

# SIGMOD *record*

## TABLE OF CONTENTS

- 1** SIGMOD Officers, Committees, and Awardees
- 3** Editor's Notes
- Database Principles**
- 5** Declarative Networking: Recent Theoretical Work on Coordination, Correctness, and Declarative Semantics  
*Tom J. Ameloot*
- Research and Vision Articles**
- 17** Public Data and Visualizations: How are Many Eyes and Tableau Public Used for Collaborative Analytics?  
*Kristi Morton, Magdalena Balazinska, Dan Grossman, Robert Kosara, and Jock Mackinlay*
- 23** ENORM: An Essential Notation for Object-Relational Mapping  
*Alexandre Torres, Renata Galante, and Marcelo Pimenta*
- 29** Locating Valid SLCAs for XML Keyword Search with NOT Semantics  
*Rung-Ren Lin, Ya-Hui Chang, and Kun-Mao Chao*
- Systems and Prototypes**
- 35** Medusa: A Parallel Graph Processing System on Graphics Processors  
*Jianlong Zhong and Bingsheng He*
- Research Centers**
- 41** The IBM Research Accelerated Discovery Lab  
*Laura Haas, Melissa Cefkin, Cheryl Kieliszewski, Wil Plouffe, and Mary Roth*
- Event Reports**
- 49** Report on the First International Workshop on Exploratory Search in Databases and the Web (ExploreDB 2014)  
*Georgia Koutrika, Laks V.S. Lakshmanan, Mirek Riedewald, and Kostas Stefanidis*
- 53** Report on the Sixth International Workshop on Cloud Data Management (CloudDB 2014)  
*Shuai Ma, Xiaofeng Meng, and Fusheng Wang*
- 57** Report on the First Workshop on Linking and Contextualizing Publications and Datasets  
*Paolo Manghi, Lukasz Bolikowski, Nikos Houssos, and Jochen Schirrwagen*



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*



[www.sigmod.org](http://www.sigmod.org)

## SIGMOD Record

SIGMOD Record is a quarterly publication of the Special Interest Group on Management of Data (SIGMOD) of the Association for Computing Machinery (ACM). SIGMOD is dedicated to the study, development, and application of database and information technology. SIGMOD Record Web Edition is also freely available online at <http://www.sigmod.org/record>.

SIGMOD Record solicits contributions of articles, reports, and proposals for special sections. Conference announcements and calls for papers are published if relevant to the interests of the group and, in most cases, are limited to one page. The SIGMOD Record editorial policy is available at: <http://www.sigmod.org/publications/sigmod-record/sigmod-record-editorial-policy>.

Papers should be submitted electronically in PDF format through the SIGMOD REcord Electronic Submission System (RECESS) at <http://db.cs.pitt.edu/recess>. Detailed information for authors is available at: <http://www.sigmod.org/publications/sigmod-record/authors>.

By submitting your article for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights:

- to publish in print on condition of acceptance by the editor
- to digitize and post your article in the electronic version of this publication
- to include the article in the ACM Digital Library and in any Digital Library related services
- to allow users to make a personal copy of the article for noncommercial, educational or research purposes

However, as a contributing author, you retain copyright to your article and ACM will refer requests for republication directly to you. Therefore, ACM is asking all newsletter authors to include their contact information in their submissions. Opinions expressed in articles and letters are those of the author(s) and do not necessarily express the opinions of the ACM or SIGMOD. Author(s) should be contacted for reprint authorization.

### SIGMOD Record Editor-in-Chief:

Yanlei Diao, Associate Professor, University of Massachusetts Amherst  
Department of Computer Science Room 232, 140 Governors Drive, Amherst, MA 01003-9264  
<yanlei AT cs.umass.edu>

### Associate Editors:

- Denilson Barbosa, University of Alberta (Research Articles), <denilson AT cs.ualberta.ca>
- Pablo Barceló, Universidad de Chile (Database Principles), <pbarcelo AT dcc.uchile.cl>
- Vanessa Braganholo, Universidade Federal Fluminense (Distinguished Profiles in Data Management), <vanessa AT icc.uff.br>
- Marco Brambilla, Politecnico di Milano (Systems and Prototypes), <mbrambil AT elet.polimi.it>
- Chee Yong Chan, National University of Singapore (Research Articles), <chancy AT comp.nus.edu.sg>
- Rada Chirkova, North Carolina State University (Research Articles), <rychirko@ncsu.edu>
- Anish Das Sarma, Google (Reports), <anish AT google.com>
- Glenn Paulley, Conestoga College (Industry Perspectives), <gpaulley AT conestogac.on.ca>
- Alkis Simitsis, HP Labs (Research Centers), <alkis AT hp.com>
- Nesime Tatbul, ETH Zurich (Research Surveys), <tatbul AT inf.ethz.ch>
- Marianne Winslett, University of Illinois (Distinguished Profiles in Data Management), <winslett AT cs.uiuc.edu>

SIGMOD Record (ISSN 0163-5808) is published quarterly by the Association for Computing Machinery, 2 Penn Plaza, Suite 701, New York, NY 10121-0701. Periodicals postage paid at New York, NY 10001, and at additional mailing offices.

