Proceedings of the VLDB Endowment

Volume 12, No. 9 – May 2019

Editors in Chief:

Lei Chen and Fatma Özcan

Associate Editors:

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

Publication Editors:

Abdul Quamar, Yongxin Tong

PVLDB – Proceedings of the VLDB Endowment

Volume 12, No. 9, May 2019.

All papers published in this issue will be presented at the 45th International Conference on Very Large Data Bases, Los Angeles, California, 2019.

Copyright 2019 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 9, May 2019

Pages i – vi and 948 - 1057

ISSN 2150-8097

Available at: http://www.pvldb.org and https://dl.acm.org.

TABLE OF CONTENTS

Front Matter

Copyright Notice i
Table of Contents ii
PVLDB Organization and Review Board – Vol. 12 iii

Research Papers

k/2-hop: Fast Mining of Convoy Patterns With Effective Pruning 948
Faisal Orakzai, Toon Calders, Torben Pedersen

Balance-Aware Distributed String Similarity-Based Query Processing System 961
Ji Sun, Zeyuan Shang, Guoliang Li, Zhifeng Bao, Dong Deng

Fine-Grained, Secure and Efficient Data Provenance for Blockchain 975
Pingcheng Ruan, Gang Chen, Anh Dinh, Qian Lin, Beng Chin Ooi, Meihui Zhang

Progressive Top-k Subarray Query Processing in Array Databases 989
Dalsu Choi, Chang-Sup Park, Yon Dohn Chung

Megaphone: Latency-conscious state migration for distributed streaming dataflows 1002
Moritz Hoffmann, Andrea Lattuada, Frank Mcsherry, Vasiliki Kalavri, John Liagouris, Timothy Roscoe

From Anomaly Detection to Rumour Detection using Data Streams of Social Platforms 1016
Thanh Tam Nguyen, Matthias Weidlich, Bolong Zheng, Hongzhi Yin, Quoc Viet Hung Nguyen, Bela Stantic

Obscure: Information-Theoretic Oblivious and Verifiable Aggregation Queries 1030
Peeush Gupta, Yin Li, Sharad Mehrotra, Nisha Panwar, Shantanu Sharma, Sumaya Almanee

Selectivity Estimation for Range Predicates using Lightweight Models 1044
Anshuman Dutt, Chi Wang, Azade Nazi, Srikanth Kandula, Vivek Narasayya, Surajit Chaudhuri

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 12

Editors in Chief of PVLDB
Lei Chen, HKUST
Fatma Özcan, IBM Research - Almaden

Associate Editors of PVLDB
Azza Abouzied, NYU Abu Dhabi
Selcuk Candan, Arizona State University
Surajit Chaudhuri, Microsoft Research
Amol Desphande, University of Maryland
Johann-Christoph Freytag, HU Berlin
Rainer Gemulla, University of Mannheim
Nick Koudas, University of Toronto
Georgia Koutrika, Athena Research Center
Yunyao Li, IBM Research - Almaden
Alexandra Meliou, University of Massachusetts
Arnab Nandi, Ohio State University
M. Tamer Özsu, University of Waterloo
Themis Palpanas, French University Institute
Alkis Polyzotis, Google
Kyuseok Shim, Seoul National University
Xiaokui Xiao, National University of Singapore
Meihui Zhang, Beijing Institute of Technology

Publication Editors
Abdul Quamar, IBM Research - Almaden
Yongxin Tong, Beihang University

PVLDB Managing Editor
Wolfgang Lehner, TU Dresden

PVLDB Advisory Committee
Sihem Amer-Yahia, CNRS
Peter Boncz, Centrum Wiskunde & Informatica
Xin Luna Dong, Amazon
Juliana Freire, New York University
Wolfgang Lehner, TU Dresden
Renée J. Miller, University of Toronto
Tova Milo, Tel Aviv University
M. Tamer Özsu, University of Waterloo
Divesh Srivastava, AT&T Labs-Research

