

VLDB2019

45th International Conference on Very Large Data Bases, Los Angeles, California



Proceedings of the VLDB Endowment

Volume 12, No. 1 – September 2018

**Proceedings of the 45th International Conference on
Very Large Data Bases, Los Angeles, California**

Program Chairs:

Lei Chen and Fatma Özcan

Associate Editors – Research Track:

Azza Abouzied, Selcuk Candan, Surajit Chaudhuri, Amol Desphande, Johann-Christoph Freytag, Rainer Gemulla, Nick Koudas, Georgia Koutrika, Yunyao Li, Alexandra Meliou, Arnab Nandi, M. Tamer Özsu, Themis Palpanas, Alkis Polyzotis, Kyuseok Shim, Xiaokui Xiao, Meihui Zhang

Proceedings Chairs:

Abdul Quamar, Yongxin Tong

PVLDB – Proceedings of the VLDB Endowment

Volume 12, No. 1, September 2018.

The 45th International Conference on Very Large Data Bases, Los Angeles, California.

Copyright 2018 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 12, Number 1, September 2018: VLDB 2019

Pages i – vi and 1 - 84

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	i
Table of Contents	ii
VLDB 2019 Organization and Review Board	iii

Research Papers

List Intersection for Web Search: Algorithms, Cost Models, and Optimizations	
.....Sunghwan Kim, Taesung Lee, Seungwon Hwang, Sameh Elnikety	1
Interactive Checks for Coordination Avoidance	Michael Whittaker, Joseph M. Hellerstein
	14
Pigeonring: A Principle for Faster Thresholded Similarity Search	Jianbin Qin, Chuan Xiao
	28
Local Algorithms for Hierarchical Dense Subgraph Discovery	
.....Ahmet Erdem Sariyuce, C. Seshadhri, Ali Pinar	43
Cost-Effective Data Annotation using Game-Based Crowdsourcing	
.....Jingru Yang, Ju Fan, Zhewei Wei, Guoliang Li, Tongyu Liu, Xiaoyong Du	57
Optimization for Active Learning-based Interactive Database Exploration.....	
.....Enhui Huang, Liping Peng, Luciano Di Palma, Ahmed Abdelkafi, Anna Liu, Yanlei Diao	71

VLDB 2019 ORGANIZATION AND REVIEW BOARD

General Chairs

Shahram Ghandeharizadeh, USC

Program Chairs and Editors in Chief of PVLDB 12

Lei Chen, HKUST

Fatma Özcan, IBM Research – Almaden

Associate Editors of PVLDB 12

Azza Abouzied, UAE

Selcuk Candan, USA

Surajit Chaudhuri, USA

Amol Desphande, USA

Johann-Christoph Freytag, Germany

Rainer Gemulla, Germany

Nick Koudas, Canada

Georgia Koutrika, Greece

Yunyao Li, USA

Alexandra Meliou, USA

Arnab Nandi, USA

M. Tamer Özsu, University of Waterloo

Themis Palpanas, France

Alkis Polyzotis, USA

Kyuseok Shim, South Korea

Xiaokui Xiao, Singapore

Meihui Zhang, China

VLDB Endowment Representative

Michael Carey, UCI

Sponsorship Committee Chairs

Xiaoyong Du, Renmin University

Volker Markl, TU Berlin

Renee Miller, Northeastern

Publicity Committee Chair

Sumita Barahmand, Microsoft

Jason Yap, Google

Tutorial Chairs

Amr El Abbadi, UCSB

Xin Luna Dong, Amazon

Industrial Chairs

Beng Chin Ooi, NU Singapore

Pat Helland, Salesforce

Wolfgang Lehner, Dresden

Demonstration Chairs

Alin Deutsch, UCSD

Nesime Tatbul, Intel Labs and MIT

Panel Chairs

Sang Kyun Cha, National Seoul University

M. Tamer Özsu, University of Waterloo

Workshop Chairs

Sharad Mehrotra, UCI

Yuanyuan Tian, IBM Research

PhD Workshop Chairs

Ilaria Bartolini, University of Bologna

Feifei Li, University of Utah

Proceedings Chairs

Abdul Quamar, IBM Research

Yongxin Tong, Beihang University

Website Chair

Mehran Barahmand, Amazon

PVLDB Managing Editor

Divesh Srivastava, AT&T Labs-Research

PVLDB Advisory Committee

Peter Boncz, Xin Luna Dong, Juliana Freire, Jayant

Haritsa, Wolfgang Lehner, Renée J. Miller, Tova Milo, M.

