

Table of Contents

Invited Talks

Tradeoffs between Parallel Database Systems, Hadoop, and HadoopDB as Platforms for Petabyte-Scale Analysis	1
<i>Daniel J. Abadi</i>	
Emerging Trends and Converging Technologies in Data Intensive Scalable Computing	4
<i>Roger S. Barga</i>	

Query Processing

Deriving Spatio-temporal Query Results in Sensor Networks	6
<i>Markus Bestehorn, Klemens Böhm, Patrick Bradley, and Erik Buchmann</i>	
Efficient and Adaptive Distributed Skyline Computation	24
<i>George Valkanias and Apostolos N. Papadopoulos</i>	
On the Efficient Construction of Multislices from Recurrences	42
<i>Romans Kasperovics, Michael H. Böhlen, and Johann Gamper</i>	
Optimizing Query Processing in Cache-Aware Wireless Sensor Networks	60
<i>Mario A. Nascimento, Romulo A.E. Alencar, and Angelo Brayner</i>	
Approximate Query Answering and Result Refinement on XML Data...	78
<i>Katja Seidler, Eric Peukert, Gregor Hackenbroich, and Wolfgang Lehner</i>	
Efficient and Scalable Method for Processing Top-k Spatial Boolean Queries	87
<i>Ariel Cary, Ouri Wolfson, and Naphtali Rishe</i>	

Scientific Data Management and Analysis

A Framework for Moving Sensor Data Query and Retrieval of Dynamic Atmospheric Events	96
<i>Shen-Shyang Ho, Wenqing Tang, W. Timothy Liu, and Markus Schneider</i>	

Client + Cloud: Evaluating Seamless Architectures for Visual Data Analytics in the Ocean Sciences	114
<i>Keith Grochow, Bill Howe, Mark Stoermer, Roger Barga, and Ed Lazowska</i>	
Scalable Clustering Algorithm for N-Body Simulations in a Shared-Nothing Cluster	132
<i>YongChul Kwon, Dylan Nunley, Jeffrey P. Gardner, Magdalena Balazinska, Bill Howe, and Sarah Loebman</i>	
Database Design for High-Resolution LIDAR Topography Data	151
<i>Viswanath Nandigam, Chaitan Baru, and Christopher Crosby</i>	
PetaScope: An Open-Source Implementation of the OGC WCS Geo Service Standards Suite	160
<i>Andrei Aiordăchioaie and Peter Baumann</i>	
Towards Archaeo-Informatics: Scientific Data Management for Archaeobiology	169
<i>Hans-Peter Kriegel, Peer Kröger, Christiaan Hendrikus van der Meijden, Henriette Obermaier, Joris Peters, and Matthias Renz</i>	
Data Mining	
DESSIN: Mining Dense Subgraph Patterns in a Single Graph	178
<i>Shirong Li, Shijie Zhang, and Jiong Yang</i>	
Discovery of Evolving Convoys	196
<i>Htoo Htet Aung and Kian-Lee Tan</i>	
Finding Top-k Similar Pairs of Objects Annotated with Terms from an Ontology	214
<i>Arnab Bhattacharya, Abhishek Bhowmick, and Ambuj K. Singh</i>	
Identifying the Most Influential User Preference from an Assorted Collection	233
<i>Hua Lu and Linhao Xu</i>	
MC-Tree: Improving Bayesian Anytime Classification	252
<i>Philipp Kranen, Stephan Günnemann, Sergey Fries, and Thomas Seidl</i>	
Non-intrusive Quality Analysis of Monitoring Data	270
<i>Mark Brightwell, Anastasia Ailamaki, and Anna Suwalska</i>	

Visual Decision Support for Ensemble Clustering	279
<i>Martin Hahmann, Dirk Habich, and Wolfgang Lehner</i>	