**POSTMASTER:** Send address changes to SIGMOD Record, ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701.

## SIGMOD Officers, Committees, and Awardees

Chair	Vice-Chair	Secretary/Treasurer
Donald Kossmann Systems Group ETH Zürich Cab F 73 8092 Zuerich SWITZERLAND +41 44 632 29 40 <donaIdk AT inf.ethz.ch>	Anastasia Ailamaki School of Computer and Communication Sciences, EPFL EPFL/IC/IIF/DIAS Station 14, CH-1015 Lausanne SWITZERLAND +41 21 693 75 64 <natassa AT epfl.ch>	Magdalena Balazinska Computer Science & Engineering University of Washington Box 352350 Seattle, WA USA +1 206-616-1069 <magda AT cs.washington.edu>

### SIGMOD Executive Committee:

Donald Kossmann (Chair), Anastasia Ailamaki (Vice-Chair), Magdalena Balazinska, K. Selçuk Candan, Yanlei Diao, Curtis Dyreson, Christian Jensen, Yannis Ioannidis, and Tova Milo.

### Advisory Board:

Raghu Ramakrishnan (Chair, Microsoft), Amr El Abbadi, Serge Abiteboul, Ricardo Baeza-Yates, Phil Bernstein, Elisa Bertino, Mike Carey, Surajit Chaudhuri, Christos Faloutsos, Alon Halevy, Joe Hellerstein, Renée Miller, C. Mohan, Beng-Chin Ooi, Z. Meral Ozsoyoglu, Sunita Sarawagi, Min Wang, and Gerhard Weikum.

### SIGMOD Information Director:

Curtis Dyreson, Utah State University <curtis.dyreson AT usu.edu>

### Associate Information Directors:

Manfred Jeusfeld, Georgia Koutrika, Wim Martens, Mirella Moro, Rachel Pottinger, and Jun Yang.

### SIGMOD Record Editor-in-Chief:

Yanlei Diao, University of Massachusetts Amherst <yanlei AT cs.umass.edu>

### SIGMOD Record Associate Editors:

Denilson Barbosa, Pablo Barceló, Vanessa Braganholo, Marco Brambilla, Chee Yong Chan, Rada Chirkova, Anish Das Sarma, Glenn Paulley, Alkis Simitsis, Nesime Tatbul, and Marianne Winslett.

### SIGMOD Conference Coordinator:

K. Selçuk Candan, Arizona State University

### PODS Executive Committee:

Rick Hull (Chair, IBM Research), Michael Benedikt, Wenfei Fan, Martin Grohe, Maurizio Lenzerini, Jan Paradaens.

### Sister Society Liaisons:

Raghu Ramakrishnan (SIGKDD), Yannis Ioannidis (EDBT Endowment), Christian Jensen (IEEE TKDE).

### Awards Committee:

Umesh Dayal (Chair, Hitachi America Ltd.), Elisa Bertino, Surajit Chaudhuri, Masaru Kitsuregawa, and Maurizio Lenzerini.

### Jim Gray Doctoral Dissertation Award Committee:

Tova Milo (Co-Chair, Tel Aviv University), Timos Sellis (Co-Chair, RMIT University), Ashraf Aboulnaga, Sudipto Das, Juliana Freire, Minos Garofalakis, Dan Suciu, Kian-Lee Tan.

