

CONTENTS

Invited Talks

Atomicity versus Anonymity: Distributed Transactions for Electronic Commerce	1
<i>J. D. Tygar (CMU and UC Berkeley, USA)</i>	
Technology and the Future of Commerce and Finance	13
<i>D. E. Shaw (D. E. Shaw & Co.)</i>	

Research Sessions

Session: Text and Semistructured Data

Determining Text Databases to Search in the Internet	14
<i>Weiye Meng (SUNY Binghamton, USA), King-Lup Liu, Clement Yu (Univ. of Illinois, Chicago, USA), Xiaodong Wang, Yuhsi Chang (SUNY Binghamton, USA), Naphtali Rishe (Florida Intl. Univ., USA)</i>	
Proximity Search in Databases	26
<i>Roy Goldman, Narayanan Shivakumar, Suresh Venkatasubramanian, Hector Garcia-Molina (Stanford University, USA)</i>	
Incremental Maintenance for Materialized Views over Semistructured Data	38
<i>Serge Abiteboul (INRIA, France), Jason McHugh, Michael Rys, Vasilis Vassalos, Janet L. Wiener (Stanford University, USA)</i>	

Session: Performance and Multimedia

Performance Measurements of Tertiary Storage Devices	50
<i>Theodore Johnson (AT&T Labs, USA), Ethan L. Miller (University of Maryland, USA)</i>	
Active Storage For Large-Scale Data Mining and Multimedia	62
<i>Erik Riedel, Garth Gibson, Christos Faloutsos (Carnegie Mellon University, USA)</i>	
Resource Scheduling for Composite Multimedia Objects	74
<i>Minos N. Garofalakis (University of Wisconsin-Madison, USA), Yannis E. Ioannidis (University of Athens, Greece), Banu Özden (Bell Labs, USA)</i>	

Session: Join Processing

Hash joins and hash teams in Microsoft SQL Server	86
<i>Goetz Graefe, Ross Bunker, Shaun Cooper (Microsoft Corporation, USA)</i>	
Diag-Join: An Opportunistic Join Algorithm for 1:N Relationships	98
<i>Sven Helmer, Till Westmann, Guido Moerkotte (University of Mannheim, Germany)</i>	
Evaluating Functional Joins Along Nested Reference Sets in Object-Relational and Object-Oriented Databases	110
<i>Reinhard Braumandl, Jens Claussen, Alfons Kemper (Universität Passau, Germany)</i>	

Session: Heterogeneity and Interoperability

Using Schema Matching to Simplify Heterogeneous Data Translation	122
<i>Tova Milo, Sagit Zohar (Tel-Aviv University, Israel)</i>	
nD-SQL: A Multi-dimensional Language for Interoperability and OLAP	134
<i>Frédéric Gingras, Laks V.S. Lakshmanan (Concordia University, Canada)</i>	
The Heterogeneity Problem and Middleware Technology: Experiences with and Performance of Database Gateways	146
<i>Fernando de Ferreira Rezende, Klaudia Hergula (Daimler-Benz AG, Germany)</i>	

Session: Query Processing

Reducing the Braking Distance of an SQL Query Engine	158
<i>Michael J. Carey (IBM Almaden Research Center, USA), Donald Kossmann (University of Passau, Germany)</i>	
Querying Continuous Time Sequences	170
<i>Ling Lin, Tore Risch (Linköping University, Sweden)</i>	
Low-Cost Compensation-Based Query Processing	182
<i>Øystein Grøvlen (ClustRa, Norway), Svein-Olaf Hvasshovd (Norwegian Univ. of Science and Tech., Norway), Øystein Torbjørnsen (ClustRa, Norway)</i>	

Session: Similarity

A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces	194
<i>Roger Weber, Hans-J. Schek (ETH, Switzerland), Stephen Blott (Bell Labs, USA)</i>	
Improving Adaptable Similarity Query Processing by Using Approximations	206
<i>Mihael Ankerst, Bernhard Braunmüller, Hans-P. Kriegel, Thomas Seidl (University of Munich, Germany)</i>	
MindReader: Querying databases through multiple examples	218
<i>Yoshiharu Ishikawa (Nara Inst. of Science and Tech., Japan), Ravishankar Subramanya (Pittsburgh Supercomputing Center, USA), Christos Faloutsos (Carnegie Mellon University, USA)</i>	

