

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Marianne Winslett (Ed.)

Scientific and Statistical Database Management

21st International Conference, SSDBM 2009
New Orleans, LA, USA, June 2-4, 2009
Proceedings

Volume Editor

Marianne Winslett
University of Illinois at Urbana-Champaign
Department of Computer Science
201 North Goodwin Avenue, Urbana, IL 61801, USA
E-mail: winslett@cs.uiuc.edu

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.2.8, H.2.5, H.3, H.4, J.1-3, I.3.7

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-642-02278-2 Springer Berlin Heidelberg New York
ISBN-13	978-3-642-02278-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12696451 06/3180 5 4 3 2 1 0

Preface

This volume contains the proceedings of SSDBM 2009, the 21st International Conference on Scientific and Statistical Database Management. SSDBM 2009 took place during June 2–4, 2009, at the Hotel Monteleone in New Orleans, USA. The SSDBM conference series brings together scientific domain experts, database researchers, practitioners, and developers for the presentation and exchange of current research concepts, tools, and techniques for scientific and statistical database applications. SSDBM organizers strive to provide a stimulating environment to encourage discussion, fellowship, and exchange of ideas in all aspects of research related to scientific and statistical databases, including both original research contributions and insights from practical system design, implementation, and evaluation.

SSDBM 2009 received 76 submissions from 18 countries. Each submission was reviewed by three Program Committee members, leading to the acceptance of 29 long papers and 12 short papers. The short papers include a mix of demonstrations, poster papers, and traditional conference presentations. This year we had the goal of increasing our acceptance rate while maintaining or increasing the quality of our papers; to this end, 17 of our accepted papers were shepherded.

This year we also benefitted from three invited talks. Our keynote presentation was from Kate Keahey of Argonne National Laboratory, who talked about scientific computing on cloud platforms. Bertram Ludaescher from the University of California Davis explained what makes scientific workflows scientific, and Arie Shoshani gave an overview of new technology developed at the Scientific Data Management Center at Lawrence Berkeley National Laboratory for exploring scientific datasets.

Organizing SSDBM 2009 was a team effort that involved many people. I thank our General Chairs Mahdi Abdelguerfi and Shengru Tu for their careful attention to so many details, and the SSDBM Steering Committee for their guidance. The Program Committee and our external referees did an excellent job with their timely review and careful discussion of all our submissions. I thank our 17 shepherds for their extra effort to ensure that SSDBM remains a premier forum. I also appreciate EasyChair's great facilities for assembling the camera-ready version of the proceedings, and Ragib Hasan's help in assembling the proceedings. We are also grateful to our conference sponsors, who provided financial support for SSDBM 2009: Louisiana Technology Group (LATG), Diamond Data Systems (DDS), Sun Microsystems, NOVACES, and the Department of Computer Science at the University of New Orleans.

I hope that you enjoy the proceedings!

Organization

Program Chair

Marianne Winslett University of Illinois at Urbana-Champaign,
USA

General Chair

Mahdi Abdelguerfi University of New Orleans, USA

General Co-chair

Shengru Tu University of New Orleans, USA

Program Committee

Walid Aref	Purdue University, USA
Ken Barker	University of Calgary, Canada
Randal Burns	Johns Hopkins University, USA
Sarah Cohen-Boulakia	University of Paris-Sud XI, France
Nilesh Dalvi	Yahoo!, USA
Amr El Abbadi	UC Santa Barbara, USA
Hakan Ferhatosmanoglu	Ohio State, USA
Minos Garofalakis	Yahoo! & UC Berkeley, USA
Dimitrios Gunopolis	UC Riverside, USA
Amarnath Gupta	UC San Diego, USA
H. V. Jagadish	University of Michigan, USA
George Kollios	Boston University, USA
Yunyao Li	IBM, USA
Xiaosong Ma	North Carolina State University, USA
Claudia Medeiros	University of Campinas, Brazil
Jignesh M. Patel	University of Michigan, USA
Sunil Prabhakar	Purdue University, USA
Louiqa Raschid	University of Maryland, USA
Rajeev Rastogi	Yahoo!, USA
Rishi Sinha	Microsoft, USA
Divesh Srivastava	AT&T Labs–Research, USA
Alex Szalay	Johns Hopkins University, USA
Kian-Lee Tan	National University of Singapore, Singapore
Anthony Tung	National University of Singapore, Singapore
Kesheng Wu	Lawrence Berkeley National Laboratory, USA
Cong Yu	Yahoo!, USA
Xiaofang Zhou	University of Queensland, Australia