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, Hewlett Packard Labs
Ambuj Singh, UC Santa Barbara
Andrew Pavlo, Carnegie Mellon University
Angela Bonifati, University of Lyon
Arijit Khan, Nanyang Technological University
Arnab Bhattacharya, IIT Kanpur
Arun Kumar, University of California, San Diego
Arvind Arasu, Microsoft Research
Ashraf Aboulnaga, QCRI
Ashwin Machanavajjhala, Duke University
Avrilia Floratou, Microsoft Research
Azade Nazi, Microsoft Research
Badrish Chandramouli, Microsoft Research
Barzan Mozafari, University of Michigan
Beng Chin OOI, National University of Singapore
Berthold Reinwald, IBM Research - Almaden
Bin Cui, Peiking University
Bobbie Cochrane, IBM
Bolin Ding, Alibaba
Boris Glavic, Illinois Institute of Technology
Bugra Gedik, Bilkent University
Byron Choi, Hong Kong Baptist University
Carlo Curino, Microsoft Research
Chee-Yong Chan, National University of Singapore
Chen Li, University of California, 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, UC Santa Barbara
Donald Kossmann, Microsoft Research
Egemen Tanin, University of Melbourne
Eser Kandogan, IBM Research - Almaden
Essam M. Mansour, QCRI
Fabio Porto, LNCC
Fei Chiang, McMaster University
Feifei Li, University of Utah
Florin Rusu, University of California, 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
Hakan Hacigumus, Google
Hanghang Tong, Arizona State University
Helen Huang, University of Queensland
Heng Tao Shen, UESTC
Hong Cheng, Chinese University of Hong Kong
Hongzhi YIn, University of Queensland
Hua Lu, Aalborg University
Huiping Cao, New Mexico State University
Ilaria Bartolini, University of Bologna
Ilkay Altintas, San Diego Supercomputing Center
Immanuel Trummer, Cornell University
Ioana Manulescu, INRIA
Ismail Sengor Altingovde, METU
James Cheng, Chinese University of Hong Kong
Jens Dittrich, University of Saarland
Jens Teubner, TU Dortmund
Jianliang Xu, Hong Kong Baptist University
Jignesh Patel, University of Wisconsin - Madison
Jinyang Gao, National University of Singapore
Johann Gamper, Free University of Bozen-Bolzano
Jun Yang, Duke University
Junjie Yao, East China Normal University
Kai Zheng, UESTC
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, Polytechnic University of Milan
Lucian Popa, IBM Research - Almaden
Luna Dong, Amazon
Manos Karpathiotakis, Facebook London
Maria Luisa Sapino, University of Torino
Mario Nascimento, University of Alberta
Martin Theobald, University of Luxemburg
Mary Roth, IBM Research - Almaden
Matthias Boehm, Graz University of Technology
Matthias Renz, George Masion University
Maya Ramanath, Indian Institute of Technology Delhi
Melanie Herschel, University of Stuttgart
Michael Böhlen , University of Zurich
Michael Hay, Colgate University
Michael Mathioudakis, University of Helsinki
Min Li, JD.com
Mirek Riedewald, Northeastern University
Mirella Moro, Universidade Federal de Minas Gerais
Mohamed Eltabakh, Worcester Polytechnic Institute
Mohamed Mokbel, QCRI
Mohamed Sarwat, Arizona State University
Murat Kantarcioglu, University of Texas at Dallas
Nan Tang, QCRI
Nicolas Anciaux, INRIA
Nikolaus Augsten, University of Salzburg
Oktie Hassanzadeh, IBM Research - Yorktown
Olga Papaemmanouil , Brandeis University
Paolo Papotti, EURECOM
Parth Nagarkar, New Mexico State University
Pelin Angin, Middle East Technical University
Philip Bernstein, Microsoft Research
Philippe Bonnet, IT University of Copenhagen
Pinar Karagoz, Middle East Technical University
Pinar Tozun, IT University of Copenhagen

Raymond Ng, University of British Columbia
Sai Wu, Zhejiang University
Sang Kyun Cha, Seoul National University
Sebastian Breß, DFKI and 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, IIAI
Spyros Blanas, Ohio State University
Stefan Mangeold, Centrum Wiskunde & Informatica
Stefano Paraboschi, University of Bergamo
Steffen Zeuch, DFKI and TU Berlin
Stratis Viglas, University of Edinburgh
Sudip Roy, Google
Tingjian Ge, University of Massachusetts Lowell
Tyson Condie, University of California, Los Angeles
Umar Farooq Minhas, Microsoft Research
Vijayshankar Raman, Google
Viktor Leis, University of Jena
Vincent Oria, New Jersey Institute of Technology
Vivek Narasayya, Microsoft Research
Wenjie Zhang, University of New South Wales
Wook-Shin Han, POSTECH
Xiang Lian, Kent State University
Xiangmin Zhou, RMIT
Xiaochun Yang, Northeastern University
Xiaofang Zhou, University of Queensland
Li Xiong, Emory University
Xu Chu, Georgia Institute of Technology
Xuemin Lin, University of New Southwales
Yael Amsterdamer, Bar-Ilan University
Yannis Velegrakis, Utrecht University
Yanyan Shen, Shanghai Jiao Tong University
Yi Chen, New Jersey Institute of Technology
Ying Zhang, University of Technology Sydney
Yinghui Wu, Washington State University
Yingjun Wu, IBM Research - Almaden
Yingxia Shao, Peking University
Yongxin Tong, Beihang University
Yoshiharu Ishikawa, Nagoay University
Ye Yuan, Northeastern University
Yuanyuan Tian, IBM Research - Almaden
Yucel Saygin, Sabancı University
Yunjun Gao, Zhejiang University
Zhiguo Gong, University of Macau

LETTER FROM THE EDITORS IN CHIEF

The Proceedings of the VLDB Endowment (PVLDB) provides a high-quality journal 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 weeks, 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.

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.

This is the ninth issue of the twelfth volume of the PVLDB. There are eight papers accepted in this issue that will be presented at 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.

For the ninth issue of PVLDB Volume 12, the review board has selected contributions proposing advances to topics on data management, such as selectivity estimation for range query, progressive top-k subarray query processing, data migration mechanism for distributed data flow, privacy-preserved aggregation query with multi-party computation and distributed string similarity search. This issue also contains papers on advanced data analytics and mining, including fast mining convoy patterns, rumor detection and data provenance for Blockchain. We hope that the readers will find the selected papers engaging, and thought provoking. We also hope that the selected papers will provide valuable insights and inspire novel systems contributions and follow-up research.

Lei Chen and Fatma Özcan

PVLDB Volume 12 Editors in Chief