Session: Query Optimization

Design and Analysis of Parametric Query Optimization Algorithms	228
<i>Sumit Ganguly (Indian Institute of Technology, Kanpur, India)</i>	
Inferring Function Semantics to Optimize Queries	239
<i>Mitch Cherniack, Stan Zdonik (Brown University, USA)</i>	
TOPAZ: a Cost-Based, Rule-Driven, Multi-Phase Parallelizer	251
<i>Clara Nippl, Bernhard Mitschang (Technische Universität München, Germany)</i>	

Session: Query Processing and Optimization

Filtering with Approximate Predicates	263
<i>Narayanan Shivakumar, Hector Garcia-Molina, Chandra S. Chekuri (Stanford University, USA)</i>	
Optimal Histograms with Quality Guarantees	275
<i>H. V. Jagadish (AT&T Labs, USA), Nick Koudas (Univ. of Toronto, Canada), S. Muthukrishnan, Viswanath Poosala (Bell Labs, USA), Ken Sevcik (Univ. of Toronto, Canada), Torsten Suel (Bell Labs)</i>	
Binding Propagation in Disjunctive Databases	287
<i>Sergio Greco (Università della Calabria, Italy)</i>	

Session: Data Mining and Warehousing

Computing Iceberg Queries Efficiently	299
<i>Min Fang, Narayanan Shivakumar, Hector Garcia-Molina, Rajeev Motwani, Jeffrey D. Ullman (Stanford University, USA)</i>	
Clustering Categorical Data: An Approach Based on Dynamical Systems	311
<i>David Gibson (UC Berkeley, USA), Jon Kleinberg (Cornell University, USA), Prabhakar Raghavan (IBM Almaden Research Center, USA)</i>	
Incremental Clustering for Mining in a Data Warehousing Environment	323
<i>Martin Ester, Hans P. Kriegel, Jörg Sander, Michael Wimmer, Xiaowei Xu (Univ. of Munich, Germany)</i>	

Session: High-Dimensional and Temporal Data

On Optimal Node Splitting for R-trees	334
<i>Yván García, Mario A. López, Scott T. Leutenegger (University of Denver, USA)</i>	
R-Tree Based Indexing of Now-Relative Bitemporal Data	345
<i>Rasa Bliujūte, Christian S. Jensen, Simonas Šaltenis, Giedrius Slivinskas (Aalborg Univ., Denmark)</i>	
Fast High-Dimensional Data Search in Incomplete Databases	357
<i>Beng Chin Ooi, Cheng Hian Goh, Kian-Lee Tan (National University of Singapore, Singapore)</i>	

Session: Data Mining I

On the Discovery of Interesting Patterns in Association Rules	368
<i>Sridhar Ramaswamy (Bell Labs, USA), Sameer Mahajan (Informix Corp., USA), Avi Silberschatz (Bell Labs, USA)</i>	
Algorithms for Mining Association Rules for Binary Segmentations of Huge Categorical Databases	380
<i>Yasuhiko Morimoto, Takeshi Fukuda, Hirofumi Matsuzawa, Takeshi Tokuyama, Kunikazu Yoda (IBM Tokyo Research Laboratory, Japan)</i>	
Algorithms for Mining Distance-Based Outliers in Large Datasets	392
<i>Edwin M. Knorr, Raymond T. Ng (University of British Columbia, Canada)</i>	

Session: Data Mining 2

PUBLIC: A Decision Tree Classifier that Integrates Building and Pruning	404
<i>Rajeev Rastogi, Kyuseok Shim (Bell Labs, USA)</i>	
RainForest - A Framework for Fast Decision Tree Construction of Large Datasets	416
<i>Johannes Gehrke, Raghu Ramakrishnan, Venkatesh Ganti (University of Wisconsin-Madison, USA)</i>	
WaveCluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases	428
<i>G. Sheikholeslami, S. Chatterjee, A. Zhang (SUNY at Buffalo, USA)</i>	

