



# Advances in Database Technology – EDBT 2012

---

15th International Conference  
on Extending Database Technology  
Berlin, Germany, March 27–30, 2012  
Proceedings

Editors:

Elke Rundensteiner (*Worcester Polytechnic Institute, USA*)  
Volker Markl (*Technische Universität Berlin, Germany*)  
Ioana Manolescu (*INRIA, France*)  
Sihem Amer-Yahia (*QCRI, Doha, Qatar*)  
Felix Naumann (*Hasso Plattner Institute, Potsdam, Germany*)  
Ismail Ari (*Ozyegin University, Turkey*)

Advances in Database Technology – EDBT 2012  
Proceedings of the 15th International Conference  
on Extending Database Technology  
Berlin, Germany, March 27–30, 2012

Editors:

Elke Rundensteiner  
Volker Markl  
Ioana Manolescu  
Sihem Amer-Yahia  
Felix Naumann  
Ismail Ari

The Association for Computing Machinery  
2 Penn Plaza, Suite 701  
New York, NY, 10121-0701

ACM COPYRIGHT NOTICE. Copyright c 2011 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept., ACM, Inc., fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, +1-978-750-8400, +1-978-750-4470 (fax).

ACM ISBN: 978-1-4503-0790-1

# Acknowledgements

The editors would like to thank the following people for their tremendous efforts to make this year's conference a success.

## **Local Arrangement Chairs**

Ralf-Detlef Kutsche

*Technische Universität Berlin, Germany*

Johann-Christoph Freytag

*Humboldt-Universität zu Berlin, Germany*

## **Publicity Chair**

Agnès Voisard

*Freie Universität Berlin and Fraunhofer ISST Berlin, Germany*

## **Sponsorship Chair**

Johann-Christoph Freytag

*Humboldt-Universität zu Berlin, Germany*

## **Website Chair**

Henning Agt

*Technische Universität Berlin, Germany*

# Table of Contents

Foreword . . . . .	viii
Program Committee Members . . . . .	ix–xi

## Invited Papers

Towards an Ecosystem for Structured Data on the Web <i>Alon Halevy</i> . . . . .	1-2
Inside "Big Data Management": Ogres, Onions, or Parfaits? <i>Michael J. Carey</i> . . . . .	3-14

## Research Sessions

### Map Reduce and Databases

Clydesdale: Structured Data Processing on MapReduce <i>Tim Kaldewey, Eugene Shekita, and Sandeep Tata</i> . . . . .	15-25
An Optimization Framework for Map-Reduce Queries <i>Leonidas Fegaras, Chengkai Li, and Upa Gupta</i> . . . . .	26-37
Efficient Parallel kNN Joins for Large Data in MapReduce <i>Chi Zhang, Feifei Li, and Jeffrey Jestes</i> . . . . .	38-49

### Privacy and Mining

Differentially Private Search Log Sanitization with Optimal Output Utility <i>Yuan Hong, Jaideep Vaidya, Haibing Lu, and Mingrui Wu</i> . . . . .	50-61
Integrating Historical Noisy Answers for Improving Data Utility under Differential Privacy <i>Shixi Chen, Shuigeng Zhou, and Sourav Bhowmick</i> . . . . .	62-73
Mining Probabilistically Frequent Sequential Patterns in Uncertain Databases <i>Zhou Zhao, Da Yan, and Wilfred Ng</i> . . . . .	74-85

### Recommendation and Social Networks

RecStore: An Extensible and Adaptive Framework for Online Recommender Queries inside the Database Engine <i>Justin Levandoski, Mohamed Sarwat, Mohamed Mokbel, and Michael Ekstrand</i> . . . . .	86-96
Supporting Top-K Item Exchange Recommendations in Large Online Communities <i>Zhan Su, Anthony Tung, and Zhenjie Zhang</i> . . . . .	97-108
Limiting Link Disclosure in Social Network Analysis through Subgraph-Wise Perturbation <i>Amin Milani Fard, Ke Wang, and Philip Yu</i> . . . . .	109-119