Steering Committee

Arie Shoshani (Chair)	Lawrence Berkeley National Laboratory, USA
Moustafa A. Hammad	University of Calgary, Canada
Nick Koudas	University of Toronto, Canada
Bertram Ludaescher	University of California at Davis, USA
Nikos Mamoulis	University of Hong Kong, Hong Kong, China

Sponsors



External Reviewers

Alhajj, Reda
Antony, Shyam
Bacarin, Evandro
Bhattacharya, Souranghsu
Canahuate, Guadalupe
Cao, Jianneng
Chakravorty, Sayantan
Chatzimilioudis, George
Chiticariu, Laura
Cho, Brian
Cormode, Graham
Das, Sudipto
Deligiannakis, Antonios
Deng, Ke
Deshpande, Amey
Dhamankar, Robin
Digiampietri, Luciano
Dinakar, Divya
Eltabakh, Mohamed
Falcou, Joel
Fung, Gabriel
Gibas, Michael
Hakkoymaz, Huseyin
Hasan, Ragib
Huang, Zi
Ji, Feng

Koc, Levent
Koehler, Henning
Koonce, Steve
Kumar, Shailesh
Lappas, Theodoros
Li, Jiangtian
Mayfield, Chris
Minami, Kazuhiro
Mitra, Soumyadeb
Papapetrou, Panagiotis
Pastorello Jr., Gilberto
Patterson, Stacy
Qi, Yinian
Ramaswamy, Prakash
Sacan, Ahmet
Santanchè, André
Sanyal, Subhajit
Sengamedu, Srinivasan
Srinivasan, Jagannath
Sundararajan, S.
Termehchy, Arash
Wang, Shiyuan
Yalniz, Ismet
Zhang, Zhe

Table of Contents

1. Invited Presentation

The Scientific Data Management Center: Providing Technologies for Large Scale Scientific Exploration	1
<i>Arie Shoshani</i>	

2. Improving the End-User Experience

Query Recommendations for Interactive Database Exploration	3
<i>Gloria Chatzopoulou, Magdalini Eirinaki, and Neoklis Polyzotis</i>	
Scientific Mashups: Runtime-Configurable Data Product Ensembles	19
<i>Bill Howe, Harrison Green-Fishback, and David Maier</i>	
View Discovery in OLAP Databases through Statistical Combinatorial Optimization	37
<i>Cliff Joslyn, John Burke, Terence Critchlow, Nick Hengartner, and Emilie Hogan</i>	
Designing a Geo-scientific Request Language - A Database Approach . . .	56
<i>Peter Baumann</i>	
SEEDEEP: A System for Exploring and Querying Scientific Deep Web Data Sources	74
<i>Fan Wang and Gagan Agrawal</i>	
Expressing OLAP Preferences	83
<i>Matteo Golfarelli and Stefano Rizzi</i>	

3. Indexing, Physical Design, and Energy

Energy Smart Management of Scientific Data	92
<i>Ekow Otoo, Doron Rotem, and Shih-Chiang Tsao</i>	
Data Parallel Bin-Based Indexing for Answering Queries on Multi-core Architectures	110
<i>Luke J. Gosink, Kesheng Wu, E. Wes Bethel, John D. Owens, and Kenneth I. Joy</i>	
Finding Regions of Interest in Large Scientific Datasets	130
<i>Rishi Rakesh Sinha, Marianne Winslett, and Kesheng Wu</i>	

Adaptive Physical Design for Curated Archives	148
<i>Tanu Malik, Xiaodan Wang, Debabrata Dash, Amitabh Chaudhary, Anastasia Ailamaki, and Randal Burns</i>	
MLR-Index: An Index Structure for Fast and Scalable Similarity Search in High Dimensions	167
<i>Rahul Malik, Sangkyum Kim, Xin Jin, Chandrasekar Ramachandran, Jiawei Han, Indranil Gupta, and Klara Nahrstedt</i>	
4. Application Experience	
B-Fabric: An Open Source Life Sciences Data Management System	185
<i>Can Türker, Fuat Akal, Dieter Joho, and Ralph Schlapbach</i>	
Design and Implementation of Metadata System in PetaShare	191
<i>Xinqi Wang and Tevfik Kosar</i>	
Covariant Evolutionary Event Analysis for Base Interaction Prediction Using a Relational Database Management System for RNA	200
<i>Weijia Xu, Stuart Ozer, and Robin R. Gutell</i>	
5. Invited Presentation	
What Makes Scientific Workflows Scientific?	217
<i>Bertram Ludäscher</i>	
6. Workflow	
Enabling Ad Hoc Queries over Low-Level Scientific Data Sets	218
<i>David Chiu and Gagan Agrawal</i>	
Exploring Scientific Workflow Provenance Using Hybrid Queries over Nested Data and Lineage Graphs	237
<i>Manish Kumar Anand, Shawn Bowers, Timothy McPhillips, and Bertram Ludäscher</i>	
Data Integration with the DaltOn Framework – A Case Study	255
<i>Stefan Jablonski, Bernhard Volz, M. Abdul Rehman, Oliver Archner, and Olivier Curé</i>	
Experiment Line: Software Reuse in Scientific Workflows	264
<i>Eduardo Ogasawara, Carlos Paulino, Leonardo Murta, Cláudia Werner, and Marta Mattoso</i>	
Tracking Files in the Kepler Provenance Framework	273
<i>Pierre Moullem, Roselyne Barreto, Scott Klasky, Norbert Podhorszki, and Mladen Vouk</i>	

