

VLDB2018

44th International Conference on Very Large Data Bases, Rio de Janeiro, Brazil



Proceedings of the VLDB Endowment

Volume 11, No. 3 – November 2017

**Proceedings of the 44th International Conference on
Very Large Data Bases, Rio de Janeiro, Brazil**

Program Chairs:

Sihem Amer-Yahia and Jian Pei

Associate Editors – Research Track:

Luc Bouganim, Juliana Freire, Johannes Gehrke, Wook-Shin Han, Chris Jermaine, Jimmy Lin, Ioana Manolescu, Renee Miller, Mohamed Mokbel, Felix Naumann, Srinivasan Parthasarathy, Andrew Pavlo, S. Sudarshan, Jens Teubner, Yuanyuan Tian, Jianliang Xu, Meihui Zhang, Xiaodong Zhang

Proceedings Chairs:

Sourav Bhowmick, Ricardo Torres

PVLDB – Proceedings of the VLDB Endowment

Volume 11, No. 3, November 2017.

The 44th International Conference on Very Large Data Bases, Rio de Janeiro, Brazil.

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.

Volume 11, Number 3, November 2017: VLDB 2018

Pages i – vi and 243 - 392

ISSN 2150-8097

Additional copies only online at: portal.acm.org, arxiv.org/corr, and www.vldb.org

TABLE OF CONTENTS

Front Matter

Copyright Notice	ii
Table of Contents	iii
VLDB 2018 Organization and Review Board	iv

Research Papers

Efficient Structural Graph Clustering: An Index-Based Approach	
..... Dong Wen, Lu Qin, Ying Zhang, Lijun Chang, Xuemin Lin	243
An Authorization Model for Multi-Provider Queries	
..... Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Giovanni Livraga, Stefano Paraboschi, Pierangela Samarati	256
Snorkel: Rapid Training Data Creation with Weak Supervision	
..... Alexander Ratner, Stephen Bach, Henry Ehrenberg, Jason Fries, Sen Wu, Christopher Re	269
VERIFAS: A Practical Verifier for Artifact Systems.....	
..... Yuliang Li, Alin Deutsch, Victor Vianu	283
A Distributed Multi-GPU System for Fast Graph Processing	
..... Zhihao Jia, Yongkee Kwon, Galen Shipman, Pat McCormick, Mattan Erez, Alex Aiken	297
Efficient Denial Constraint Discovery with Hydra	
..... Tobias Bleifuß, Sebastian Kruse, Felix Naumann	311
ReCache: Reactive Caching for Fast Analytics over Heterogeneous Data.....	
..... Tahir Azim, Manos Karpathiotakis, Anastasia Ailamaki	324
Effective and Efficient Dynamic Graph Coloring.....	
..... Long Yuan, Lu Qin, Xuemin Lin, Lijun Chang, Wenjie Zhang	338
GPU Rasterization for Real-Time Spatial Aggregation over Arbitrary Polygons.	
.... Eleni Tzirita Zacharatou, Harish Doraiswamy, Anastasia Ailamaki, Claudio Silva, Juliana Freire	352
Are Key-Foreign Key Joins Safe to Avoid when Learning High-Capacity Classifiers?.....	
..... Vraj Shah, Arun Kumar, Xiaojin Zhu	366
Worker Recommendation for Crowdsourced Q&A Services: A Triple-Factor Aware Approach.	
..... Zheng Liu, Lei Chen	380

VLDB 2018 ORGANIZATION AND REVIEW BOARD

General Chairs

Alberto Laender, Universidade Federal de Minas Gerais
Fabio Porto, LNCC
Marco Antonio Casanova, PUC Rio

Honorary Chair

Antonio Furtado, PUC Rio
Nivio Ziviani, UFMG

Organization Committee Chair

Stephan Günnemann, TUM
Alfons Kemper, TUM
Thomas Neumann, TUM

Program Chairs and Editors in Chief of PVLDB 11

Jian Pei, Simon Fraser University
Sihem Amer-Yahia, University of Grenoble Alpes, CNRS

