CONTENTS

Invited Talks

Keynote Address: The Changing Landscape of the Software Industry and its Implications for India 2
Umang Gupta (Centura Corporation)

Very Large Databases in a Commercial Application Environment 2
Karl-Heinz Hess (SAP AG Corporation)

Of Objects and Databases: A Decade of Turmoil 3
Michael J. Carey (IBM Almaden Research Center), David J. DeWitt (University of Wisconsin)

Research Sessions

(Asterisk (*) preceding a title indicates an experience/application paper)

Session: Spatial Access Methods

Filter Trees for Managing Spatial Data over a Range of Size Granularities 16
K. Sevcik, N. Koudas

The X-tree : An Index Structure for High-Dimensional data 28
S. Berchtold, D. A. Keim, H. -P. Kriegel

Analysis of n-Dimensional Quadtrees using the Hausdorff Fractal Dimension 40
C. Faloutsos, V. Gaede

Session: Data Warehousing

WATCHMAN : A Data Warehouse Intelligent Cache Manager 51
P. Scheuermann, J. Shim, R. Vingralek

Efficient Snapshot Differential Algorithms for Data Warehousing 63
W. J. Labio, H. Garcia-Molina

Incremental Maintenance of Externally Materialized Views 75
M. Staudt, M. Jarke

Session: Query Optimization

Optimization of Queries with User-defined Predicates 87
S. Chaudhuri, K. Shim

The Design and Implementation of a Sequence Database System 99
P. Seshadri, M. Livny, R. Ramakrishnan

*EROC: A Toolkit for Building NEATO Query Optimizers 111
W. McKenna, L. Burger, C. Hoang, M. Truong

Session: Data Mining

A new SQL-like Operator for Mining Association Rules 122
Rosa Meo, Giuseppe Psaila, Stefano Ceri

Sampling Large Databases for Association Rules 134
H. Toivonen

Constructing Efficient Decision Trees by Using Optimized Numeric Association Rules 146
Takeshi Fukuda, Yasuhiko Morimoto, Shinichi Morishita, Takeshi Tokuyama

Session: Query Processing - 1

Reordering Query Execution in Tertiary Memory Databases 156
Sunita Sarawagi, Michael Stonebraker

Query Processing Techniques for Multiversion Access Methods 168
J.v.d. Bercken, B. Seeger

Coalescing in Temporal Databases 180
M. H. Böhlen, R. T. Snodgrass, M. D. Soo

Session: Querying and Information Retrieval

Effective & Efficient Document Ranking without using a Large Lexicon 192
Y. Ogawa

PESTO : An Integrated Query/Browser for Object Databases 203
M. Carey, L. Haas, V. Maganty, J. Williams

Fast Nearest Neighbor Search in Medical Image Databases 215
F. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, Z. Protopapas

Session: Heterogeneous Databases - 1

Query Decomposition and View Maintenance for Query Languages
for Unstructured Data 227
D. Suciu

SchemaSQL - A Language for Interoperability in Relational Multi-database Systems 239
L. V. S. Lakshmanan, F. Sadri, I. N. Subramanian

Querying Heterogeneous Information Sources Using Source Descriptions 251
A. Levy, A. Rajaraman, J. Ordille

Session: Database Applications

*Supporting State-Wide Immunization Tracking Using Multi-Paradigm Workflow Technology 263
A. Sheth, K. Kochut, J. Miller, D. Worah, S. Das, C. Lin, D. Palaniswami, J. Lynch, I. Shevchenko

ZOO : A Desktop Experiment Management Environment 274
Y. E. Ioannidis, M. Livny, S. Gupta, N. Ponnekanti

*Applying Data Mining Techniques to a Health Insurance Information System 286
M. S. Viveros, J. P. Nearhos, M. J. Rothman

Session: Query Processing - 2

Querying Multiple Features of Groups in Relational Databases 295
D. Chatziantoniou, K. A. Ross

Modeling Skewed Distribution Using Multifractals and the '80-20' Law 307
C. Faloutsos, Y. Matias, A. Silberschatz

Answering Queries with Aggregation Using Views 318
D. Srivastava, S. Dar, H. V. Jagadish, A. Levy

Session: I/O Optimization

Semantic Data Caching and Replacement 330
S. Dar, M. J. Franklin, B. T. Jónsson, D. Srivastava, M. Tan

Clustering Techniques for Minimizing External Path Length 342
A. A. Diwan, S. Rane, S. Seshadri, S. Sudarshan

Disseminating Updates on Broadcast Disks 354
S. Acharya, M. J. Franklin, S. Zdonik

Session: Object Oriented Query Processing

Implementation and Analysis of a Parallel Collection Query Language 366
D. Suciu

Calibrating the Query Optimizer Cost Model of IRO-DB, an Object-Oriented Federated Database System 378
G. Gardarin, F. Sha, Z. -H. Tang

Cost-based Selection of Path Expression Processing Algorithms in Object-Oriented Databases 390
G. Gardarin, J. -R. Gruser, Z. -H. Tang

Session: Heterogeneous Databases - 2

Obtaining Complete Answers from Incomplete Databases 402
A. Levy

Object Fusion in Mediator Systems 413
Y. Papakonstantinou, S. Abiteboul, H. Garcia-Molina

The Role of Integrity Constraints in Database Interoperation 425
Mark W.W. Vermeer, Peter M.G. Apers

