

Manuel Kolp, Catholic University of Louvain, Belgium

## **Technical Program**

### **Technical Program Chair**

M. Tamer Özsu, University of Waterloo, Canada

### **Core Database Technology Program Chair**

Donald Kossmann, University of Heidelberg, Germany

### **Infrastructure for Information Systems Program Chair**

Renée J. Miller, University of Toronto, Canada

### **Industrial and Applications Program Co-Chairs**

José A. Blakeley, Microsoft Corporation, USA

Berni Schiefer, IBM Laboratories, Canada

### **Tutorial Program Co-Chairs**

Raymond Ng, University of British Columbia, Canada

Matthias Jarke, RWTH Aachen, Germany

### **Panel Program Co-Chairs**

Jarek Gryz, York University, Canada

Fred Lochovsky, HKUST, Hong Kong

### **Demonstrations Co-Chairs**

Bettina Kemme, McGill University, Canada

David Toman, University of Waterloo, Canada

### **Proceedings Editor**

Mario A. Nascimento, University of Alberta, Canada

### **eProceedings Editor**

Patricia Rodriguez-Gianolli, University of Toronto, Canada

### **Workshops Co-Chairs**

Alberto Mendelzon, University of Toronto, Canada

S. Sudarshan, Indian Institute of Technology, Bombay, India

## Ordering Information

Morgan Kaufmann Publishers is the exclusive worldwide distributor for the VLDB proceedings volumes listed below:

	ISBN
2004 Toronto, Canada	0-12-088469-0
2003 Berlin, Germany	0-12-722442-4
2002 Hong Kong, China	1-55860-869-9
1990 Brisbane, Australia	1-55860-149-X
1989 Amsterdam, The Netherlands	1-55860-101-5
1988 Los Angeles, USA	0-934613-75-3
1985 Stockholm, Sweden	0-934613-17-6
1984 Singapore	0-934613-16-8
1983 Florence, Italy	0-934613-15-X

Prices are \$50.00 per copy for the 2000 - 2004 volumes, \$40.00 per copy for all other volumes.

Shipping is free from Morgan Kaufmann within the U.S. on prepaid orders. International shipping costs are \$7 per volume via DHL/regular mail combination, or \$20.00 per volume via international overnight courier. Morgan Kaufmann accepts credit card payments: the buyer should provide card number, expiration date, and name as it appears on the card for Visa, MasterCard, or American Express credit cards. Morgan Kaufmann also accepts cheque payments in U.S. dollars only; cheques must be drawn on a U.S. bank.

Order from Morgan Kaufmann Publishers

By Mail: Morgan Kaufmann Publishers/Elsevier Science  
Attention: Order Fulfillment Department  
11830 Westline Industrial Drive  
St. Louis, MO 63146-9938

By Phone: 1-800-545-2522 (from within US & Canada) and 1-314-453-7010 (International)  
By Fax: 1-800-535-9935 or 1-314-453-7095 (International)  
By Email: [custserv.mkp@elsevier.com](mailto:custserv.mkp@elsevier.com)  
By Web: <http://www.mkp.com>