### Query Optimization

On Optimizing Relational Self-Joins <i>Yu Cao, Yongluan Zhou, Chee-Yong Chan, and Kian-Lee Tan</i> . . . . .	120-131
Transitive Closure and Recursive Datalog Implemented on Clusters <i>Foto Afrati and Jeffrey Ullman</i> . . . . .	132-143
Shortest-Path Queries for Complex Networks: Exploiting Low Tree-width Outside the Core <i>Takuya Akiba, Christian Sommer, and Ken-ichi Kawarabayashi</i> . . . . .	144-155

### Road Networks

User Oriented Trajectory Search for Trip Recommendation <i>Shuo Shang, Ruogu Ding, Bo Yuan, Kexin Xie, Kai Zheng, and Panos Kalnis</i> . . . . .	156-167
Top-k Spatial Keyword Queries on Road Networks <i>João B. Rocha-Junior, and Kjetil Nørkvåg</i> . . . . .	168-179
Relevance Search in Heterogeneous Networks <i>Chuan Shi, Xiangnan Kong, Philip Yu, Sihong Xie, and Bin Wu</i> . . . . .	180-191

### Data Summarization

An Adaptive Algorithm for Online Time Series Segmentation with Error Bound Guarantee <i>Zhenghua Xu, Rui Zhang, Ramamohanarao Kotagi, and Udaya Parampalli</i> . . . . .	192-203
---	---------

## Stream Data Processing

Transactional Stream Processing <i>Irina Botan, Peter Fischer, Donald Kossmann, and Nesime Tatbul</i> . . . . .	204-215
Dynamic Diversification of Continuous Data <i>Marina Drosou and Evaggelia Pitoura</i> . . . . .	216-227
Towards “Intelligent Compression” in Streams: A Biased Reservoir Sampling based Bloom Filter Approach <i>Sourav Dutta, Souvik Bhattacharjee, and Ankur Narang</i> . . . . .	228-239

## Preference Databases

Efficient Approximation of The Maximal Preference Scores by Lightweight Cubic Views <i>Yueguo Chen, Bin Cui, Xiaoyong Du, and Anthony Tung</i> . . . . .	240-251
Skyline-Sensitive Joins with LR-Pruning <i>Mithila Nagendra and K. Selcuk Candan</i> . . . . .	252-263
Top-k Interesting Phrase Mining in Ad-hoc Collections using Sequence Pattern Indexing <i>Chuancong Gao and Sebastian Michel</i> . . . . .	264-275

## Spatio Temporal Databases and OLAP

Optimizing Index Deployment Order for Evolving OLAP <i>Hideaki Kimura, Carleton Coffrin, Alexander Rasin, and Stanley Zdonik</i> . . . . .	276-287
Distance Histogram Computation Based on Spatiotemporal Uniformity in Scientific Data <i>Anand Kumar, Vladimir Grupcev, Yongke Yuan, Yi-Cheng Tu, and Gang Shen</i> . . . . .	288-299
A Generic Data Model and Query Language for Spatiotemporal OLAP Cube Analysis <i>Silvia Gómez, Leticia Gómez, and Alejandro Vaisman</i> . . . . .	300-311

## Web Systems

Subscription Indexes for Web Syndication Systems <i>Cedric Du Mouza, Zeinab Hmedhe, Harris Kourdounakis, Vassilis Christophides, Michel Scholl, and Nicolas Travers</i> . . . . .	312-323
Heuristics- based Query Optimisation for SPARQL <i>Petros Tsaliamanis, Lefteris Sidirourgos, Irimi Fundulaki, Vassilis Christophides, and Peter Boncz</i> . . . . .	324-335
See What’s enBlogue – Real-time Emergent Topic Identification in Social Media <i>Foteini Alvanaki, Sebastian Michel, Krithi Ramamritham, and Gerhard Weikum</i> . . . . .	336-347

## Textual Databases and Search

Searching by Corpus with Fingerprints <i>Charu Aggarwal, Wangqun Lin, and Philip Yu</i> . . . . .	348-359
Aggregate Queries on Probabilistic Record Linkages <i>Ming Hua and Jian Pei</i> . . . . .	360-371
Efficient Distributed Query Processing for Autonomous RDF Databases <i>Fabian Prasser, Alfons Kemper, and Klaus Kuhn</i> . . . . .	372-383