Associate Editors of PVLDB 11

Andrew Pavlo, Carnegie Mellon University
Chris Jermaine, Rice University
Felix Naumann, Hasso Plattner Institute
Ioana Manolescu, INRIA Saclay
Jens Teubner, TU Dortmund
Jianliang Xu, Hong Kong Baptist U.
Jimmy Lin, University of Waterloo
Johannes Gehrke, Microsoft
Juliana Freire, New York University
Luc Bouganim, INRIA
Meihui Zhang, SUTD
Mohamed Mokbel, University of Minnesota
Renee Miller, University of Toronto
Srinivasan Parthasarathy, Ohio State University
S. Sudarshan, IIT Bombay
Wook-Shin Han, Postech
Xiaodong Zhang, Ohio State University
Yuanyuan Tian, IBM Almaden

VLDB Endowment Representative

Divesh Srivastava, AT&T Labs-research

Sponsorship Committee Chairs

Artur Ziviani, LNCC
Anand Deshpande, Persistent
Mike Carey, University of California, UCI
Patrick Valduriez, INRIA

Publicity Committee Chair

Carmem Hara, Universidade Federal do Paraná
Mahashweta Das, Visa Research

Tutorial Chairs

Nick Koudas, University of Toronto
Sergio Lifschitz, PUC Rio

Industrial Chairs

Karin Breitman
Rakesh Agrawal, Data Insight Laboratories

Demonstration Chairs

Ming Hua, Facebook
Vanessa Braganholo, Universidade Federal Fluminense

Panel Chairs

Letizia Tanca, Politecnico di Milano
Mario Nascimento, Univ. Alberta

Workshop Chairs

Mirella Moro, Universidade Federal Minas Gerais
Xuemin Lin, University of New South Wales

PhD Workshop Chairs

Altigran Silva, Universidade Federal do Amazonas
Senjuti Basu Roy, New Jersey Institute of Technology

Proceedings Chairs

Ricardo Torres, University of Campinas
Sourav Bhowmick, Nanyang Technological University

Website Chair

Daniel De Oliveira, Universidade Federal Fluminense
Enver Anibal Choque Cayo, LNCC

PVLDB Managing Editor

Divesh Srivastava, AT&T Labs-research

PVLDB Advisory Committee

Juliana Freire, Jayant Haritsa, Wolfgang Lehner, Chen Li,
Renée J. Miller, Tova Milo, M. Tamer Özsu, Divesh
Srivastava, Kian-Lee Tan

Research Track Review Board

Abdul Quamar, IBM Almaden
Aijun An, York University, Canada
Alan Fekete, University of Sydney
Alex Thomo, University of Victoria
Alexandra Meliou, University of Massachusetts Amherst
Allison Holloway, Oracle
Anastasia Ailamaki, EPFL
Andrea Cali, Birkbeck Univ. of London, UK
Andrew Pavlo, Carnegie Mellon University
Anja Gruenheid, Google Research
Anshumali Shrivastava, Rice University
Antonios Deligiannakis, Technical University of Crete
Arbee Chen, Asia University, Taiwan
Aristides Gionis, Aalto University
Arnab Bhattacharya, IIT Kanpur
Arun Kumar, University of California, San Diego
Ashraf Aboulmaga, Qatar Computing Research Institute
Ashwin Machanavajjhala, Duke
Asterios Katsifodimos, TU Berlin, Germany
Atsuyuki Morishima, University of Tsukuba
Avrilia Floratou, Microsoft
Azza Abouzied, New York University Abu Dhabi
Baihua Zheng, Singapore Management University
Barzan Mozafari, University of Michigan
Beng Chin Ooi, NUS
Bernd Amann, Université Pierre et Marie Curie, France
Bolin Ding, Microsoft Research
Byron Choi, HKBU
Carlo Curino, Microsoft
Ce Zhang, ETH
Chee-Yong Chan, National University of Singapore
Chi Wang, Microsoft Research
Chris Jermaine, Rice University
Christopher Re, Stanford University
Chuan Xiao, Nagoya Univ., Japan
Cristian Bizer, University of Mannheim
Da Yan, University of Alabama
Daisy Zhe Wang, University of Florida
Dan Ports, Univ. of Washington USA
Dario Colazzo, U. Paris Dauphine
David Koop, U Mass Dartmouth
De-Nian Yang, Academia Sinica
Denilson Barbosa, University of Alberta - Canada
Divesh Srivastava, AT&T Labs Research
Elke Rundensteiner, WPI
Emmanuel Muller, Hasso Plattner Institute
Essam Mansour, QCRI
Fatma Ozcan, IBM Almaden
Feida Zhu, Singapore Management University
Feifei Li, University of Utah
Felix Naumann, Hasso Plattner Institute
Fernando Chirigati, NYU
Florian Rusu, University of California, Merced
Floris Geerts, University of Antwerp, USA