VLDB 2004 ISBN 0-12-722442-4  
ISSN 0730-9317

Copyright © 2004 VLDB Endowment

# Conference Officers

## **General Chair**

John Mylopoulos, University of Toronto, Canada

## **VLDB Endowment Liaison**

Kyu-Young Whang, KAIST, Korea

## **Area Coordinator, North America**

Philip Bernstein, Microsoft Research, USA

## **Area Coordinator, South America**

Alberto Laender, Federal University of Minas Gerais, Brazil

## **Area Coordinator, Europe, Mideast & Africa**

Avigdor Gal, Technion, Israel

## **Area Coordinator, Far East & Australia**

Hongjun Lu, Hong Kong University of Science and Technology, China

## **Publicity and Publications Chair**

Mariano Consens, University of Toronto, Canada

## **Web Coordinator**

Manuel Kolp, Catholic University of Louvain, Belgium

## **Technical Program**

### **Technical Program Chair**

M. Tamer Özsu, University of Waterloo, Canada

### **Core Database Technology Program Chair**

Donald Kossmann, University of Heidelberg, Germany

### **Infrastructure for Information Systems Program Chair**

Renée J. Miller, University of Toronto, Canada

### **Industrial and Applications Program Co-Chairs**

José A. Blakeley, Microsoft Corporation, USA

Berni Schiefer, IBM Laboratories, Canada

### **Tutorial Program Co-Chairs**

Raymond Ng, University of British Columbia, Canada

Matthias Jarke, RWTH Aachen, Germany

### **Panel Program Co-Chairs**

Jarek Gryz, York University, Canada

Fred Lochovsky, HKUST, Hong Kong

### **Demonstrations Co-Chairs**

Bettina Kemme, McGill University, Canada

David Toman, University of Waterloo, Canada

### **Proceedings Editor**

Mario A. Nascimento, University of Alberta, Canada

### **eProceedings Editor**

Patricia Rodriguez-Gianolli, University of Toronto, Canada

### **Workshops Co-Chairs**

Alberto Mendelzon, University of Toronto, Canada

S. Sudarshan, Indian Institute of Technology, Bombay, India

## **Local Organization**

### **Local Arrangements Chair**

Nick Koudas, AT&T Labs Research, USA

### **Exhibits Chair**

H.-Arno Jacobsen, University of Toronto, Canada

### **Treasurer**

Kenneth Barker, University of Calgary, Canada

### **Fund Raising Chair**

Victor DiCiccio, University of Waterloo, Canada

### **Registration Chair**

Grant Weddell, University of Waterloo, Canada



# Program Committee Members

## Core Database Technology

Karl Aberer, EPFL, Switzerland  
Anastassia Ailamaki, CMU, USA  
Demet Aksoy, UC Davis, USA  
Paul Aoki, PARC, USA  
Walid Aref, Purdue University, USA  
Remzi Arpaci-Dusseau, University of Wisconsin, USA  
Philip Bohannon, Bell Labs, USA  
Christian Böhm, University of Munich, Germany  
Philippe Bonnet, University of Copenhagen, Denmark  
Stephane Bressan, National University of Singapore, Singapore  
Mike Carey, BEA Systems, USA  
Stefano Ceri, Politecnico Milan, Italy  
Surajit Chaudhuri, Microsoft Research, USA  
Arbee Chen, National Dong Hwa University, Taiwan  
Bobby Cochrane, IBM Almaden, USA  
Laurent Daynes, Sun Microsystems, France  
Alin Deutsch, UC San Diego, USA  
Klaus Dittrich, University of Zurich, USA  
Wenfei Fan, Bell Labs, USA  
Mary Fernandez, ATT Research, USA  
Elena Ferrari, University of Milano, Italy  
Daniela Florescu, BEA Systems, USA  
Mike Franklin, UC Berkeley, USA  
Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany  
Johannes Gehrke, Cornell University, USA  
Goetz Graefe, Microsoft, USA  
Ralf Güting, Fernuni Hagen, Germany  
Peter Haas, IBM Almaden, USA  
Alon Halevy, University of Washington, USA  
Sven Helmer, University of Mannheim, Germany  
Alexander Hinneburg, University of Halle, Germany  
Wie Hong, Intel Research Lab, USA  
H.V. Jagadish, University of Michigan, USA  
Christian S. Jensen, Aalborg University, Denmark  
Björn Þór Jónsson, Reykjavík University, Iceland  
Alfons Kemper, University of Passau, Germany  
Martin Kersten, CWI, Netherlands  
Masaru Kitsuregawa, Kitsuregawa Lab, Japan  
George Kollios, Boston University, USA  
Hank Korth, Lehigh University, USA  
Alexandros Labrinidis, University of Pittsburgh, USA  
Paul Larson, Microsoft Research, USA  
David Lomet, Microsoft Research, USA  
Hongjun Lu, HKUST, China  
Ioana Manolescu, INRIA Futurs, France  
Volker Markl, IBM Almaden, USA  
Guido Moerkotte, University of Mannheim, Germany  
Mario Nascimento, University of Alberta, Canada  
Chris Olston, Carnegie Mellon University, USA  
Fatma Ozcan, IBM Almaden, USA  
Dimitris Papadias, HKUST, China  
Yannis Papakonstantinou, UC San Diego, USA  
Krithi Ramamritham, IIT Bombay, India  
Tore Risch, Uppsala University, Sweden  
Michael Rys, Microsoft, USA  
Ken Salem, University of Waterloo, Canada  
Sunita Sarawagi, IIT Bombay, India  
Oded Shmueli, Technion, Israel  
Ioana Stanoi, IBM Almaden, USA  
S. Sudarshan, IIT Bombay, India  
Kian-Lee Tan, National University of Singapore, Singapore  
Dirk van Gucht, University of Indiana, USA  
Jun Yang, Duke University, USA  
Clement Yu, University of Illinois, USA  
Masatoshi Yoshikawa, Nagoya University, Japan  
Carlo Zaniolo, UCLA, USA  
Justin Zobel, RMIT, Australia