Tamer Özsu

Research Track Review Board

Abdul Quamar, IBM Research - Almaden
Ada Waichee Fu, Chinese University of Hong Kong
Ahmet Erdem Sariyuce, University at Buffalo
Alan Fekete, University of Sydney
Alkis Simitsis, Microfocus
Ambuj Singh, UCSB
Andrew Pavlo, CMU
Angela Bonifati, University of Lyon - France
Arijit Khan, Nanyang Technological University
Arnab Bhattacharya, IIT Kanpur
Arun Kumar, UC - San Diego
Arvind Arasu, Microsoft
Ashraf Aboulnaga, QCRI
Ashwin Machanavajjhala, Duke University
Avrilia Floratou, Microsoft
Azade Nazi, Microsoft Research
Badrish Chandramouli, Microsoft Research
Barzan Mozafari, University of Michigan
Beng Chin OOI, NUS - Singapore
Berthold Reinwald, IBM Research - Almaden
Bin Cui, Peiking University - China
Bobbie Cochrane, IBM
Bolin Ding, Alibaba
Boris Glavic, Illinois Institute of Technology
Bugra Gedik, Bilkent University - Turkey
Byron Choi, Hong Kong Baptist University
Carlo Curino, Microsoft Research
Chee-Yong Chan, National University of Singapore (NUS) - Singapore
Chen Li, UC - Irvine
Chengkai Li, UT Arlington
Chuan Lei, IBM Research - Almaden
Cong Yu, Google
Curtis Dyreson, Utah State University
Danica Probic, Oracle
Daniel Kifer, Penn State University
Davide Mottin, Hasso-Plattner Institute
Demetrios Zeinalipour-Yazti, University of Cyprus
Dimitris Papadias, HKUST
Diptikalyan Saha, IBM Research - India
Divyakant Agrawal, UCSB
Donald Kossmann, Microsoft Research
Egemen Tanin, U. Melbourne - Australia
Eser Kandogan, IBM Research - Almaden
Essam M. Mansour, QCRI
Fabio Porto, LNCC - Brazil
Fei Chiang, McMaster University
Feifei Li, University of Utah
Florin Rusu, UC Merced
Floris Geerts, University of Antwerp
George Papadakis, University of Athens
Goetz Graefe, Google
Guoliang Li, Tsinghua University
H. V. Jagadish, University of Michigan
Hakan Ferhatosmanoglu, Bilkent University - Turkey
Hakan Hacigumus, Google
Hanghang Tong, ASU

Helen Huang, The University of Queensland
Heng Tao Shen, UESTC - China
Hong Cheng, Chinese University of Hong Kong
Hongzhi Yin, The University of Queensland
Hua Lu, Aalborg University
Huiping Cao, NMSU
Ilaria Bartolini, University of Bologna
Ilkay Altintas, San Diego Supercomputing Center
Immanuel Trummer, Cornell University
Ioana Manulescu, French Institute for Research in Computer Science and Automation (INRIA)
Ismail Sengor Altingovde, METU - Turkey
James Cheng, Chinese University of Hong Kong
Jens Dittrich, University of Saarland - Germany
Jens Teubner, TU Dortmund
Jianliang Xu, HKBU
Jignesh Patel, University of Wisconsin - Madison
Jinyang Gao, National University of Singapore (NUS) - Singapore
Johann Gamper, Free University of Bolzano - Italy
Jun Yang, Duke University
Junjie Yao, East China Normal University
Kai Zheng, University of Electronic Science and Technology of China
Karthik Sankaranarayanan, IBM Research - India
Katja Hose, Aalborg University
Khuzaima Daudjee, University of Waterloo
Kostas Stefanidis, University of Tampere
Kostas Zoumpatianos, Harvard University
Letizia Tanca, Politecnico di Milano
Lucian Popa, IBM Research - Almaden
Luna Dong, Amazon
Manos Karpathiotakis, EPFL
Maria Luisa Sapino, U. Torino - Italy
Mario Nascimento, U. Alberta - Canada
Martin Theobald, University of Luxemburg
Mary Roth, IBM Research - Almaden
Matthias Boehm, IBM Research - Almaden
Matthias Renz, George Mason University
Maya Ramanath, IIT Delhi
Melanie Herschel, University of Stuttgart - Germany
Michael Böhlen, University of Zurich
Michael Hay, Colgate University
Michael Mathioudakis, University of Helsinki
Min Li, IBM Research - Almaden
Mirek Riedewald, Northeastern University
Mirella Moro, Universidade Federal de Minas Gerais
Mohamed Eltabakh, WPI
Mohamed Mokbel, Qatar Computing Research Institute
Mohamed Sarwat, ASU
Murat Kantarcioglu, University of Texas at Dallas
Nan Tang, QCRI
Nicolas Ancaux, French Institute for Research in Computer Science and Automation (INRIA)
Nikolaus Augsten, University of Salzburg
Oktie Hassanzadeh, IBM Research - Yorktown
Olga Papaemmanouil, Brandeis University
Paolo Papotti, EURECOM - France

