Proceedings

# 14th International Conference on Data Engineering

Proceedings

# 14th International Conference on Data Engineering

February 23-27, 1998

Orlando, Florida

*Sponsored by*

IEEE Computer Society Technical Committee on Data Engineering

Los Alamitos, California

Washington • Brussels • Tokyo

Copyright © 1998 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved

*Copyright and Reprint Permissions*: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

*The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.*

IEEE Catalog Number 98CB36164
ISBN: 0-8186-8289-2 (Softbound)
ISBN: 0-7803-4750-1 (Casebound)
ISBN: 0-8186-8291-4 (Microfiche)
ISSN: 1063-6382
PR8289

*Additional copies may be ordered from:*

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1-714-821-8380
Fax: + 1-714-821-4641
E-mail: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1-908-981-1393
Fax: + 1-908-981-9667
mis.custserv@computer.org

IEEE Computer Society
13, Avenue de l'Aquilon
B-1200 Brussels
BELGIUM
Tel: + 32-2-770-2198
Fax: + 32-2-770-8505
euro.ofc@computer.org

IEEE Computer Society
Ooshima Building
2-19-1 Minami-Aoyama
Minato-ku, Tokyo 107
JAPAN
Tel: + 81-3-3408-3118
Fax: + 81-3-3408-3553
tokyo.ofc@computer.org

Editorial production by Regina Spencer Sipple

Cover art production by Joseph Daigle/Studio Productions

Printed in the United States of America by The Printing House

# Table of Contents


**14th International Conference on Data Engineering (ICDE '98)**

**General Chair's Message** ..... xii
**Program Co-Chairs' Message** ..... xiii
**Organizing Committee** ..... xv
**Steering Committee** ..... xvii
**Program Committee Vice Chairs** ..... xviii
**Program Committee** ..... xix
**External Reviewers** ..... xxi

**Keynote Address**

**Databases: The Next Generation**
***Andrew Mendelsohn, Oracle Corporation***

**Session 1: Semistructured Data**

***Chair: Boris Chidlovskii, Rank Xerox Research Center***

Representing and Querying Changes in Semistructured Data ..... 4
*S.S. Chawathe, S. Abiteboul, and J. Widom*

Optimizing Regular Path Expressions Using Graph Schemas ..... 14
*M. Fernández and D. Suciu*

WebOQL: Restructuring Documents, Databases and Webs ..... 24
*G.O. Arocena and A.O. Mendelzon*

**Session 2: Mobile and Disconnected Database Operations**

***Chair: Anupam Joshi, Missouri University***

Concurrent Operations in a Distributed and Mobile Collaborative Environment ..... 36
*M. Suleiman, M. Cart, and J. Ferrié*

Grouping Techniques for Update Propagation in Intermittently Connected Databases ..... 46
*S. Mahajan, M. Donahoo, S. Navathe, M. Ammar, and S. Malik*

Cache Management for Mobile Databases: Design and Evaluation ..... 54
*B.Y. Chan, A. Si, and H.V. Leong*

**Industry Session 1: Database Engines**

***Chair: Surajit Chaudhuri, Microsoft Research***

Messaging/Queuing in Oracle8 ..... 66
*D. Gawlick*

The New Database Imperatives ..... 69
*G. Graefe*

**Panel Session 1:**

**WWW and the Internet — Did We (the Database Community) Miss the Boat?** ..... 74
***Chair: Michael Rabinovich, AT&T Laboratories***
***Panelists:*** *Mic Bowman, Transarc*
*Alon Levy, University of Washington*
*Susan Malaika, IBM Santa Teresa Lab*
*Alberto Mendelson, University of Toronto*
*Hector Garcia-Molina, Stanford University*

**Session 3: Parallel Database Systems**

***Chair: Kia Makki, University of Southwestern Louisiana***

Performance Analysis of Parallel Hash Join Algorithms on a Distributed Shared Memory Machine: Implementation and Evaluation on HP Exemplar SPP 1600 ..... 76
*M. Nakano, H. Imai, and M. Kitsuregawa*