## Infrastructure for information Systems

Sihem Amer-Yahia, AT&T Labs Research, USA  
Paolo Atzeni, Università di Roma Tre, Italy  
Phil Bernstein, Microsoft Research, USA  
Leo Bertossi, Carleton University, Canada  
Peter Buneman, University of Edinburgh, UK  
Tiziana Catarci, Università di Roma La Sapienza, Italy  
Soumen Chakrabarti, Indian Institute of Technology, Bombay, India  
Mitch Cherniack, Brandeis University, USA  
Panos Chrysanthis, University of Pittsburgh, USA  
Mariano Consens, University of Waterloo, Canada  
Isabel Cruz, University of Illinois at Chicago, USA  
Tamraparni Dasu, AT&T Labs-Research, USA  
Asuman Dogac, Middle East Tech Univ., Turkey  
Lise Getoor, University of Maryland, USA  
Dimitrios Gunopulos, University of California, Riverside, USA  
Laura Haas, IBM Silicon Valley Lab, USA  
Jayant Haritsa, Indian Institute of Science, Bangalore, India  
Wynne Hsu, National University of Singapore, Singapore  
Carlos Hurtado, Universidad de Chile, Santiago, Chile  
Chris Jermaine, University of Florida, USA  
Christoph Koch, University of Edinburgh, UK  
Nick Koudas, AT&T Labs Research, USA  
Hans-Peter Kriegel, University of Munich, Germany  
Mong Li Lee, National University of Singapore, Singapore  
Maurizio Lenzerini, Università di Roma, La Sapienza, Italy  
Qiong Luo, Hong Kong Univ. Science & Tech., Hong Kong, China  
Pat Martin, Queens University, Canada  
David Maier, Oregon Health & Science University, USA  
Alberto Mendelzon, University of Toronto, Canada  
Tova Milo, Tel Aviv University, Israel  
Jeff Naughton, University of Wisconsin, USA  
Raymond Ng, UBC, Canada  
Joann Ordille, Avaya, USA  
Themis Palpanas, University of California, Riverside, USA  
Jignesh Patel, University of Michigan, USA  
Lucian Popa, IBM Almaden Research Center, USA  
Alexandra Poullovassilis, Univ. of London, UK  
Davood Rafiei, University of Alberta, Canada  
Raghu Ramakrishnan, University of Wisconsin, USA  
Rajeev Rastogi, Bell Labs, USA  
Prasan Roy, Indian Institute of Technology, Bombay, India  
Elke Rundensteiner, Worcester Polytechnic Institute, USA  
Timos Sellis, National Technical Univ. of Athens, Greece  
Dennis Shasha, New York University, USA  
Kyuseok Shim, Seoul National Univ., South Korea  
Eric Simon, Inria Rocquencourt, France  
Avi Silberschatz, Yale University, USA  
Divesh Srivastava, AT&T Labs Research, USA  
Wang-Chiew Tan, University of California, Santa Cruz, USA  
David Toman, University of Waterloo, Canada  
Alejandro Vaisman, Universidad de Buenos Aires, Argentina  
Victor Vianu, Univ. California, San Diego, USA  
Limsoon Wong, Institute for Infocomm Research, Singapore

## Industrial and Applications Program

Achim Kraiss, SAP, Germany  
Alejandro P. Buchmann, Darmstadt University of Technology, Germany  
Anand Deshpande, Persistent Systems, India  
Anil K. Nori, Microsoft Corporation, USA  
Erhard Rahm, Universität Leipzig, Germany  
Francois J. Raab, TPC Certified Auditor, USA  
Glenn Paulley, iAnywhere Solutions, Canada  
Guy M. Lohman, IBM Almaden Research Center, USA  
Jacob Slonim, Dalhousie University, Canada  
Joseph M. Hellerstein, University of California, Berkeley, USA  
Josep-L. Larriba-Pey, Universitat Politècnica de Catalunya, Spain  
Michael L. Brodie, Verizon, USA  
Srinivasan Seshadri, Strand Genomics, India  
Seckin Unlu, Intel Corporation, USA  
Sophie Cluet, INRIA, France  
Umeshwar Dayal, Hewlett-Packard Laboratories, USA  
Vishu Krishnamurthy, Oracle Corporation, USA