Session: System Issues 1

An Asynchronous Avoidance-Based Cache Consistency Algorithm for Client Caching DBMSs ..	440
<i>M. Tamer Özsu (GMD-IPSI, Germany), Kaladhar Voruganti (University of Alberta, Canada), Ronald C. Unrau (Cygnus Solutions, USA)</i>	
Design, Implementation, and Performance of the LHAM Log-Structured History Data Access Method	452
<i>Peter Muth (University of the Saarland, Germany), Patrick O'Neil (UMass/Boston, USA), Achim Pick, Gerhard Weikum (University of the Saarland, Germany)</i>	
Secure Buffering in Firm Real-Time Database Systems	464
<i>Binto George, Jayant R. Haritsa (Indian Inst. of Science, India)</i>	

Session: Data Warehousing

Small Materialized Aggregates: A Light Weight Index Structure for Data Warehousing	476
<i>Guido Moerkotte (Universität Mannheim, Germany)</i>	
Materialized View Selection for Multidimensional Datasets	488
<i>Amit Shukla, Prasad Deshpande, Jeffrey F. Naughton (University of Wisconsin-Madison, USA)</i>	
Expiring Data in a Warehouse	500
<i>Hector Garcia-Molina, Wilburt J. Labio, Jun Yang (Stanford University, USA)</i>	

Session: System Issues 2

Safely and Efficiently Updating References During On-line Reorganization	512
<i>Chendong Zou (Informix Software, USA), Betty Salzberg (Northeastern University, USA)</i>	
Buffering and Read-Ahead Strategies for External Mergesort	523
<i>Weiye Zhang, Per-Åke Larson (Microsoft Corporation, USA)</i>	
Bulk-Loading Techniques for Object Databases and an Application to Relational Data	534
<i>Sihem Amer-Yahia, Sophie Cluet (INRIA, France), Claude Delobel (University of Paris XI, France)</i>	

Session: Spatial Data

Algorithms for Querying by Spatial Structure	546
<i>Dimitris Papadias, Nikos Mamoulis (Hong Kong Univ. of Science & Tech., Hong Kong), Vasilis Delis (University of Patras, Greece)</i>	
A Raster Approximation For Processing of Spatial Joins	558
<i>Geraldo Zimbrão, Jano Moreira de Souza (Federal University of Rio de Janeiro, Brazil)</i>	
Scalable Sweeping-Based Spatial Join	570
<i>Lars Arge, Octavian Procopiuc (Duke University, USA), Sridhar Ramaswamy, Torsten Suel (Bell Labs, USA), Jeffrey S. Vitter (Duke University, USA)</i>	

Session: Data Mining 3

Ratio Rules: A New Paradigm for Fast, Quantifiable Data Mining	582
<i>Flip Korn, Alexandros Labrinidis, Yannis Kotidis (University of Maryland, USA), Christos Faloutsos (Carnegie Mellon Univ., USA)</i>	
Scalable Techniques for Mining Causal Structures	594
<i>Craig Silverstein, Sergey Brin, Rajeev Motwani, Jeff Ullman (Stanford University, USA)</i>	
Mining surprising patterns using temporal description length	606
<i>Soumen Chakrabarti, Sunita Sarawagi, Byron Dom (IBM Almaden Research Center, USA)</i>	

Industrial Sessions

Session: Database Vendors Meet Complex Query Languages

Massive Stochastic Testing of SQL	618
<i>Don Slutz (Microsoft Research, USA)</i>	
Selectivity Estimation in Extensible Databases — A Neural Network Approach	623
<i>Seetha Lakshmi, Shaoyu Zhou (Informix Inc., USA)</i>	
The Drill Down Benchmark	628
<i>Peter A. Boncz, Tim Rühl, Fred Kwakkel (Data Distilleries)</i>	

Session: Industrial Database Experience with Large Data Warehouses

Issues in Developing Very Large Data Warehouses	633
<i>Lyman Do, Pamela Drew, Wei Jin, Vish Jumani, David Van Rossum (Boeing, USA)</i>	
The National Medical Knowledge Bank	637
<i>Warren Sterling (NCR Parallel Systems, USA)</i>	
Bank of America Case Study: The Information Currency Advantage	641
<i>Felipe Cariño Jr., Mark Jahnke (NCR Teradata, USA)</i>	