Parallelizing Loops in Database Programming Languages ..... 86
*D.F. Lieuwen*

Cyclic Allocation of Two-Dimensional Data ..... 94
*S. Prabhakar, K. Abdel-Ghaffar, D. Agrawal, and A. El Abbadi*

**Industry Session 2: Data Mining**

***Chair: Meichun Hsu, Hewlett Packard Labs***

Employing Intelligent Agents for Knowledge Discovery ..... 104
*E. Stahl*

Industry Applications of Data Mining: Challenges & Opportunities ..... 105
*E. Simoudis*

**Session 4: Queries and Views**

***Chair: Ed Fernandez, Florida Atlantic University***

Fuzzy Triggers: Incorporating Imprecise Reasoning into Active Databases ..... 108
*A. Wolski and T. Bouaziz*

Graph Structured Views and Their Incremental Maintenance ..... 116
*Y. Zhuge and H. Garcia-Molina*

Query Folding with Inclusion Dependencies ..... 126
*J. Gryz*

On Querying Spreadsheets ..... 134
*L.V.S. Lakshmanan, S.N. Subramanian, N. Goyal, and R. Krishnamurthy*

**Session 5: Recovery and Buffer Management**

***Chair: Wojtek Cellary, University of Economics at Poznan***

Data Logging: A Method for Efficient Data Updates in Constantly Active RAIDs ..... 144
*E. Gabber and H.F. Korth*

Persistent Applications Using Generalized Redo Recovery ..... 154
*D.B. Lomet*

The Effect of Buffering on the Performance of R-Trees ..... 164
*S.T. Leutenegger and M.A. Lopez*

**Industry Session 3: OLAP and Data Warehousing**

***Chair: Kyuseok Shim, Bell Laboratories***

Red Brick Vista™: Aggregate Computation and Management ..... 174
*L.S. Colby, R.L. Cole, E. Haslam, N. Jazayeri, G. Johnson, W.J. McKenna, L. Schumacher, and D. Wilhite*

Outstanding Challenges in OLAP ..... 178
*J.A. Bedell*

Ending the ROLAP/MOLAP Debate: Usage Based Aggregation and Flexible HOLAP ..... 180
*C. Salka*

**Keynote Address**

**Safeguarding and Charging for Information on the Internet** ..... 182
***Hector Garcia-Molina, Steven P. Ketchpel, and Narayanan Shivakumar, Stanford University***

**Session 6: Spatial and Temporal Databases**

***Chair: Mostafa Bassiouni, University of Central Florida***

Point- Versus Interval-Based Temporal Data Models ..... 192
*M.H. Böhlen, R. Busatto, and C.S. Jensen*

Efficient Retrieval of Similar Time Sequences under Time Warping ..... 201
*B.-K. Yi, H.V. Jagadish, and C. Faloutsos*

Fast Nearest Neighbor Search in High-Dimensional Space ..... 209
*S. Berchtold, B. Ertl, D.A. Keim, H.-P. Kriegel, and T. Seidl*

**Session 7: Indexing in Data Warehousing**

***Chair: Ravi Mukkamala, Old Dominion University***

Encoded Bitmap Indexing for Data Warehouses ..... 220
*M.-C. Wu and A.P. Buchmann*

Coarse Indices for a Tape-Based Data Warehouse ..... 231
*T. Johnson*

Array-Based Evaluation of Multi-Dimensional Queries in Object-Relational Database Systems ..... 241
*Y. Zhao, K. Ramasamy, K. Tufte, and J.F. Naughton*

**Industry Session 4: Multimedia and Workflow Systems**

***Chair: Hector Garcia-Molina, Stanford University***

Content-Based Multimedia Information Management ..... 252
*R. Jain*

Building a Robust Workflow Management System with Persistent Queues and Stored Procedures ..... 254
*F. Leymann and D. Roller*

An Extended Object-Oriented Database Approach to Networked Multimedia Applications ..... 259
*H. Ishikawa, K. Kato, M. Ono, N. Yoshizawa, K. Kubota, and A. Kanaya*