[Last updated : June 30th, 2014]

## SIGMOD Edgar F. Codd Innovations Award

*For innovative and highly significant contributions of enduring value to the development, understanding, or use of database systems and databases.* Formerly known as the "SIGMOD Innovations Award", it now honors Dr. E. F. (Ted) Codd (1923 - 2003) who invented the relational data model and was responsible for the significant development of the database field as a scientific discipline. Recipients of the award are the following:

Michael Stonebraker (1992)	Jim Gray (1993)	Philip Bernstein (1994)
David DeWitt (1995)	C. Mohan (1996)	David Maier (1997)
Serge Abiteboul (1998)	Hector Garcia-Molina (1999)	Rakesh Agrawal (2000)
Rudolf Bayer (2001)	Patricia Selinger (2002)	Don Chamberlin (2003)
Ronald Fagin (2004)	Michael Carey (2005)	Jeffrey D. Ullman (2006)
Jennifer Widom (2007)	Moshe Y. Vardi (2008)	Masaru Kitsuregawa (2009)
Umeshwar Dayal (2010)	Surajit Chaudhuri (2011)	Bruce Lindsay (2012)
Stefano Ceri (2013)	Martin Kersten (2014)	

## SIGMOD Contributions Award

*For significant contributions to the field of database systems through research funding, education, and professional services.* Recipients of the award are the following:

Maria Zemankova (1992)	Gio Wiederhold (1995)	Yahiko Kambayashi (1995)
Jeffrey Ullman (1996)	Avi Silberschatz (1997)	Won Kim (1998)
Raghu Ramakrishnan (1999)	Michael Carey (2000)	Laura Haas (2000)
Daniel Rosenkrantz (2001)	Richard Snodgrass (2002)	Michael Ley (2003)
Surajit Chaudhuri (2004)	Hongjun Lu (2005)	Tamer Özsu (2006)
Hans-Jörg Schek (2007)	Klaus R. Dittrich (2008)	Beng Chin Ooi (2009)
David Lomet (2010)	Gerhard Weikum (2011)	Marianne Winslett (2012)
H.V. Jagadish (2013)	Kyu-Young Whang (2014)	

## SIGMOD Jim Gray Doctoral Dissertation Award

SIGMOD has established the annual SIGMOD Jim Gray Doctoral Dissertation Award to *recognize excellent research by doctoral candidates in the database field.* Recipients of the award are the following:

- **2006 Winner:** Gerome Miklau, University of Washington. *Runners-up:* Marcelo Arenas and Yanlei Diao.
- **2007 Winner:** Boon Thau Loo, University of California at Berkeley. *Honorable Mentions:* Xifeng Yan and Martin Theobald.
- **2008 Winner:** Ariel Fuxman, University of Toronto. *Honorable Mentions:* Cong Yu and Nilesh Dalvi.
- **2009 Winner:** Daniel Abadi, MIT. *Honorable Mentions:* Bee-Chung Chen and Ashwin Machanavajjhala.
- **2010 Winner:** Christopher Ré, University of Washington. *Honorable Mentions:* Soumyadeb Mitra and Fabian Suchanek.
- **2011 Winner:** Stratos Idreos, Centrum Wiskunde & Informatica. *Honorable Mentions:* Todd Green and Karl Schnaitterz.
- **2012 Winner:** Ryan Johnson, Carnegie Mellon University. *Honorable Mention:* Bogdan Alexe.
- **2013 Winner:** Sudipto Das, University of California, Santa Barbara. *Honorable Mention:* Herodotos Herodotou and Wenchao Zhou.
- **2014 Winners:** Aditya Parameswaran, Stanford University, and Andy Pavlo, Brown University.

A complete listing of all SIGMOD Awards is available at: <http://www.sigmod.org/awards/>

[Last updated : June 30th, 2014]

## Editor's Notes

Welcome to the June 2014 issue of the ACM SIGMOD Record!

The issue opens with a Database Principles article by Ameloot on recent theoretical work on declarative networking. This article addresses three important aspects of declarative networking: (a) *coordination* where nodes of a cloud are trying to obtain a global consensus, and in particular, the CALM conjecture by Hellerstein, (b) *correctness* of distributed computation, including decidability results, and (c) *languages* in declarative networking and *declarative semantics* for such languages, including the CRON conjecture by Hellerstein. The article concludes by outlining several directions for future work. This article presents a timely review of the recent theoretical work on declarative networking and offers a great introduction to this topic for the interested reader.

The Research and Vision Articles column features three articles. First, Morton, Balazinska, Grossman, Kosara, and Mackinlay, present an analysis of usage patterns of Many Eyes and Tableau Public, two popular Web-based, collaborative visual analytics systems. The analysis explores a number of primary dimensions of online visual analytics including the types of users, how users collaborate and interact, and how they analyze single datasets versus multiple data sources. The results of this study offer valuable information and insights towards building online visual analytics systems in the future. The second article, by Torres, Galante, and Pimenta, addresses the issue that today's Object-Relational Mappings are often platform dependent and deeply embedded in the code, which is difficult to read, understand, or evolve. This article presents ENORM, a notation that extends class models representing all the essential mappings and does so in a platform-independent manner. The third article, by Lin, Chang, and Chao, addresses keyword search queries over XML documents which may contain arbitrary combinations of AND, OR, and NOT operators. This article presents the concept of *valid Smallest Least Common Ancestors (SLCAs)* as query results, which eliminates erroneous results returned by previous algorithms, as well as an efficient algorithm to process such queries.