Session: New Technology Database Vendor Offerings

DTL's DataSpot: Database Exploration Using Plain Language	645
<i>Shaul Dar, Gadi Entin, Shai Geva, Eran Palmon (Data Technologies Ltd. Israel)</i>	
From Data Independence to Knowledge Independence : An on-going Story	650
<i>Laurent Vieille (Next Century Media Inc., France)</i>	
Experiences in Federated Databases: From IRO-DB to MIRO-Web	655
<i>P. Fankhauser (GMD, Darmstadt), G. Gardarin (Univ. of Versailles, France), M. Lopez (GIE Dyade, France), J. Munoz (Ibermatica, Spain), A. Tomasic (INRIA, France)</i>	

Session: Database Vendor Internals: the Oracle Story

Materialized Views In Oracle	659
<i>Randall G. Bello, Karl Dias, Alan Downing, James Feenan, Jim Finnerty, William D. Norcott, Harry Sun, Andrew Witkowski, Mohamed Ziauddin</i>	
Checkpointing in Oracle	665
<i>Ashok Joshi, William Bridge, Juan Loaiza, Tirthankar Lahiri</i>	
Architecture of Oracle Parallel Server	669
<i>R. Bamford, D. Butler, B. Klots, and N. Macnaughton</i>	

Session: Database Vendor Kernels

KODA — The Architecture And Anterface Of A Data Model Independent Kernel	671
<i>Gopalan Arun, Ashok Joshi (Oracle Corp., USA)</i>	
The ADABAS Buffer Pool Manager	675
<i>Harald Schöning (Software AG, Darmstadt)</i>	
A Database System for Real-Time Event Aggregation in Telecommunication	680
<i>Jerry Baulier, Stephen Blott, Henry F. Korth, Avi Silberschatz (Bell Labs, USA)</i>	

Session: Database Vendor Support for Data Warehouses

Building Petabyte Databases with Objectivity/DB	not in proceedings
<i>Leong Guzenda (Objectivity Inc., USA)</i>	
Heterogeneous Database Query Optimization in DB2 Universal DataJoiner	685
<i>Shivakumar Venkataraman, Tian Zhang (IBM Santa Teresa Labs, USA)</i>	
Plan-Per-Tuple Optimization Solution — Parallel Execution of Expensive User-Defined Functions	690
<i>Felipe Cariño and William O'Connell (NCR Teradata, USA)</i>	

Panels

Is Web-site management a Database Problem?	696
<i>Daniela Florescu (INRIA, France), Alon Levy (Univ. of Washington, USA), Dan Suciu (AT&T Labs, USA)</i>	
Information, Communication, and Money: For What Can We Charge and How Can We Meter It?	697
<i>Stephen Blott, Henry F. Korth, Abraham Silberschatz (Bell Laboratories, USA)</i>	
Starting (and Sometimes Ending) a Database Company	697.5
<i>Jack Orenstein (Novera Software Inc., USA)</i>	

Industrial Exhibits

Overview	698
Bridging Heterogeneity: Research and Practice of Database Middleware Technology	699
<i>Daimler-Benz AG, Germany</i>	
The Cubetree Storage Organization	700
<i>Advanced Communication Technology, Inc., USA</i>	
DataBlitz: A High Performance Main-Memory Storage Manager	701
<i>Bell Laboratories, USA</i>	
DMS: A Parallel Data Mining Server	702
<i>Compaq, Ireland</i>	
IBM Industrial Exhibit	703
<i>IBM Corp., USA</i>	
MapInfo SpatialWare A Spatial Information Server for RDBMS	704
<i>MapInfo Corp., USA</i>	
Objectivity Industrial Exhibit	705
<i>Objectivity Inc., USA</i>	
Oracle Industrial Exhibit	706
<i>Oracle Corp., USA</i>	
A Single Pass Computing Engine for Interactive Analysis of VLDBs	707
<i>Mihalisin Associates, USA</i>	
Author Index	708