**Session 8: Video and Image Databases**

***Chair: Ming-Syan Chen, National Taiwan University***

Distributed Video Presentations ..... 268
*E. Hwang, V.S. Subrahmanian, and B. Prabhakaran*

Query Processing in a Video Retrieval System ..... 276
*K.L. Liu, P. Sistla, C. Yu, and N. Rishe*

SEMCOG: A Hybrid Object-based Image Database System and Its Modeling, Language, and Query ..... 284
*W.-S. Li and K.S. Candan*

**Panel Session 2:**

**Data Warehousing Lessons from Experience** ..... 294

***Chair: Patrick O'Neil, University of Massachusetts, Boston***

***Panelists:*** *Richard Winter, Winter Consulting*
*Clark French, Sybase IQ*
*Dan Crowley, Informix Software*
*William McKenna, Red Brick Systems*

**Industry Session 5: The Web and Databases**

***Chair: Alex Buchmann, T.H. Darmstadt***

Junglee: Integrating Data of All Shapes and Sizes ..... 296
*A. Gupta*

Virtual Database Technology ..... 297
*A. Gupta, V. Harinarayan, and A. Rajaraman*

Data-Intensive Intra- & Internet Applications — Experiences Using Java and CORBA in the World Wide Web ..... 302
*J. Sellentin and B. Mitschang*

**Panel Session 3:**

**Migrating Legacy Databases and Applications** ..... 314

***Chair: Bhavani Thiraisingham, MITRE Corporation***

***Panelists:*** *Sandra Heiler, GTE Laboratories*
*Arnon Rosenthal, MITRE Corporation*
*Susan Malaika, IBM Santa Teresa Lab*
*Others TBD*

**Session 9: Architectural Issues in Data Mining Systems**

***Chair: Lois Delcambre, Oregon Graduate Center***

A Tightly-Coupled Architecture for Data Mining ..... 316
*R. Meo, G. Psaila, and S. Ceri*

A Distribution-Based Clustering Algorithm for Mining in Large Spatial Databases ..... 324
*X. Xu, M. Ester, H.-P. Kriegel, and J. Sander*

**Session 10: Cooperation and Workflow Management**

***Chair: Niki Pissinou, University of Southwestern Louisiana***

Failure Handling and Coordinated Execution of Concurrent Workflows ..... 334
*M. Kamath and K. Ramamritham*

Methodical Restructuring of Complex Workflow Activities ..... 342
*L. Liu and C. Pu*

Leveraging Mediator Cost Models with Heterogeneous Data Sources ..... 351
*H. Naacke, G. Gardarin, and A. Tomasic*

**Session 11: Access Structures**

***Chair: Thomas Mueck, University of Wien***

The LSD$^h$-tree: An Access Structure for Feature Vectors ..... 362
*A. Henrich*

Compressing Relations and Indexes ..... 370
*J. Goldstein, R. Ramakrishnan, and U. Shaft*

Generalizing "Search" in Generalized Search Trees ..... 380
*P.M. Aoki*

**Session 12: Association Rules and Dependencies**

***Chair: X. Hu, Sybase Inc.***

Efficient Discovery of Functional and Approximate Dependencies Using Partitions ..... 392
*Y. Huhtala, J. Kärkkäinen, P. Porkka, and H. Toivonen*

Online Generation of Association Rules ..... 402
*C.C. Aggarwal and P.S. Yu*

Cyclic Association Rules ..... 412
*B. Özden, S. Ramaswamy, and A. Silberschatz*

**Session 13: Concurrency Control**

***Chair: Vijay Atluri, Rutgers University***

Asynchronous Version Advancement in a Distributed Three Version Database ..... 424
*H.V. Jagadish, I.S. Mumick, and M. Rabinovich*

Design and Performance of an Assertional Concurrency Control System ..... 436
*A.J. Bernstein, D.S. Gerstl, W.-H. Leung, and P.M. Lewis*

