

# VLDB2017

43<sup>rd</sup> International Conference on Very Large Data Bases, Munich, Germany



## Proceedings of the VLDB Endowment

Volume 10, No. 11 – August 2017

**Proceedings of the 43rd International Conference on  
Very Large Data Bases, Munich, Germany**

Program Chairs:

**Peter Boncz and Ken Salem**

Associate Editors – Research Track:

**Ashraf Aboulnaga, Shimin Chen, Gautam Das, Amol Deshpande, Zack Ives, Qiong Luo, Stefan Manegold, Ioana Manolescu, Sharad Mehrotra, Fatma Ozcan, Themis Palpanas, Rachel Pottinger, Ken Ross, Gerhard Weikum**

Proceedings Chairs:

**Alvin Cheung, Aaron Elmore**

PVLDB – Proceedings of the VLDB Endowment

Volume 10, No. 11, August 2017.

The 43rd International Conference on Very Large Data Bases, Munich, Germany.

## **Copyright 2017 VLDB Endowment**

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>. For any use beyond those covered by this license, obtain permission by emailing [info@vldb.org](mailto:info@vldb.org).

Volume 10, Number 11, August 2017: VLDB 2017

Pages i – vii and 1166 – 1597

ISSN 2150-8097

Additional copies only online at: [portal.acm.org](http://portal.acm.org), [arxiv.org/corr](http://arxiv.org/corr), and [www.vldb.org](http://www.vldb.org)

## TABLE OF CONTENTS

### Front Matter

Copyright Notice .....	i
Table of Contents .....	ii
VLDB 2016 Organization and Review Board .....	v

### Research Papers

Memory Management Techniques for Large-Scale Persistent-Main-Memory Systems..... Ismail Oukid, Daniel Booss, Adrien Lespinasse, Wolfgang Lehner, Thomas Willhalm, Grégoire Gomes .....	1166
Trajectory Similarity Join in Spatial Networks .....	1178
.....Shuo Shang, Lisi Chen, Zhewei Wei, Christian Jensen, Kai Zheng, Panos Kalnis	
HoloClean: Holistic Data Repairs with Probabilistic Inference.....	1190
.....Theodoros Rekatsinas, Xu Chu, Ihab Ilyas, Chris Re	
Caribou: Intelligent Distributed Storage .....	1202
.....Zsolt Istvan, David Sidler, Gustavo Alonso	
Towards Linear Algebra over Normalized Data .....	1214
..... Lingjiao Chen, Arun Kumar, Jeffrey Naughton, Jignesh Patel	
Comparative Evaluation of Big-Data Systems on Scientific Image Analytics Workloads .....	1226
..... Parmita Mehta, Sven Dorkenwald, Dongfang Zhao, Tomer Kaftan, Alvin Cheung, Magdalena Balazinska, Ariel Rokem, Andrew Connolly, Jacob Vanderplas, Yusra AlSayyad .....	
Revenue Maximization in Incentivized Social Advertising .....	1238
..... Cigdem Aslay, Francesco Bonchi, Laks Lakshmanan, Wei Lu	
SquirrelJoin: Network-Aware Distributed Join Processing with Lazy Partitioning .....	1250
..... Lukas Rupperecht, William Culhane, Peter Pietzuch	
I've Seen "Enough": Incrementally Improving Visualizations to Support Rapid Decision Making ...	1262
..... Sajjadur Rahman, Maryam Aliakbarpour, Hidy Kong, Eric Blais, Karrie Karahalios, Aditya Parameswaran, Ronitt Rubinfeld .....	
Minimal On-Road Time Route Scheduling on Time-Dependent Graphs .....	1274
..... Lei Li, Wen Hua, Xingzhong Du, Xiaofang Zhou	
A holistic view of stream partitioning costs.....	1286
..... Nikos R. Katsipoulakis, Alexandros Labrinidis, Panos Chrysanthis	
Truss-based Community Search: a Truss-equivalence Based Indexing Approach .....	1298
..... Esra Akbas, Peixiang Zhao	
Query Optimization for Dynamic Imputation .....	1310
.....Jose Cambronero, John Feser, Micah Smith, Samuel Madden	