Francesco Bonchi, Yahoo Labs Barcelona, Spain
François Goasdoué, U. Rennes 1
Gao Cong, Nanyang Technological University
George Fletcher, Eindhoven University of Technology the Netherlands
George Kollios, Boston University
Georgia Koutrika, ATHENA Research Center, Greece
Gillian Dobbie, Univ. of Auckland, New Zealand
Guoliang Li, Tsinghua University
H. Jagadish, University of Michigan
Haibo Hu, Hong Kong Polytechnic
Haixun Wang, Facebook
Hakan Ferhatosmanoglu, Bilkent University
Harish Doraiswamy, NYU
Haryadi Gunawi, Univ. of Chicago, USA
Heiko Mueller, NYU
Herodotos Herodotou, Cyprus University of Technology
Holger Pirk, MIT
Hong Cheng, Chinese University of Hong Kong
Huy Vo, CUNY
Ihab Ilyas, University of Waterloo
Immanuel Trummer, EPFL, Switzerland
Ioana Manolescu, INRIA Saclay, France
Ira Assent, University of Aarhus
Isabel Cruz, University of Illinois at Chicago, USA
Jaideep Vaidya, Rutgers University
James Cheng, Chinese University of Hong Kong, Hong Kong
Jeffrey Xu Yu, Chinese University of Hong Kong, Hong Kong
Jens Dittrich, Saarland University
Jens Teubner, TU Dortmund
Jianliang Xu, Hong Kong Baptist U.
Jiannan Wang, SFU
Jignesh Patel, UW - Madison
Jimmy Lin, University of Waterloo
Jing Gao, State University of New York at Buffalo
Johannes Gehrke, Microsoft
Jonathan Goldstein, Microsoft
Josep Domingo-Ferrer, Universitat Rovira i Virgili, Catalonia
Ju Fan, Remin University
Juliana Freire, New York University
Justin Levandoski, Microsoft Research
Karthik Ramachandra, Microsoft Gray Systems Lab
Ke Wang, SFU
Ken Barker, University of Calgary, Canada
Khuzaima Daudjee, University of Waterloo
Kyuseok Shim, Seoul National University
Laks V.S. Lakshmanan, The University of British Columbia
Lei Chen, Hong Kong University of Science and Technology
Letizia Tanca, Politecnico di Milano, Italy
Li Xiong, Emory University
Luc Bouganim, INRIA
Lukasz Golab, University of Waterloo

Luna Dong, Amazon.com
Magdalena Balanziska, University of Washington
Marco Serafini, Qatar Computing Research Institute, Qatar
Maria Damiani, University of Milano, Italy
Martin Theobald, University of Luxembourg
Masaru Kitsuregawa, University of Tokyo, Japan
Matteo Golfarelli, University of Bologna, Italy
Matthias Boehm, IBM Almaden
Matthias Renz, George Mason Univ., USA
Matthieu Latapy, LIP6, France
Meihui Zhang, SUTD
Melanie Herschel, Universität Stuttgart
Michael Cafarella, University of Michigan
Mohamed Mokbel, University of Minnesota
Mohamed Sharaf, University of Queensland, Australia
Mohamed Sarwat, ASU
Mohammad Sadoghi, Purdue University
Nesime Tatbul, Intel Labs and MIT
Nicolas Ancaux, INRIA
Niketani Pansare, IBM Almaden
Nikos Mamoulis, The University of Hong Kong
Ninghui Li, Purdue University
Norman May, SAP Research
Oktie Hassanzadeh, IBM Research
Olga Papaemmanouil, Brandeis University
Oliver Kennedy, University of Buffalo
Panagiotis Bouros, Aarhus University
Panagiotis Karras, Aalborg University
Panagiotis Papapetrou, Stockholm University
Pankaj Agarwal, Duke University
Paolo Papotti, Arizona State University
Peter Bailis, MIT and Stanford
Philippe Lamarre, INSA-Lyon France
Pinar Tozun, IBM Research
Quanquan Gu, University of Virginia
Rainer Gemulla, Universität Mannheim
Rajeev Rastogi, Amazon
Ralf Schenkel, University of Trier
Rebecca Taft, MIT
Rene Mueller, IBM Research - Almaden
Renee Miller, University of Toronto
Reynold Cheng, The University of Hong Kong, China
Reza Akbarinia, INRIA
Ryan Johnson, Logic Blox USA
S. Sudarshan, IIT Bombay
Sandeep Tata, Google, USA
Sang Won Lee, Sungkyunkwan University, Korea
Sara Cohen, The Hebrew University of Jerusalem
Selçuk Candan, Arizona State University
Sergey Melnik, Google, USA
Sergio Greco, Univ. of Calabria, Italy
Shady Elbassuoni, AUB