Dynamic Granular Locking Approach to Phantom Protection in R-trees ..... 446
*K. Chakrabarti and S. Mehrotra*

**Plenary Panel Session:**

**Future Directions in Database Research** ..... 456

***Chair: Surajit Chaudhuri, Microsoft Corporation***

***Panelists:*** *Hector Garcia-Molina, Stanford University*
*Hank Korth, Bell Laboratories*
*Guy Lohman, IBM Almaden Research Center*
*David Lomet, Microsoft Research*
*David Maier, Oregon Graduate Institute*

**Session 14: Query Processing in Spatial Databases**

***Chair: Sharma Chakravarthy, University of Florida***

Processing Incremental Multidimensional Range Queries in a Direct Manipulation Visual Query Environment ..... 458
*S. Hibino and E. Rundensteiner*

High Dimensional Similarity Joins: Algorithms and Performance Evaluation ..... 466
*N. Koudas and K.C. Sevcik*

Cost Models for Join Queries in Spatial Databases ..... 476
*Y. Theodoridis, E. Stefanakis, and T. Sellis*

**Session 15: Mining of Association Rules**

***Chair: Michael Soo, University of South Florida***

Mining Association Rules: Anti-Skew Algorithms ..... 486
*J.-L. Lin and M.H. Dunham*

Mining for Strong Negative Associations in a Large Database of Customer Transactions ..... 494
*A. Savasere, E. Omiecinski, and S. Navathe*

Mining Optimized Association Rules with Categorical and Numeric Attributes ..... 503
*R. Rastogi and K. Shim*

**Session 16: Server-Based Database Systems**

***Chair: Anoop Singhal, AT&T Laboratories***

Remote Load-Sensitive Caching for Multi-Server Database Systems ..... 514
*S. Venkataraman, J.F. Naughton, and M. Livny*

DB-MAN: A Distributed Database System Based on Database Migration in ATM Networks ..... 522
*T. Hara, K. Harumoto, M. Tsukamoto, and S. Nishio*

Network Latency Optimizations in Distributed Database Systems ..... 532
*S. Banerjee and P.K. Chrysanthis*

**Session 17: Visualization of Multimedia Data**

***Chair: Tiziana Catarci, Universitá di Roma "La Sapienza"***

Global Integration of Visual Databases ..... 542
*W. Chang, D. Murthy, A. Zhang, and T.F. Syeda-Mahmood*

The Alps at Your Fingertips: Virtual Reality and Geoinformation Systems ..... 550
*R. Pajarola, T. Ohler, P. Stucki, K. Szabo, and P. Widmayer*
Design and Implementation of Display Specifications for Multimedia Answers ..... 558
*C. Baral, G. Gonzalez, and T.C. Son*

**Session 18: Management of Objects**

***Chair: Arbee Chen, National Tsing Hua University***

Flattening an Object Algebra to Provide Performance ..... 568
*P. Boncz, A.N. Wilschut, and M.L. Kersten*
Back to the Future: Dynamic Hierarchical Clustering ..... 578
*C. Zou, B. Salzberg, and R. Ladin*
Cost and Imprecision in Modeling the Position of Moving Objects ..... 588
*O. Wolfson, S. Chamberlain, S. Dao, L. Jiang, and G. Mendez*

**Research Prototype Demonstrations**

ROL: A Prototype for Deductive and Object-Oriented Databases ..... 598
*M. Liu, W. Yu, M. Guo, and R. Shan*
A Graphical Editor for the Conceptual Design of Business Rules ..... 599
*P. Lang, W. Obermair, W. Kraus, and T. Thalhammer*
The Active HYpermedia Delivery System (AHYDS) using the PHASME Application-Oriented DBMS ..... 600
*F. Andres and K. Ono*
ECA Rule Support for Distributed Heterogeneous Environments ..... 601
*S. Chakravarthy and R. Le*
ZEBRA Image Access System ..... 602
*S. Mudumbai, K. Shah, A. Sheth, K. Parasuraman, and C. Bertram*

**Author Index** ..... 603