Parth Nagarkar, NMSU
Pelin Angin, METU - Turkey
Philip Bernstein, Microsoft Research
Philippe Bonnet, ITU - Copenhagen
Pinar Karagoz, METU - Turkey
Pinar Tozun, ITU - Copenhagen
Raymond Ng, UBC
Sai Wu, Zhejiang University
Sang Kyun Cha, Seoul National University
Sebastian Breß, DFKI - TU Berlin
Semih Salihoglu, University of Waterloo
Senjuti Basu Roy, New Jersey Institute of Technology
Seung-Won Hwang, Yonsei University
Shaoxu Song, Tsinghua University
Shuo Shang, King Abdullah University of Science and Technology
Spyros Blanas, Ohio State University
Stefan Mangeold, CWI Amsterdam
Stefano Paraboschi, Università degli Studi di Bergamo
Steffen Zeuch, DFKI - TU Berlin
Stratis Viglas, University of Edinburgh
Sudip Roy, Google
Tingjian Ge, University of Massachusetts - Lowell
Tyson Condie, Microsoft
Umar Farooq Minhas, Microsoft Research
Vijayshankar Raman, IBM Research - Almaden
Viktor Leis, TU Munich

Vincent Oria, NJIT
Vivek Narasayya, Microsoft Research
Wenjie Zhang, UNSW
Wook-Shin Han, Postech - Korea
Xiang Lian, Kent State University
Xiangmin Zhou, RMIT
Xiaochun Yang, Northeastern University
Xiaofang Zhou, University of Queensland
Li Xiong, Emory University
Xu Chu, Georgia Tech
Xuemin Lin, University of New Southwales
Yael Amsterdamer, Bar-Ilan University
Yannis Velegarakis, University of Trento - Italy
Yanyan Shen, Shanghai Jiao Tong University
Yi Chen, NJIT
Ying Zhang, UTS
Yinghui Wu, Washington State University
Yingjun Wu, IBM Research - Almaden
Yingxia Shao, Peking University
Yongxin Tong, Beihang University
Yoshiharu Ishikawa, Nagoay University
Ye Yuan, NEU - China
Yuanyuan Tian, IBM Research - Almaden
Yucel Saygin, Sabanci Uni. Turkey
Yunjun Gao, Zhejiang University
Zhiguo Gong, University of Macau

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 day 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 first issue of the twelfth volume of the PVLDB. The papers accepted in this volume will be presented in the 45th International Conference on Very Large Data Bases (VLDB 2019), to be held in Los Angeles, California during August 26 to August 30, 2019.

PVLDB attracts many submissions spanning diverse data management topics, and the PVLDB reviewing process is implemented by a large team of dedicated researchers. The Review Board of PVLDB Volume 12 consists of 166 expert researchers, and reviewing is coordinated by 17 Associate Editors. Review Board members provide timely (within a 4-week deadline) high-quality reviews, and participate actively in online discussions led by the Associate Editors for each paper.

For the first issue of PVLDB Volume 12, the review board has selected contributions proposing advances to topics such as Web search, similarity search, and subgraph discovery as well as advanced data management topics such as crowdsourcing, consistent distributed systems, and database exploration. We hope that the selected papers will provide valuable insights to the readers and create impact by inspiring novel systems contributions or follow-up research.

Lei Chen and Fatma Özcan
PVLDB Volume 12 Editors in Chief
VLDB 2019 Program Committee Chairs