Shahram Ghandeharizadeh, UCS
Shel Finkelstein, University of California at Santa Cruz, USA
Sourav S Bhowmick, Nanyang Technological University
Spyros Blanas, Ohio State University
Srikanta Bedathur, IBM India
Srikanta Tirthapura, Iowa State Univ., USA
Stéphane Bressan, National University of Singapore, Singapore
Steven Whang, Google, USA
Sudipto Das, Microsoft Research
Sungpack Hong, Oracle
Theodore Johnson, AT&T Labs, USA
Thomas Neumann, TU Munich
Tiark Rumpf, Purdue University
Tim Kraska, Brown University
Tingjian Ge, University of Massachusetts at Lowell, USA
Ulf Leser, Humboldt-Universität zu Berlin
Umar Farooq Minhas, Microsoft Research
Venkatesan Chakaravarthy, IBM Research, India
Vijayshankar Raman, IBM Research - Almaden
Viktor Leis, Technische Universität München
Vincent Leroy, University Grenoble-Alps, CNRS, LIG, France
Vineet Chaoji, Amazon
Walid Aref, Purdue Univ., USA
Wang-Chien Lee, Pennsylvania State University, USA
Wei Lu, Renmin University of China
Wei Wang, National University of Singapore
Willis Lang, Microsoft Gray Systems
Wook-Shin Han, Postech
Wynne Hsu, National University of Singapore
Xiaochun Yang, Northeast University
Xiaodong Zhang, Ohio State University
Xiaofang Zhou, University of Queensland
Xuan Liu, Baidu
Xuemin Lin, University of New
Yael Amsterdamer, Tel-aviv Univ., Israel
Yizhou Sun, UCLA
Yongxin Tong, Beihang University, China
Yoshiharu Ishikawa, Nagoya University
Yuanyuan Tian, IBM Almaden
Yufei Tao, Univ. of Queensland
Zhenhui Li, Penn State University
Zhifeng Bao, RMIT University
Ziawasch Abedjan, TU Berlin
Zoi Kaoudi, Qatar Computing Research Institute, Qatar

LETTER FROM THE PROGRAM CHAIRS

The Proceedings of the VLDB Endowment (PVLDB) provides a high-quality publication service to the data management research community. Each volume offers twelve monthly submission deadlines on the first of each month and a quick, six week, reviewing cycle. This publication model was pioneered by PVLDB and combines a journal-style reviewing process, which includes a three month revision cycle, with the agility and visibility provided by rapid on-line publication and presentation at the annual VLDB conference.

This is the third issue of the eleventh volume of the PVLDB.

PVLDB attracts many submissions, and the PVLDB reviewing process is implemented by a large team. The Review Board of PVLDB Volume 11 consists of 201 expert researchers, and reviewing is coordinated by 18 Associate Editors. Board members provide timely (within a 4-week deadline) high-quality reviews, and participate in online discussions with the Associate Editors and the other reviewers assigned to each paper. In order to honor these efforts, this year PVLDB will be recognizing outstanding reviewers. For this purpose, we are gathering feedback from the Associate Editors as well as from the authors, who in a feedback phase are asked to judge the quality of the reviews. These outstanding reviewers will be honored at the 44th VLDB conference, to be held in Rio de Janeiro, Brazil, on August 27-31, 2018. Here also the large majority of the PVLDB Volume 11 papers will be presented.

For the third issue of PVLDB Volume 11, the review board has selected contributions in domains such as learning, graph processing, access control as well as core data management topics such as indexing and caching. We hope that the selected papers will provide insight to the readers and create impact by inspiring additional novel contributions.

Sihem Amer-Yahia and Jian Pei
PVLDB Volume 11 Editors in Chief
VLDB 2018 Program Committee Chairs