## Searching

CRSI: A Compact Randomized Similarity Index for Set-Valued Features <i>Petros Venetis, Yannis Sismanis, and Berthold Reinwald</i> . . . . .	384-395
VAST-Tree: A Vector-Advanced and Compressed Structure for Massive Data Tree Traversal <i>Takeshi Yamamuro, Makoto Onizuka, Toshio Hitaka, and Masashi Yamamuro</i> . . . . .	396-407
Repair-Oriented Relational Schemas for Multidimensional Databases <i>Mahkameh Yaghmaie, Leopoldo Bertossi, and Sina Ariyan</i> . . . . .	408-419

## Database Engines and Architectures

Adaptive MapReduce using Situation-Aware Mappers <i>Rares Vernica, Andrey Balmin, Kevin Beyer, and Vuk Ercegovic</i> . . . . .	420-431
"Cut Me Some Slack": Latency-Aware Live Migration for Databases <i>Sean Barker, Yun Chi, Hyun Moon, Hakan Hacigumus, and Prashant Shenoy</i> . . . . .	432-443
Peak Power Plays in Database Engines <i>Mayuresh Kunjir, Puneet Birwa, and Jayant Haritsa</i> . . . . .	444-455

## Graph Databases

Finding Top-K Similar Graphs in Graph Databases <i>Yuanyuan Zhu, Lu Qin, Jeffrey Xu Yu, and Hong Cheng</i> . . . . .	456-467
I/O Cost Minimization: Reachability Queries Processing over Massive Graphs <i>Zhiwei Zhang, Jeffrey Xu Yu, Lu Qin, Qing Zhu, and Xiaofang Zhou</i> . . . . .	468-479
Finding Maximal k-Edge-Connected Subgraphs from a Large Graph <i>Rui Zhou, Chengfei Liu, Jeffrey Xu Yu, Weifa Liang, Baichen Chen, and Jianxin Li</i> . . . . .	480-491

## Multi-dimensional Data Sets

SIMP: Accurate and Efficient Near Neighbor Search in High Dimensional Spaces <i>Vishwakarma Singh and Ambuj Singh</i> . . . . .	492-503
Effectively Indexing the Multi-Dimensional Uncertain Objects for Range Searching <i>Ying Zhang, Wenjie Zhang, Qianlu Lin, and Xuemin Lin</i> . . . . .	504-515
SFA: A Symbolic Fourier Approximation and Index for Similarity Search in High Dimensional Datasets <i>Patrick Schäfer and Mikael Högvist</i> . . . . .	516-527

## Industry and Applications Sessions

Introducing MapLan to Map Banking Survey Data into a Time Series Database <i>Manuel Günter</i> . . . . .	528-533
Extending a General-Purpose Streaming System for XML <i>Mark Mendell, Howard Nasgaard, Eric Bouillet, Martin Hirzel, and Bugra Gedik</i> . . . . .	534-539
Mining Search Behavior and User-Generated Content <i>Carlos Castillo</i> . . . . .	540-541
Data Management with SAPs In-Memory Computing Engine <i>Joos-Hendrik Böse and Cafer Tosun</i> . . . . .	542-544
Tailoring entity resolution for matching product offers <i>Hanna Köpcke, Andreas Thor, Stefan Thomas, and Erhard Rahm</i> . . . . .	545-550
Towards Principled Design Support for Scalable OLTP Workloads <i>Bin Liu, Junichi Tatemura, and Hakan Hacigümüs</i> . . . . .	551-553

## Tutorials

Similarity in (Spatial, Temporal and) Spatio-Temporal Databases <i>Dimitrios Gunopulos, and Goce Trajcevski</i> . . . . .	554-557
Distributed Skyline Processing: a Trend in Database Research Still Going Strong <i>Katja Hose and Akrivi Vlachou</i> . . . . .	558-561
Indexing and mining topological patterns for drug discovery <i>Sayan Ranu and Ambuj Singh</i> . . . . .	562-565
Adaptive Indexing in Modern Databases <i>Stratos Idreos, Stefan Manegold, and Goetz Graefe</i> . . . . .	566-569