## Workshop Selection

Nicolas Bruno, Microsoft, USA  
Gianni Mecca, Università della Basilicata, Italy  
Alberto Mendelzon, University of Toronto, Canada (co-chair)  
Beng Chin Ooi, National University of Singapore  
S. Sudarshan, IIT Bombay, India (co-chair)  
Janet Wiener, Hewlett-Packard, USA  
Peter Wood, Birkbeck College, UK  
Calisto Zuzarte, IBM Toronto, Canada

## Additional Reviewers

Ashraf Aboulnaga	Georges Dupret	Richard Chang Luo	Markus Sinnwell
Shipra Agrawal	Tanmoy Dutta	Stefan Manegold	Zhexuan Song
Josep Aguilar	Takeharu Eda	Ulrich Marquard	Jessica Staddon
Bugrahan Akcay	Said Elnaffar	Anurag Maskey	Jan Steffan
Juan M. Ale	Maged El-Sayed	Sergey Melnik	Uta Störl
Mehmet Altinel	Mehmet Erkanar	Paolo Merialdo	Liyang Sui
Toshiyuki Amagasa	Mauricio Minnuto Espil	Laurent Mignet	Sandeep Tata
Bernd Amann	Martin Ester	Gerome Miklau	Igor Tatarinov
Cristiana Amza	Daniel J. Farrar	Diego Milano	Yannis Theodoridis
Periklis Andritsos	Gabriela Flores	Paolo Missier	Wenhu Tian
Subramanian	Kei Fujimoto	Jun Miyazaki	Mehmet Nuri Tike
Arumugam	Michael Gillmann	Yasuhiro Mori	Leonardo Tininini
Ricardo Baeza-Yates	Jordi Gomez	Andreea E. Munteanu	Riccardo Torlone
Dirk Balfanz	Sergio Gomez	Victor Muntès	Panayiotis Tsaparas
Denilson Barbosa	Zeus Gomez	Jussi Myllymaki	Anthony Tung
Jonathan L. Beaver	Ozgür Gülderen	Benjamin Nguyen	Alexandra
Srikanta Bedathur	Yavuz Gürçan	Andrew Nierman	Uitdenbogerd
Michael Benedikt	Claudio Gutierrez	Baoning Niu	Seda Ünal
Omar Benjelloun	Marios Hadjieleftheriou	William O'Connell	Peter van Oosterom
Daniela Berardi	Maria Halkidi	Tadashi Ohmori	Alexander Vaschillo
Enrico Bertini	Richard A. Hankins	Beng Chin Ooi	Panos Vassiliadis
Kevin Beyer	Lilian Harada	Pietro Pala	Jeroen Vermeulen
Bodo Billerbeck	Kenji Hatano	HweeHwa Pang	Antonino Virgillito
Peter Boncz	Katia Hayati	Spiros Papadimitriou	Florian Waas
Roque Bonilla	Bingsheng He	Dimitris Papadopoulos	Jason Wang
Ivan T. Bowman	Marti Hearst	Cris Pedregal-Martin	Xiaoyu Wang
Paul Brown	Mauricio Hernandez	Michalis Petropoulos	Amol Wanjari
Luca Cabibbo	Tim Hoad	Abhijit Pol	Leejay Wu
Andrea Cali	Vagelis Hristidis	Mercel Pons	Tzu-Chiang Wu
Diego Calvanese	Jen-Chieh Huang	Vishy Poosala	Vilas Wuwongse
Adam Cannane	Yoshiharu Ishikawa	Rachel Pottinger	Ping Xia
Mario Cannataro	Anoop Jain	Wendy Powley	Yu Xu
Carlos Castillo	Dani Jimenez	Vikram Pudi	Wenwei Xue
Barbara Catania	Shantanu Joshi	Huiming Qu	John Yiannis
Ugur Cetintemel	Yildirak Kabak	Mukund Raghavachari	Ali Yildiz
Chee Yong Chan	Panagiotis Kalnis	Vijayshankar Raman	Haruo Yokota
Shyamal S. Chandra	Ibrahim Kamel	Pradeep Ravikumar	Masatoshi
Adriane Chapman	Nikos Karayannidis	Fred Reiss	Yoshikawa
Yun Chen	Norio Katayama	Flavio Rizzolo	Xiaohui Yu
Jin Chen	Stefan Katzenbeisser	Marie-Christine	Murat Yükselen
Krishna Chitrapura	Dao Dinh Kha	Rousset	Markos
Gregory Cobena	Jerry Kiernan	Esther Ryvkina	Zaharioudakis
Alex Coman	Stephen Kimani	Yasushi Sakurai	Jenny Zhang
Valter Crescenzi	Hiroyuki Kitagawa	Simonas Saltenis	Wei Zhang
Istvan Cseri	Johannes Klein	Parag Sarda	Dyce Jing Zhao
Susan Davidson	Yannis Kotidis	Monica Scannapieco	Farhana Zulkernine
Giuseppe De	Sailesh Krishnamurthy	Ken Sevcik	
Giacomo	A. Kumaran	Mehul Shah	
Amol Deshpande	Zoe Lacroix	Mohamed A. Sharaf	
Christian Digout	Gökce Banu Laleci	Christian Shelton	
Jens-Peter Dittrich	Philip Levis	Reza Sherkat	
Alin Dobra	Jessica Lin	Toshiyuki Shimizu	
	Bertram Ludaescher	Siyamed Sinir	