The Systems and Prototypes Column features the Medusa system developed by Zhong and He, which is a parallel graph processing system on graphics processors (GPUs). The core design of Medusa is to enable developers to leverage the massive parallelism and other hardware features of GPUs by writing sequential C/C++ code for a small set of APIs. The runtime system of Medusa then automatically executes the user-defined APIs in parallel on the GPU, with a number of optimizations based on the architecture features of GPUs and the characteristics of graph applications. A case study in social network analysis shows how Medusa improves both the coding productivity and the performance of graph operations.

In the Research Centers column, Haas, Cefkin, Kieliszewski, Plouffe, Roth, et al. describe the design and activities of the IBM Research Accelerated Discovery Lab. The lab is built on an analytics cloud environment with a unique software system that supports the process of discovery, facilitating collaboration and fostering insight. While many other groups have or are creating institutes that focus in one way or another in data-driven discovery, the IBM Research Accelerated Discovery Lab is unique in its emphasis on supporting the overall discovery process and the focus on understanding, from a social science perspective, how discovery happens and how it may be accelerated. These aspects are illustrated through a diversity of analytic and systems research projects that span disciplines and institutions, as well as through a study of the practice of discovery, with the goal to use the findings to better enable and accelerate discovery.

This issue features three event reports. Koutrika, Lakshmanan, Riedewald, and Stefanidis report on the First International Workshop on Exploratory Search in Databases and the Web (ExploreDB 2014), held in conjunction with EDBT/ICDT 2014. The report highlights the keynote by Prof. Keim on "Exploring Big

Data using Visual Analytics”, outlines a collection of six papers on various topics ranging from data exploration with structured database queries to search and ranking on the Web, and summarizes the panel discussion at the workshop. The report concludes by pointing out a number of research directions in areas including databases, Web search, multimedia exploration. In the second article, Ma, Meng, and Wang report on the Sixth International Workshop on Cloud Data Management (CloudDB 2014), co-located with ICDE 2014. The report covers the two keynote speeches, “Building Big Data Processing Systems under a New Computing Model” by Prof. Xiaodong Zhang and “Multi-faceted Classification of Big Data Uses and Proposed Architecture Integrating High Performance Computing and the Apache Stack” by Prof. Fox, as well as 10 research papers covering a wide range of topics from Quality of Service, to Query Processing, System Architecture, and Benchmarks. The report closes by pointing out some open problems including big data management in the cloud and cloud data security and privacy. The third article, by Manghi, Bolikowski, Houssos, and Schirrwagen, reports on the First Workshop on Linking and Contextualizing Publications and Datasets, held in conjunction with the 3<sup>rd</sup> International Conference on Theory and Practice of Digital Libraries (TPDL). This workshop addresses the need to interlink and contextualize datasets and scientific publications in sectors of scholarly communication and digital libraries. The report covers two keynote speeches and 10 research papers on topics including dataset contextualization, interlinking datasets and publications, and representing and visualizing datasets. The report concludes by outlining main considerations and future issues with respect to publications and datasets.

On behalf of the SIGMOD Record Editorial board, I hope that you will all enjoy reading the June 2014 issue of the SIGMOD Record.

Your submissions to the Record are welcome via the submission site:

<http://sigmod.hosting.acm.org/record>

Prior to submitting, you are encouraged to read the Editorial Policy on the SIGMOD Record’s Web site (<http://www.sigmod.org/publications/sigmod-record/sigmod-record-editorial-policy>).

Yanlei Diao

June 2014

#### Past SIGMOD Record Editors:

Harrison R. Morse (1969)  
Daniel O’Connell (1971 – 1973)  
Randall Rustin (1974-1975)  
Douglas S. Kerr (1976-1978)  
Thomas J. Cook (1981 – 1983)  
Jon D. Clark (1984 – 1985)  
Margaret H. Dunham (1986 – 1988)  
Arie Segev (1989 – 1995)  
Jennifer Widom (1995 – 1996)  
Michael Franklin (1996 – 2000)  
Ling Liu (2000 – 2004)  
Mario Nascimento (2005 – 2007)  
Alexandros Labrinidis (2007 – 2009)  
Ioana Manolescu (2009-2013)