## Demonstrations

A Probabilistic Convex Hull Query Tool for Animal Tracking <i>Zhou Zhao, Da Yan, and Wilfred Ng</i> . . . . .	570-573
ColisTrack: Testbed for a Pervasive Environment Management System <i>Yann Gripay, Frederique Laforest, Francois Lesueur, Nicolas Lumineau, Jean-Marc Petit, Vasile-Marian Scuturici, Samir Sebah, and Sabina Surdu</i> . . . . .	574-577
The Mainframe Strikes Back: Elastic Multi-Tenancy Using Main Memory Database Systems On A Many-Core Server <i>Henrik Muehe, Alfons Kemper, and Thomas Neumann</i> . . . . .	578-581
SOS (Save Our Systems): A uniform programming interface for non-relational systems <i>Paolo Atzeni, Francesca Bugiotti, and Luca Rossi</i> . . . . .	582-585
PeerTrack: A Platform for Tracking and Tracing Objects in Large-Scale Traceability Networks <i>Yanbo Wu, Quan Zheng Sheng, Damith Ranasinghe, and Lina Yao</i> . . . . .	586-589
Fault-Tolerant Complex Event Processing using Customizable State Machine-based Operators <i>Thomas Heinze, Zbigniew Jerzak, André Martin, Lenar Yazdanov, and Christof Fetzer</i> . . . . .	590-593
Knowledge-Based Processing of Complex Stock Market Events <i>Kia Teymourian, Malte Rohde, and Adrian Paschke</i> . . . . .	594-597

Private-HERMES: A Benchmark Framework for Privacy-Preserving Mobility Data Querying and Mining Methods	
<i>Nikos Pelekis, Aris Gkoulalas-Divanis, Marios Vondas, Anargyros Plemenos, Despina Kopanaki, and Yannis Theodoridis</i>	598-601
Evaluating Hybrid Queries through Service Coordination in HYPATIA	
<i>Victor Cuevas-Vicenttin, Genoveva Vargas-Solar, and Christine Collet</i>	602-605
A Desktop Interface over Distributed Document Repositories	
<i>Camélia Constantin, Cedric du Mouza, Philippe Rigaux, Virginie Thion, and Nicolas Travers</i>	606-609
SPARQL-RW: Transparent Query Access over Mapped RDF Data Sources	
<i>Konstantinos Makris, Nikos Bikakis, Nektarios Gioldasis, and Stavros Christodoulakis</i>	610-613
Intention Insider: Discovering People's Intentions in the Social Channel	
<i>Malu Castellanos, Meichun Hsu, Umesh Dayal, Ridhiman Ghosh, Mohamed Dekhil, Carlos Ceja, Marcial Puchi, and Perla Ruiz</i>	614-617
QUASAR: Querying Annotation, Structure, and Reasoning	
<i>Luying Chen, Michael Benedikt, and Evgeny Kharlamov</i>	618-621
Realtime Healthcare Services Via Nested Complex Event Processing Technology	
<i>Medhabi Ray, Mo Liu, Elke Rundensteiner, Chetan Gupta, Ismail Ari, Song Wang, and Daniel Dougherty</i>	622-625
Distributed Data Management for Large-Scale Wireless Sensor Networks Simulations	
<i>Stephen Wylie, James Heide, Besim Avci, Dennis Vaccaro, Oliviu Ghica, and Goce Trajcevski</i>	626-629
Knowing: A Generic Data Analysis Application	
<i>Thomas Bernecker, Franz Graf, Hans-Peter Kriegel, Nepomuk Seiler, Christoph Tuermer, and Dieter Dill</i>	630-633

# Foreword

Welcome to the 2012 Edition of the International Conference on Extending Database Technology (EDBT). This year, EDBT is taking place in Berlin, Germany on March 26-30, 2012. The International Conference on Extending Database Technology is known as a leading international forum for database researchers, practitioners, developers, and users to discuss cutting-edge ideas, and to exchange techniques, tools, and experiences related to data management. Data management constitutes the essential enabling technology for scientific, engineering, business, and social communities; and its technology is driven by the needs in applications ranging from the social web, scientific discoveries, virtual libraries, to embedded systems. We are very pleased to offer you an exciting program on these important and timely topics. This year, as in prior years, EDBT is co-located with the International Conference on Database Theory (ICDT).