Session: Parallel Query Processing

Dynamic Load Balancing in Hierarchical Parallel Database Systems 436
L. Bouganim, D. Florescu, P. Valduriez

Estimation of Query-Result Distribution and its Application in Parallel-Join Load Balancing 448
V. Poosala, Y. Ioannidis

Intra-Transaction Parallelism in the Mapping of an Object Model to a Relational Multi-Processor System 460
M. Rys, M. C. Norrie, H. -J. Schek

Session: Data Integrity/Security

Supporting Periodic Authorizations and Temporal Reasoning in Database Access Control 472
E. Bertino, C. Bettini, E. Ferrari, P. Samarati

Querying a Multilevel Database: A Logical Analysis 484
F. Cuppens

Further Improvement on Integrity Constraint Checking for Stratifiable Deductive Databases 495
S. Y. Lee, T. W. Ling

Session: OLAP

On the Computation of Multidimensional Aggregates 506
S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, S. Sarawagi

Storage Estimation for Multidimensional Aggregates in the Presence of Hierarchies 522
A. Shukla, P. M. Deshpande, J. Naughton, K. Ramaswamy

Information Retrieval from an Incomplete Data Cube 532
C. Dyreson

Session: Potpourri

SPRINT: A Scalable Parallel Classifier for Data Mining 544
J. Shafer, R. Agrawal, M. Mehta

Modeling Design Versions 556
R. Ramakrishnan, D. Janaki Ram

*Integrating Triggers and Declarative Constraints in SQL Database Systems 567
R. Cochrane, H. Pirahesh, N. Mattos

Industry/Application Sessions

Industry/Application Session - A

Practical Issues with Commercial Use of Federated Databases 580
Jim Kleewein (IBM)

Supporting Procedural Constructs in SQL Compilers 581
Nelson Mattos (IBM)

Extracting Large Data Sets using DB2 Parallel Edition 582
Sriram Padmanabhan (IBM)

Industry/Application Session - B

Cache Coherency in Oracle Parallel Server 583
Boris Klots (Oracle)

Scalablity and Availability in Oracle7 7.3 584
Dieter Gawlick (Oracle)

Bellcore's ADAPT/X Harness System for Managing Information on Internet and Intranets 585
Amit Sheth (University of Georgia)

Industry/Application Session - C

DISNIC-PLAN: A NICNET Based Distributed Database for Micro-level Planning in India 586
M. Moni (National Informatics Centre, New Delhi)

Large Databases for Remote Sensing and GIS 587
A. R. Dasgupta (Space Applications Centre - ISRO)

DWMS: Data Warehouse Management System 588
Narendra Mohan (Software AG of Far East, Inc., Japan)

Industry/Application Session - D

The XPS Approach to Loading and Unloading Terabyte Databases 589
Sanket Atal (Informix Software Inc.)

How System 11 SQL Server Became Fast 590
T.K. Rengarajan (Sybase)

Using Referential Integrity To Easily Define Consistent Subset Replicas 591
Brad Hammond (Microsoft)

Industry/Application Session - E

The Query Optimizer in Tandem's new ServerWare SQL Product 592
Pedro Celis (Tandem Computers)

TPC-D: The Challenges, Issues and Results 593
Ramesh Bhashyam (NCR Corporation)

Tribeca: A Stream Database Manager for Network Traffic Analysis 594
Mark Sullivan (Bellcore)

Industry/Application Session - F

MineSet™: A System for High-End Data Mining and Visualization 595
Database Mining and Visualisation Group, SGI inc.

The Structured Information Manager: A Database System for SGML Documents 596
R. Sacks-Davis (RMIT, Australia)

Loading the Data Warehouse Across Various Parallel Architectures 597
Vijay Raghavan (Red Brick Systems, USA)

Panels

Performance of Future Database Systems: Bottlenecks and Bonananzas 599
Chaitanya K. Baru (San Diego Supercomputer Center)

DataBase Management Systems and the Internet 600
Susan Malaika (IBM Hursley, UK)

What is the data warehousing problem? (Are materialized views the answer?) 602
Ashish Gupta (Junglee Corpn.), Inderpal S. Mumick (AT&T Research)

Index of Authors 603

Vendor Session 2

DB2 Query Parllelism: Staging and Implementation 686
Y. Wang (IBM Corporation)

Managing a DB2 Parallel Edition Database 692
G. Fecteau (IBM corporation)

Vendor Session 3

DB2 Common Server:Technology Progress & Directions 694
B. Lindsay (IBM Corporation)

OS Support for VLDBs: Unix Enhancements for the Teradata Database 696
J. Catozzi, S. Rabinovici (ATT, Teradata)

Vendor Session 4

Accessing a Relational Database through an Object-Oriented Database Interface 702
J. Orenstein (Object Design Incorporated)

Vendor Session 5

Informix-Online XPS: A Dynamically Scalable RDBMS for Open Parallel Platforms 706
H. Spintzik (Informix Incorporated)

The Oracle Warehouse 707
G. Hallmark (Oracle Corporation)

Vendor Session 6

Efficient Search of Multidimensional B-Tree 710
H. Leslie, R. Jain, D. Birdsall, H. Yaghmai (Tandem Computers)

BigSur: A System for the Management of Earth Science Data 720
M. Stonebraker, P. Brown (Illustra)