## Premier Sponsor



## Platinum Sponsors

ORACLE®

## Gold Sponsors



SYBASE®  
The Enterprise. Unwired.

## Silver Sponsors



InterSystems  
CACHE  
Make Applications Faster

## Academic Sponsors



University of Waterloo



University of Toronto

# **VLDB Endowment Board of Trustees**

The Very Large Data Base Endowment Inc. (VLDB Endowment) is a non-profit organization incorporated in the USA for the sole purpose of promoting and exchanging scholarly work in database system research and related fields throughout the world. The VLDB Endowment Board of Trustees has 21 members. The election procedure is documented elsewhere. The Board elects an Executive of 4 from among its members (see below). Some of the Trustees have specific responsibilities (as indicated below). All give their time freely to the management of the Endowment and to serving on special sub-committees for specific (time-limited) purposes.

The VLDB Endowment trustees are the legal guardians of its activities and charters. The trustees are elected among internationally distinguished researchers and professionals in the database field who have contributed to the objectives of the Endowment with dedication and distinction, and who are willing to commit their own time and resources. Trustees are elected for a six-year period, with one third up for election by the trustees every two years. The Endowment makes it a policy to spread the composition of the trustees across all continents. The Endowment does not have individuals as its members, nor does it have as members national computer societies; however, it closely co-operates with many international and national computer societies and other organizations.

## **President**

Gerhard Weikum

## **Vice-President**

Laura M. Haas

## **Secretary**

Hongjun Lu

## **Treasurer**

Michael J. Franklin

## **Members**

Amr El Abbadi  
Peter M. G. Apers  
Paolo Atzeni  
Philip A. Bernstein  
Elisa Bertino  
Peter Buneman  
Sophie Cluet  
Umeshwar Dayal  
Johann Christoph Freytag

H. V. Jagadish  
Christian S. Jensen  
Renée J. Miller  
Shojiro Nishio  
Krithi Ramamritham  
Raghu Ramakrishnan  
Hans-Jörg Schek  
Stanley B. Zdonik

Further information on the VLDB Endowment, its role and its activities is available on the World Wide Web at <http://www.vldb.org>.

## Welcome to VLDB'04

---

The international conference series on Very Large Data Bases (VLDB) was launched in 1975 in Framingham MA, about 20 miles from Boston. The conference was a huge success for its time, attracting almost 100 papers and more than 150 participants.

We have come a long way since then! VLDB conferences attract today hundreds of submissions and participants. Thanks to the efforts of programme committees, authors, and the VLDB Endowment over the years, VLDB conferences constitute today a prestigious scientific forum for the presentation and exchange of research results and practical experiences among members of the international Databases community.