The review process was conducted by a dedicated team of about 90 experts in the database field serving on the EDBT'2012 program committee. Having a large program committee allowed us to keep the workload for each committee member fairly low, and thus enabled the members to spend sufficient time on the review of each manuscript. We thank our colleagues on the program committee for their dedicated effort. These program committee members committed a huge amount of time to not only produce high-quality reviews but also to engage in subsequent on-line discussions to assure that each submitted manuscript was carefully considered. When deemed appropriate, additional rebuttals were requested from the authors to help clarify possible misunderstanding and enrich the discussion with additional knowledge. To assure highest quality, furthermore, 6 of the papers were accepted under shepherding, meaning that one program committee member volunteered their time to guide the authors to further improve their manuscripts based on the suggestions provided by the corresponding program committee members.

After careful deliberations, 43 out of the 193 submissions were accepted as full papers for final presentation at the conference and for publication in the EDBT'2012 proceedings. This, being to a highly competitive rate of a roughly 22.5% acceptance ratio, has no doubt resulted in a high-quality program for the 2012 EDBT meeting. Furthermore, for the EDBT/ICDT'2012 program, a record number of 42 demonstrations were submitted, out of which 16 demonstrations were accepted. The industrial program had 10 direct submissions as well as several invited industrial submissions, out of which 6 high quality works were selected for the final program. The program also includes 4 tutorials on timely topics. We are particularly proud to feature four keynote talks (joint with the ICDT conference) from distinguished researchers from both academia and industry, namely, Erich Graedel (RWTH Aachen University, Germany), Alon Halevy (Google Inc, USA), Michael J. Carey (UC Irvine, USA), and Wenfei Fan (University of Edinburgh, UK).

As in prior years, the EDBT'2012 meeting is augmented by a rich set of EDBT/ICDT workshops in timely topics, namely, Energy Data Management (EnDM2012), Data Analytics in the Cloud (DanaC), Linked Web Data Management (LWDM), Privacy and Anonymity in the Information Society (PAIS), and Business intelligence and the WEB (BEWEB). These workshops offer an excellent opportunity for discussions and brainstorming toward new areas of research. In addition, we again offer the "Joint EDBT/ICDT Ph.D. Workshop 2012" that provides a forum for Ph.D. students to present their dissertation work to the community and to receive expert feedback. In this context, we would like to thank NSF for providing partial travel funding for students to participate in EDBT/ICDT'2012.

The complete review process was managed using the Conference Management Toolkit (CMT), sponsored by Microsoft Research. We are grateful for the professional responsiveness we were repeatedly offered by the CMT Microsoft team throughout the process -- helping us with all aspects of the system from setup to particulars of the overall workflow process down to pulling material together for the final publication with ACM. Being extremely useful, CMT has become a de-facto standard in our community for handling such large number of papers and reviews.

We look forward to welcoming you at EDBT'2012, and hope you will enjoy the event and you have a pleasant stay in Berlin.

Elke Rundensteiner (Program Chair)  
Ioana Manolescu (Tutorials Chair)  
Sihem Amer-Yahia (Industrial & Applications Chair)

Felix Naumann (Demonstrations Chair)  
Volker Markl (General Chair)  
Ismail Ari (Proceedings Chair)