Indexes and Data Representation

An Indexing Scheme for Fast and Accurate Chemical Fingerprint Database Searching	288
<i>Zeyar Aung and See-Kiong Ng</i>	
BEMC: A Searchable, Compressed Representation for Large Seismic Wavefields	306
<i>Julio López, Leonardo Ramírez-Guzmán, Jacobo Bielak, and David O'Hallaron</i>	
Dynamic Data Reorganization for Energy Savings in Disk Storage Systems	322
<i>Ekow Otoo, Doron Rotem, and Shih-Chiang Tsao</i>	
Organization of Data in Non-convex Spatial Domains	342
<i>Eric Perlman, Randal Burns, Michael Kazhdan, Rebecca R. Murphy, William P. Ball, and Nina Amenta</i>	
PrefIndex: An Efficient Supergraph Containment Search Technique	360
<i>Gaoping Zhu, Xuemin Lin, Wenjie Zhang, Wei Wang, and Haichuan Shang</i>	
Supporting Web-Based Visual Exploration of Large-Scale Raster Geospatial Data Using Binned Min-Max Quadtree	379
<i>Jianting Zhang and Simin You</i>	

Scientific Workflow and Provenance

Bridging Workflow and Data Provenance Using Strong Links	397
<i>David Koop, Emanuele Santos, Bela Bauer, Matthias Troyer, Juliana Freire, and Cláudio T. Silva</i>	
LIVE: A Lineage-Supported Versioned DBMS	416
<i>Anish Das Sarma, Martin Theobald, and Jennifer Widom</i>	
Optimizing Resource Allocation for Scientific Workflows Using Advance Reservations	434
<i>Christoph Langguth and Heiko Schuldt</i>	
A Fault-Tolerance Architecture for Kepler-Based Distributed Scientific Workflows	452
<i>Pierre Mouallem, Daniel Crawl, Ilkay Altintas, Mladen Vouk, and Ustun Yildiz</i>	

Provenance Context Entity (PaCE): Scalable Provenance Tracking for Scientific RDF Data	461
<i>Satya S. Sahoo, Olivier Bodenreider, Pascal Hitzler, Amit Sheth, and Krishnaprasad Thirunarayan</i>	

Taverna, Reloaded	471
<i>Paolo Missier, Stian Soiland-Reyes, Stuart Owen, Wei Tan, Alexandra Nenadic, Ian Dunlop, Alan Williams, Tom Oinn, and Carole Goble</i>	

Similarity

Can Shared-Neighbor Distances Defeat the Curse of Dimensionality? ...	482
<i>Michael E. Houle, Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek</i>	

Optimizing All-Nearest-Neighbor Queries with Trigonometric Pruning	501
<i>Tobias Emrich, Franz Graf, Hans-Peter Kriegel, Matthias Schubert, and Marisa Thoma</i>	

Prefix Tree Indexing for Similarity Search and Similarity Joins on Genomic Data	519
<i>Astrid Rheinländer, Martin Knobloch, Nicky Hochmuth, and Ulf Leser</i>	

Similarity Estimation Using Bayes Ensembles	537
<i>Tobias Emrich, Franz Graf, Hans-Peter Kriegel, Matthias Schubert, and Marisa Thoma</i>	

Subspace Similarity Search: Efficient k-NN Queries in Arbitrary Subspaces	555
<i>Thomas Bernecker, Tobias Emrich, Franz Graf, Hans-Peter Kriegel, Peer Kröger, Matthias Renz, Erich Schubert, and Arthur Zimek</i>	

Data Stream Processing

Continuous Skyline Monitoring over Distributed Data Streams	565
<i>Hua Lu, Yonghuan Zhou, and Jonas Haustad</i>	

Propagation of Densities of Streaming Data within Query Graphs	584
<i>Michael Daum, Frank Lauterwald, Philipp Baumgärtel, and Klaus Meyer-Wegener</i>	

Spatio-temporal Event Stream Processing in Multimedia Communication Systems	602
<i>Mingyan Gao, Xiaoyan Yang, Ramesh Jain, and Beng Chin Ooi</i>	

Stratified Reservoir Sampling over Heterogeneous Data Streams	621
<i>Mohammed Al-Kateb and Byung Suk Lee</i>	
Tree Induction over Perennial Objects	640
<i>Zaigham Faraz Siddiqui and Myra Spiliopoulou</i>	
Author Index	659