BioBrowsing: Making the Most of the Data Available in Entrez	283
<i>Sarah Cohen-Boulakia and Kevin Masini</i>	
Using Workflow Medleys to Streamline Exploratory Tasks	292
<i>Emanuele Santos, David Koop, Huy T. Vo, Erik W. Anderson, Juliana Freire, and Cláudio Silva</i>	

7. Query Processing

Experiences on Processing Spatial Data with MapReduce	302
<i>Ariel Cary, Zhengguo Sun, Vagelis Hristidis, and Naphtali Rish</i>	
Optimization and Execution of Complex Scientific Queries over Uncorrelated Experimental Data	320
<i>Ruslan Fomkin and Tore Risch</i>	
Comprehensive Optimization of Declarative Sensor Network Queries	339
<i>Ixent Galpin, Christian Y.A. Brenninkmeijer, Farhana Jabeen, Alvaro A.A. Fernandes, and Norman W. Paton</i>	
Efficient Evaluation of Generalized Tree-Pattern Queries with Same-Path Constraints	361
<i>Xiaoying Wu, Dimitri Theodoratos, Stefanos Souldatos, Theodore Dalamagas, and Timos Sellis</i>	
Mode Aware Stream Query Processing	380
<i>Mingrui Wei and Elke Rundensteiner</i>	
Evaluating Reachability Queries over Path Collections	398
<i>Panagiotis Bouros, Spiros Skiadopoulos, Theodore Dalamagas, Dimitris Sacharidis, and Timos Sellis</i>	

8. Similarity Search

Easing the Dimensionality Curse by Stretching Metric Spaces	417
<i>Ives R.V. Pola, Agma J.M. Traina, and Caetano Traina Jr.</i>	
Probabilistic Similarity Search for Uncertain Time Series	435
<i>Johannes Aßfalg, Hans-Peter Kriegel, Peer Kröger, and Matthias Renz</i>	
Reverse k-Nearest Neighbor Search Based on Aggregate Point Access Methods	444
<i>Hans-Peter Kriegel, Peer Kröger, Matthias Renz, Andreas Züfle, and Alexander Katzdobler</i>	
Finding Structural Similarity in Time Series Data Using Bag-of-Patterns Representation	461
<i>Jessica Lin and Yuan Li</i>	

9. Keynote Address

Cloud Computing for Science	478
<i>Kate Keahey</i>	

10. Mining

Classification with Unknown Classes	479
<i>Chetan Gupta, Song Wang, Umeshwar Dayal, and Abhay Mehta</i>	
HSM: Heterogeneous Subspace Mining in High Dimensional Data	497
<i>Emmanuel Müller, Ira Assent, and Thomas Seidl</i>	
Split-Order Distance for Clustering and Classification Hierarchies	517
<i>Qi Zhang, Eric Yi Liu, Abhishek Sarkar, and Wei Wang</i>	
Combining Multiple Interrelated Streams for Incremental Clustering....	535
<i>Zaigham Faraz Siddiqui and Myra Spiliopoulou</i>	
Improving Relation Extraction by Exploiting Properties of the Target Relation	553
<i>Eric Normand, Kevin Grant, Elias Ioup, and John Sample</i>	
<i>Cor-Split</i> : Defending Privacy in Data Re-publication from Historical Correlations and Compromised Tuples	562
<i>Daniele Riboni and Claudio Bettini</i>	
A Bipartite Graph Framework for Summarizing High-Dimensional Binary, Categorical and Numeric Data	580
<i>Guanhua Chen, Xiuli Ma, Dongqing Yang, Shiwei Tang, and Meng Shuai</i>	

11. Spatial Data

Region Extraction and Verification for Spatial and Spatio-temporal Databases	598
<i>Mark McKenney</i>	
Identifying the Most Endangered Objects from Spatial Datasets	608
<i>Hua Lu and Man Lung Yiu</i>	
Constraint-Based Learning of Distance Functions for Object Trajectories	627
<i>Wei Yu and Michael Gertz</i>	
Author Index	647