In Search of an Entity Resolution OASIS: Optimal Asymptotic Sequential Importance Sampling ... .....	Neil Marchant, Benjamin Rubinstein	1322
Flexible Online Task Assignment in Real-Time Spatial Data ..... .....	Yongxin Tong, Libin Wang, Zimu Zhou, Bolin Ding, Lei Chen, Jieping Ye, Ke Xu	1334
A Forward Scan based Plane Sweep Algorithm for Parallel Interval Joins..... .....	Panagiotis Bouros, Nikos Mamoulis	1346
ASAP: Prioritizing Attention via Time Series Smoothing ..... .....	Kexin Rong, Peter Bailis	1358
Knowledge Verification for LongTail Verticals ..... .....	Furong Li, Xin Luna Dong, Anno Langen, Yang Li	1370
SkyGraph: Retrieving Regions of Interest using Skyline Subgraph Queries ..... .....	Shiladitya Pande, Sayan Ranu, Arnab Bhattacharya	1382
Reverse Engineering Aggregation Queries ..... .....	Wei Chit Tan, Meihui Zhang, Hazem Elmeleegy, Divesh Srivastava	1394
LDA*: A Robust and Large-scale Topic Modeling System ..... .....	Lele Yu, Bin Cui, Ce Zhang, Yingxia Shao	1406
Social Hash Partitioner: A Scalable Distributed Hypergraph Partitioner ..... ..	Igor Kabiljo, Brian Karrer, Mayank Pundir, Sergey Pupyrev, Alon Shalita, Yaroslav Akhremtsev, Alessandro Presta	1418
On Sampling from Massive Graph Streams..... .....	Nesreen Ahmed, Nick Duffield, Theodore Willke, Ryan Rossi	1430
Pyramid Sketch: a Sketch Framework for Frequency Estimation of Data Streams..... .....	Tong Yang, Yang Zhou, Hao Jin, Shigang Chen, Xiaoming Li	1442
Reconciling Skyline and Ranking Queries ..... .....	Paolo Ciaccia, Davide Martinenghi	1454
CleanM: An Optimizable Query Language for Unified Scale-Out Data Cleaning..... .....	Stella Giannakopoulou, Manos Karpathiotakis, Benjamin Gaidioz, Anastasia Ailamaki	1466
Distributed Trajectory Similarity Search ..... .....	Dong Xie, Feifei Li, Jeff Phillips	1478
Runtime Optimization of Join Location in Parallel Data Management Systems..... .....	Bikash Chandra, S. Sudarshan	1490
Stitching Web Tables for Improving Matching Quality ..... .....	Oliver Lehmberg, Christian Bizer	1502
DigitHist: a Histogram-Based Data Summary with Tight Error Bounds..... .....	Michael Shekelyan, Anton Dignös, Johann Gamper	1514

Fast Scans on Key-Value Stores .....	1526
..... Markus Pilman, Kevin Bocksrocker, Lucas Braun, Renato Marroquín, Donald Kossmann	
Finding the maximum clique in massive graphs.....	1538
..... Can Lu, Jeffrey Yu, Hao Wei, Yikai Zhang	
Privacy-preserving Network Provenance .....	1550
..... Yuankai Zhang, Adam O'Neill, Micah Sherr, Wenchao Zhou	
Truth Discovery for SpatioTemporal Events from Crowdsourced Data .....	1562
..... Daniel Garcia Ulloa, Li Xiong, Vaidy Sunderam	
Data Vocalization: Optimizing Voice Output of Relational Data .....	1574
..... Immanuel Trummer, Jiancheng Zhu, Mark Bryan	
NoScope: Optimizing Deep CNN-Based Queries over Video Streams at Scale.....	1586
..... Daniel Kang, John Emmons, Firas Abuzaid, Peter Bailis, Matei Zaharia	

## VLDB 2017 ORGANIZATION AND REVIEW BOARD

### General Chairs

Alfons Kemper, TUM  
Thomas Neumann, TUM

### Honorary Chair

Johann-Christoph Freytag, HU Berlin

### Organization Committee Chair

Stephan Günemann, TUM  
Alfons Kemper, TUM  
Thomas Neumann, TUM

### Program Chairs and Editors in Chief of PVLDB 10

Peter Boncz, CWI  
Ken Salem, University of Waterloo

### Associate Editors of PVLDB 10

Ashraf Aboulnaga, Qatar Computing Research Institute  
Shimin Chen, Chinese Academy of Sciences  
Gautam Das, University of Texas at Arlington  
Amol Deshpande, University of Maryland  
Zack Ives, University of Pennsylvania  
Qiong Luo, HKUST  
Stefan Manegold, CWI  
Ioana Manolescu, INRIA  
Sharad Mehrotra, UC Irvine  
Fatma Ozcan, IBM Research  
Themis Palpanas, Paris Descartes University  
Rachel Pottinger, University of British Columbia  
Ken Ross, Columbia University  
Gerhard Weikum, MPI

### VLDB Endowment Representative

Volker Markl, TU Berlin

### Sponsorship Committee Chairs

Mike Carey, UC Irvine  
Sang Kyun Cha, University of Seoul  
Wolfgang Lehner, TU Dresden

### Publicity Committee Chair

Jens Dittrich, Saarland University

### Tutorial Chairs

Vagelis Hristidis, UC Riverside  
Aristides Gionis, Aalto University

### Industrial Chairs

Felix Naumann, HPI  
Jonathan Goldstein, Microsoft Research  
Jingren Zhou, Alibaba