# Program Committee Members

## Research Reviewers

Karl Aberer	Paolo Atzeni	Shivnath Babu
Denilson Barbosa	Zohra Bellahsene	Elisa Bertino
Pedro Bizarro	Michael H. Boehlen	Angela Bonifati
Philippe Bonnet	Nicolas Bruno	Alejandro Buchmann
Malu Castellanos	Ugur Cetintemel	Badrish Chandramouli
Surajit Chaudhuri	Lei Chen	Yi Chen
Hong Cheng	Kajal Claypool	Isabel Cruz
Philippe Cudre-Mauroux	Alfredo Cuzzocrea	Mohamed Eltabakh
Shel Finkelstein	Johann-Christoph Freytag	Irini Fundulaki
Minos Garofalakis	Floris Geerts	Lukasz Golab
Goetz Graefe	Torsten Grust	Chetan Gupta
Theo Haerder	Jaiwei Han	Melanie Herschel
Vagelis Hristidis	Milena Ivanova	Chris Jermaine
Jaewoo Kang	Verena Kantere	Bettina Kemme
Alfons Kemper	Martin Kersten	Masaru Kitsuregawa
Yannis Kotidis	Georgia Koutrika	Hans-Peter Kriegel
Harumi Kuno	Laks Lakshmanan	Mong Li Lee
Kristen LeFevre	Xuemin Lin	Eric Lo
Qiong Luo	Murali Mani	Claudia Medeiros
Mirella M. Moro	Rimma Nehme	Beng Chin Ooi
Fatma Ozcan	Esther Pacitti	Norman Paton
Rachel Pottinger	Sunil Prabhakar	Christopher Re
Mirek Riedewald	Peter Scheuermann	Timos Sellis
Mohamed A. Sharaf	Liuba Shrira	Richard Snodgrass
Kian-Lee Tan	Nesime Tatbul	Evimaria Terzi
Tore Risch	Maurice Van-Keulen	Aparna S. Varde
Maria Esther Vidal	Stratis Viglas	Florian Waas
Song Wang	Cong Yu	Roberto V. Zicari
Calisto Zuzarte		

## **External Reviewers**

Zhaochen Guo	Gregory Leighton	Filipe Mesquita
Kostas Patroumpas	Dimitris Sacharidis	Manolis Terrovitis
Ying Zhang	Xiang Zhao	Zhitao Shen
Muhammad Aamir Cheema	Francesca Bugiotti	Luca Cabibbo
Alexander Ulanov	Sergey Serebryakov	Jaime Britton
Sireesha Gurralla	Kinnari Shah	Arunlal Kalyanasundaram
Kevin Ly	Sara Sutton	Theo Gantos
Muralidharan Venkatapathy	Manikanta Gunda	Alper Okcan
Silvia Stefanova	Andrej Andrejev	Thanh Truong
Kevin Luikens	Kyri Pavlou	Rui Zhang
Nihal Dindar	Andre Santanche	Matheus Motta
Karen Works	Medhabi Ray	Mikalai Tsytsarau
Sebastian Bdchle	Ekaterini Ioannou	Odysseas Papapetrou
George Giorgidze	Tom Schreiber	Alex Ulrich
Francois Goasdoui	Nicolas Faessel	Fabien Duchateau
Kjell Orsborn	Michele Catasta	Nguyen Quoc Viet Hung
Mehdi Riahi	Surender Yerva	Narendula Rammohan
Federico Caimi	Alessio Fabiani	

## **Industrial and Applications**

Michael Carey	Luna Dong	Jayant Haritsa
Jian Huang	Vanja Josifovski	Gjergji Kasneci
Robert Morton	C. Mohan	Victor Muntés-Mulero
Jose Carlos Cortizo	Jeremy Pickens	Philippe Pucheral
James G. Shanahan	Alkis Simitsis	Ioana Stanoiu
Fabian Suchanek	Oystein Torbjornsen	Yu Xu
Sergei Vassilvitskii		

## **Demonstrations**

Bernd Amann	Walid G. Aref	Mokrane Bouzeghoub
Stefan Desseloch	Wolfgang Gatterbauer	Francesco Guerra
Gregor Hackenbroich	Stefan Kramer	Wolfgang Lehner
Stefan Manegold	Amélie Marian	Daniela Nicklas
Meikel Poess	Ralf Schenkel	Marc Scholl
Thomas Seidl	Jayavel Shanmugasundaram	Pavel Shvaiko
Alkis Simitsis	Vasilis Vassalos	

**External Reviewers**

Dennis Geesen

Christian Beecks

Anca Ivanescu

Jens Kamenik

Sergej Fries

Philipp Kranen

Roland Assam

Marwan Hassani

Kevin Wilkinson