VLDB conference regulars are familiar with Canada. This is the third time the conference is visiting, after Montreal (1980) and Vancouver (1992.) In all three occasions, the Canadian Databases community served as backbone for the organizing and the programme committees. At the same time, as with other years, the committees that put together this conference were international, with participation from all regions of the globe. We are grateful to all members of these committees for their time and efforts. Special thanks to Tamer Özsu, the general programme chair of this year's conference; also Kyu-Young Whang, who served as the VLDB Endowment liaison. Their spirit of cooperation throughout was invaluable.

Of course, the technical programme is not the only attraction of this year's conference. The conference hotel is located in the core downtown area of Toronto, within walking distance of museums, parks, shopping areas and tourist attractions such as the CN tower, the Harbourfront or Lake Ontario. Toronto has been called the most multicultural city in the world and while here, you will have the opportunity to visit neighborhoods that have the distinctive ethnic flavour of parts of the Far East and Europe. More than that, Toronto is a well-run, safe city that visitors enjoy visiting and revisiting.

Welcome to VLDB'04 and Toronto. We hope that you enjoy both the technical programme and the city!!

John Mylopoulos  
**General Chair, VLDB'04**

## Foreword from the Program Chairs

---

Welcome to the 30th International Conference on Very Large Databases (VLDB'04). VLDB Conferences are among the premier database meetings for dissemination of research results and for the exchange of latest ideas in the development and practice of database technology. The program includes two keynote talks, a 10-year award presentation, 81 research papers, 26 industrial papers (9 of which are invited), 5 tutorials, 2 panels and 34 demonstrations. It is a very rich program indeed.

This year we witnessed a significant jump of submissions. There were 504 research and industrial paper submissions, accounting for about 10% increase over last year and about 7% increase over 2002. Consequently the competition was fierce with an acceptance rate of 16.1% for research papers and about 40% for industrial papers.

The first keynote talk is by David Yach, who is the Senior Vice President of Software at Research in Motion (RIM). RIM is a leading designer, manufacturer and marketer of innovative wireless solutions for the worldwide mobile communications market. Their best-known product is the BlackBerry line of wireless handhelds. David oversees and manages the development of all lines of software at RIM. In his talk entitled "Databases in a Wireless World", David addresses the emerging environment where "information is stored not only in these central databases, but on a myriad of computers and computer-based devices in addition to the central storage. These range from desktop and laptop computers to PDA's and wireless devices such as cellular phones and BlackBerry's. The combination of large centralized databases with a large number and variety of associated edge databases effectively form a large distributed database, but one where many of the traditional rules and assumptions for distributed databases are no longer true." His talk discusses some of the new and challenging attributes of this new environment, particularly focusing on the challenges of wireless and occasionally connected devices.

Alon Halevy of the University of Washington gives the second keynote talk. His talk is entitled "Structures, Semantics and Statistics" and addresses the issues in integrating data from multiple sources. This is a problem that has occupied our community for a long time and has gained renewed importance with the emergence of the World Wide Web and the very many (and diverse) data sources that have become available on the Web. Alon's talk begins by "highlighting some of the significant recent achievements in the field of data integration, both in research and in industry." He then focuses on the main challenge going forward, namely, large-scale reconciliation of semantic heterogeneity, and on-the-fly information integration.

The ten-year best paper award this year goes to and Ramakrishnan Srikant for their paper entitled "Fast Algorithms for Mining Association Rules in Large Databases" that appeared in the 1994 VLDB Conference Proceedings. The Awards Committee (consisting of Masaru Kitsuregawa, Johann-Christoph Freytag, Raghu Ramakrishnan, Anastassia Ailamaki, Paolo Atzeni, and Limsoon Wong and chaired by Tamer Özsu) considered this to be one of the seminal papers in data mining. The paper identifies association rule mining and they discover a very nice property (a priori) that helps in pruning candidates in association rule mining. Rakesh and Ramakrishnan present a talk at this year's conference that focuses on the future of data mining.



The same committee has selected the paper "Model-Driven Data Acquisition in Sensor Networks" by Amol Deshpande, Carlos Guestrin, Samuel R. Madden, Joseph M. Hellerstein, and Wei Hong as the best paper.