### Demonstration Chairs

Martin Theobald, Ulm University  
Bingsheng He, NUS  
Reynold Xin, Databricks

### Panel Chairs

Stratos Idreos, Harvard  
Michael Brodie, MIT

### Workshop Chairs

Christian Jensen, Aalborg University

### PhD Workshop Chairs

Erhard Rahm, University of Leipzig  
Peter Christen, ANU  
Bettina Kemme, McGill University

### Proceedings Chairs

Aaron Elmore, University of Chicago  
Alvin Cheung, University of Washington

### Website Chair

Felix Martin Schuhknecht, Saarland University

### PVLDB Managing Editor

Divesh Srivastava, AT&T Labs

### PVLDB Information Director

Gerald Weber, University of Auckland

### PVLDB Advisory Committee

H.V. Jagadish, Tan Kian Lee, Renee Miller, S. Sudarshan,  
Juliana Freire, Tamer Ozsu, Chen Li, Wolfgang Lehner

## Research Track Review Board

Alan Fekete, Sydney University  
Alekh Jindal, Microsoft  
Alexander Löser, Beuth University of Applied Sciences Berlin  
Alexandros Labrinidis, University of Pittsburgh  
Allison Holloway, Oracle  
Angela Bonifati, Université Lyon 1  
Ansger Scherp, Kiel University  
Anthony Tung, NUS Singapore  
Aris Anagnostopoulos, Sapienza University of Rome  
Arnab Nandi, Ohio State University  
Arvind Arasu, Microsoft Research  
Asterios Katsifodimos, TU Berlin  
Atsuyuki Morishima, University of Tsukuba  
Avrilia Floratou, IBM Research Almaden  
Azza Abouzied, NYU Abu Dhabi  
Barzan Mozafari, University of Michigan  
Bernhard Seeger, University of Marburg  
Berthold Reinwald, IBM Research Almaden  
Bin Cui, Peking University  
Bingsheng He, Nanyang Technological University  
Bolin Ding, Microsoft Research  
Bongki Moon, SNU  
Boris Glavic, Illinois Institute of Technology  
Carmem Hara, Universidade Federal do Parana  
Chee-Yong Chan, National University of Singapore  
Chengkai Li, University of Texas at Arlington  
Chi Wang, Microsoft Research  
Chris Jermaine, Rice University  
Christian König, Microsoft Research  
Christina Lioma, Copenhagen University  
Cong Yu, Google  
Curtis Dyreson, Utah State University  
Cyrus Shahabi, University of Southern California  
Daisy Zhe, Wang, University of Florida  
Dan Olteanu, University of Oxford  
Daniel de Oliveira, Universidade Federal Fluminense  
David Koop, University of Massachusetts  
Davide Mottin, HPI  
Dmitri Kalashnikov, AT&T Labs Research  
Eli Cortez, Microsoft  
Elisa Bertino, Purdue University  
Eric Lo, Hong Kong Polytechnic University  
Essam Mansour, Qatar Computing Research Institute  
Eugene Wu, Columbia University  
Fabrizio Silvestri, Yahoo Research London  
Fei Chiang, McMaster University  
Feifei Li, University of Utah  
Florent Masegla, INRIA  
Florian Kerschbaum, SAP  
George Papadakis, University of Athens  
Georgia Koutrika, Hewlett Packard Labs  
Giansalvatore Mecca, University Basilicata  
Goetz Graefe, Hewlett Packard Labs  
Guoliang Li, Tsinghua University  
Hakan Ferhatosmanoglu, Bilkent University  
Hannes Voigt, TU Dresden