The five tutorials that are scheduled cover a wide range of topics including core database topics as well as emerging issues in data management. The tutorials are the following:

- Database Architectures for New Hardware by Anastassia Ailamaki of Carnegie Mellon University
- Security of Shared Data in Large Systems by Arnon Rosenthal of Mitre Corporation and Marianne Winslett of University of Illinois at Urbana Champaign
- Self-Managing Technology in Database Management Systems by Surajit Chaudhuri of Microsoft Research, Benoit Dageville of Oracle, and Guy Lohman of IBM Almaden Research Lab.
- Architectures and Algorithms for Internet-Scale (P2P) Data Management by Joseph M. Hellerstein of University of California, Berkeley and Intel Research Berkeley
- The Continued Saga of DB-IR Integration by Ricardo Baeza-Yates of University of Chile and Mariano Consens of University of Toronto

There are two panels scheduled at this year's conference. The first panel is moderated by Thodoros Topaloglou of MDS Proteomics and is on "Biological Data Management: Research, Practice and Opportunities". The panel focuses on the data management problems that arise in the field of biological research. The panelists (Susan B. Davidson, H. V. Jagadish, Victor M. Markowitz, Evan W. Steeg, and Mike Tyers) discuss the ways in which database researchers can better serve the needs of biomedical research.

The second panel is entitled "Where is Business Intelligence taking today's database systems" and is moderated by William O'Connell of IBM Canada. The panelists are Andy Witkowski, Ramesh Bhashyam, Surajit Chaudhuri, Nigel Campbell. The panel addresses issues that arise in the production level deployment of business intelligence solutions (e.g., data mining, OLAP) over relational systems.

The technical program is the result of efforts by a large group of people. Three Program Committees were formed along themes (core database, infrastructure for information systems, and industrial and applications) consisting of 137 colleagues, each of whom reviewed about 13 papers. Raymond Ng and Matthias Jarke handled the tutorials, Jarek Gryz and Fred Lochovsky selected the panels, Bettina Kemme and David Toman assembled the demonstrations program. We thank them all for helping us put together an exciting program. We also thank Mario Nascimento for the excellent work he has done in putting together these Proceedings. We, along with Mario, also want to extend our thanks to DCC/UFAM in Brazil for the local support they have provided during the compilation of these proceedings while he was in Brazil.

M. Tamer Özsu  
Donald Kossmann  
Renée J. Miller  
José A. Blakeley  
Berni Schiefer

# CONTENTS

## KEYNOTES

Databases in a Wireless World .....	3
<i>David Yach (Research in Motion)</i>	
Structures, Semantics and Statistics.....	4
<i>Alon Halevy (Univ. of Washington)</i>	

## TEN YEAR BEST PAPER AWARD

Whither Data Mining?.....	9
<i>Rakesh Agrawal and Ramakrishnan Srikant (IBM Almaden)</i>	

## RESEARCH SESSIONS

### Research Session 1: Compression & Indexing

Compressing Large Boolean Matrices using Reordering Techniques .....	13
<i>David S. Johnson, Shankar Krishnan (AT&amp;T Labs-Research), Jatin Chhugani, Subodh Kumar (John Hopkins Univ.), Suresh Venkatasubramanian (AT&amp;T Labs-Research)</i>	
On the Performance of Bitmap Indices for High Cardinality Attributes .....	24
<i>Kesheng Wu, Ekow Otoo, Arie Shoshani (Lawrence Berkeley National Laboratory)</i>	
Practical Suffix Tree Construction.....	36
<i>Sandeep Tata, Richard A. Hankins, Jignesh M. Patel (Univ. of Michigan)</i>	

### Research Session 2: XML Views and Schemas

Answering XPath Queries over Networks by Sending Minimal Views .....	48
<i>Keishi Tajima, Yoshiki Fukui (JAIST)</i>	
A Framework for Using Materialized XPath Views in XML Query Processing.....	60
<i>Andrey Balmin, Fatma Özcan, Kevin Beyer, Roberta Cochrane, Hamid Pirahesh (IBM Almaden)</i>	
Schema-Free XQuery .....	72
<i>Yunyao Li, Cong Yu, H. V. Jagadish (Univ. of Michigan)</i>	

### Research Session 3: Controlling Access

Client-Based Access Control Management for XML Documents.....	84
<i>Luc Bouganim (INRIA Rocquencourt), François Dang Ngoc, Philippe Pucheral (PRISM Laboratory)</i>	
Secure XML Publishing without Information Leakage in the Presence of Data Inference .....	96
<i>Xiaochun Yang (Northeastern Univ.), Chen Li (Univ. of California, Irvine)</i>	