Hannes Mühleisen, CWI  
Harumi Kuno, Hewlett Packard Labs  
Henrik Muehe, Google  
Holger Pirk, MIT  
Huy Vo, CUNY-CCNY  
Ihab Ilyas, University of Waterloo  
Indrakshi Ray, Colorado State University  
Ingmar Weber, Qatar Computing Research Institute  
Ippokratis Pandis, Amazon Web Services  
Ira Assent, Aarhus University  
Jaewoo Kang, Korea University  
James Cheng, Chinese University of Hong Kong  
Jeff Pound, SAP  
Jeffrey Yu, Chinese University of Hong Kong  
Jennie Duggan, Northwestern University  
Jens Teubner, TU Dortmund  
Jiaheng Lu, University of Helsinki  
Jianliang Xu, Hong Kong Baptist University  
Jignesh Patel, University of Wisconsin  
Johann Gamper, Free University of Bozen-Bolzano  
Joseph Gonzalez, UC Berkeley  
Julia Stoyanovich, Drexel University  
Julien Leblay, AIST, Japan  
Kai-Uwe Sattler, TU Ilmenau  
Karthik Sankaranarayanan, IBM Research India  
Katja Hose, Aalborg University  
Khuzaima Daudjee, University of Waterloo  
Konstantinos Karanasos, Microsoft  
Kostis Kyzirakos, CWI  
Lee Mong Li, NUS Singapore  
Lefteris Sidirourgos, CWI  
Lei Zou, Peking University  
Li Xiong, Emory University  
Luc Bouganim, INRIA  
Luciano Barbosa, IBM Research Brazil  
Lucja Kot, Cornell University  
Mahashweta Das, Hewlett Packard Labs  
Marco Serafini, Qatar Computing Research Institute  
Martin Kersten, CWI  
Masatoshi Yoshikawa, Kyoto University  
Maurice Van Keulen, TU Twente  
Maya Ramanath, IIT Delhi  
Meichun Hsu, Hewlett Packard Labs  
Meikel Poess, Oracle  
Melanie Herschel, University of Stuttgart  
Michael Benedikt, Oxford University  
Michael Bohlen, University of Zurich  
Michael Hay, Colgate University  
Michael Grossniklaus, University of Konstanz  
Mirella Moro, Universidade Federal de Minas Gerais  
Mohamed Eltabakh, Worcester Polytechnic Institute  
Mohamed Mokbel, University of Minnesota  
Mohamed Sarwat, Arizona State University  
Mohammad Sadoghi, IBM Research T.J. Watson  
Mourad Ouzzani, Qatar Computing Research Institute  
Murat Kantarcioglu, UT Dallas  
Nan Zhang, George Washington University  
Nick Koudas, University of Toronto  
Nicolas Bruno, Microsoft Research

Nikolaus Augsten, University of Salzburg  
Nikos Mamoulis, Hong Kong University  
Norman Paton, University of Manchester  
Oliver Kennedy, University at Buffalo  
Panagiotis Papapetrou, Stockholm University  
Panos Kalnis, KAUST  
Panos Chrysanthis, University of Pittsburgh  
Paolo Meriardo, Roma Tre University  
Paris Koutris, University of Wisconsin-Madison  
Patricia Arocena, University of Toronto  
Peter Fischer, Universität Freiburg  
Peter Bailis, Stanford University  
Peter Alvaro, University of California, Santa Cruz  
Philippe Cudre-Mauroux, University of Fribourg  
Pierangela Samarati, University of Milan  
Pinar Tozun, IBM Research  
Raghav Kaushik, Microsoft Research  
Raluca Ada Popa, UC Berkeley  
Raymond Ng, University of British Columbia  
Reynold Cheng, Hong Kong University  
Ricardo Torres, UNICAMP Brazil  
S. Sudarshan, IIT Bombay  
Sai Wu, Zhejiang University  
Sebastian Michel, University of Kaiserslautern  
Selcuk Candan, Arizona State University  
Semih Salihoglu, University of Waterloo  
Senjuti Basu Roy, University of Washington Tacoma  
Seung-won Hwang, Yonsei University  
Sourav Bhowmick, Nanyang Technological University  
Spyros Blanas, Ohio State University  
Srikanta Bedathur, IBM Research India  
Stavros Papadopoulos, Intel Labs and MIT  
Stefanie Scherzinger, Ostbayerische Technische Hochschule Regensburg  
Stratis Viglas, University of Edinburgh  
Sudeepa Roy, Duke University  
Sudipto Das, Microsoft Research  
Sven Helmer, Free University of Bozen-Bolzano  
Tamer Ozsu, University of Waterloo  
Theodoros Rekatsinas, Stanford University  
Thomas Heinis, Imperial College  
Todd Green, Logicblox  
Torsten Grust, University of Tuebingen  
Tyson Condie, UCLA  
Umar Farooq Minhas, IBM Research  
Uwe Röhm, University of Sydney  
Verena Kantere, University of Geneva  
Viktor Leis, TU München  
Vivek Narasayya, Microsoft Research  
Wei Wang, University of New South Wales  
Wenchao Zhou, Georgetown University  
Wendy Wang, Stevens Institute of Technology  
Xiaochun Yang, Northeastern University, China  
Xiaodong Zhang, Ohio State University  
Xiaofang Zhou, University of Queensland  
Xiaohui Yu, York University  
Xiaoyang Wang, Fudan University  
Xin Luna Dong, Google

Yannis Manolopoulos, Aristotle University of Thessaloniki  
Yannis Velegrakis, University of Trento  
Yeye He, Microsoft Research  
Yi Chen, New Jersey Institute of Technology  
Yinan Li, Microsoft Research  
Yizhou Sun, Northeastern University  
Yoshiharu Ishikawa, Nagoya University  
Yuanyuan Tian, IBM Research (Almaden)  
Yufei Tao, University of Queensland  
Zhao Cao, Beijing Institute of Technology  
Zhifeng Bao, RMIT University  
Zoi Kaoudi, Qatar Computing Research Institute