Proceedings of the VLDB Endowment

Volume 13, No. 4 – December 2019

Editors in Chief:

Magdalena Balazinska and Xiaofang Zhou

Associate Editors:

Azza Abouzied, Amr El Abbadi, Phil Bernstein, Xin Luna Dong, Zi (Helen) Huang, Nick Koudas, Georgia Koutrika, Guoliang Li, Alexandra Meliou, Felix Naumann, Dan Olteanu, M. Tamer Özsu, Aditya Parameswaran, Andy Pavlo, Xiaokui Xiao, Jeffrey Xu Yu, Meihui Zhang, Jingren Zhou

Publication Editors:

Hiroaki Shiokawa and Sen Wang

PVLDB – Proceedings of the VLDB Endowment

Volume 13, No. 4, December 2019.

All papers published in this issue will be presented at the 46th International Conference on Very Large Data Bases, Tokyo, Japan, 2020.

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 13, Number 4, December 2019

Pages i – vi and 421 - 587

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. 13 iii

Research Papers

DPTree: Differential Indexing for Persistent Memory 421
Xinjing Zhou, Lidan Shou, Ke Chen, Wei Hu, Gang Chen

AJoin: Ad-hoc Stream Joins at Scale 435
Suraj Shetiya, Abolfazl Asudeh, Sadia Ahmed, Gautam Das

On Performance Stability in LSM-based Storage Systems 449
Chen Luo, Michael Carey

Hop-constrained s-t Simple Path Enumeration: Towards Bridging Theory and Practice 463
You Peng, Ying Zhang, Xuemin Lin, Wenjie Zhang, Lu Qin, Jingren Zhou

Panorama: A Data System for Unbounded Vocabulary Querying over Video 477
Yuhao Zhang, Arun Kumar

Planting Trees for scalable and efficient Canonical Hub Labeling 492
Kartik Lakhotia, Rajgopal Kannan, Qing Dong, Viktor K Prasanna

Operationalizing Individual Fairness with Pairwise Fair Representations 506
Preethi Lahoti, Krishna Gummadi, Gerhard Weikum

Optimizing Databases by Learning Hidden Parameters of Solid State Drives 519
Aarati Kakaraparthy, Jignesh Patel, Kwanghyun Park, Brian Kroth

BlazeIt: Optimizing Declarative Aggregation and Limit Queries for Neural Network-Based Video Analytics 533
Daniel Kang, Peter D Bailis, Matei Zaharia

Joins on Samples: A Theoretical Guide for Practitioners 547
Dawei Huang, Dong Young Yoon, Seth Pettie, Barzan Mozafari

Mining an "Anti-Knowledge Base" from Wikipedia Updates with Applications to Fact Checking and Beyond 561
Georgios Karagiannis, Immanuel Trummer, Saehan Jo, Shubham Khandelwal, Xuezhi Wang, Cong Yu

Evaluating Persistent Memory Range Indexes 574
Lucas Lersch, Xiangpeng Hao, Ismail Oukid, Tianzheng Wang, Thomas Willhalm

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 13

Editors in Chief of PVLDB
Magdalena Balazinska (University of Washington)
Xiaofang Zhou (University of Queensland)

Associate Editors of PVLDB
Azza Abouzied (New York University Abu Dhabi)
Amr El Abbadi (University of California, Santa Barbara)
Phil Bernstein (Microsoft Research)
Xin Luna Dong (Amazon)
Zi (Helen) Huang (University of Queensland)
Nick Koudas (University of Toronto)
Georgia Koutrika (Athena Research Center)
Guoliang Li (Tsinghua University)
Alexandra Meliou (University of Massachusetts, Amherst)
Felix Naumann (Hasso Plattner Institute, University of Potsdam)
Dan Olteanu (University of Oxford)
M. Tamer Özsu (University of Waterloo)
Aditya Parameswaran (University of California, Berkeley)
Andy Pavlo (Carnegie Mellon University)
Xiaokui Xiao (National University of Singapore)
Jeffrey Xu Yu (The Chinese University of Hong Kong)
Meihui Zhang (Beijing Institute of Technology)
Jingren Zhou (Alibaba Group)

Publication Editors
Hiroaki Shiokawa (University of Tsukuba)
Sen Wang (University of Queensland)

PVLDB Managing Editor
Wolfgang Lehner (TU Dresden, Germany)

PVLDB Advisory Committee
Divesh Srivastava (AT&T Labs-Research)
M. Tamer Özsu (University of Waterloo)
Juliana Freire (New York University)
Xin Luna Dong (Amazon)
Peter Boncz (CWI)
Xiaofang Zhou (University of Queensland)
Magdalena Balazinska (University of Washington)
Lei Chen (Hong Kong University of Science and Technology)
Fatma Ozcan (IBM Almaden)
Graham Cormode (University of Warwick)
Felix Naumann (HPI, Germany)

Review Board

Ziawasch Abedjan (TU Berlin)
Ashraf Aboulnaga (Qatar Computing Research Institute)
Pelin Angın (Middle East Technical University)
Arvind Arasu (Microsoft Research)
Joy Arulraj (Georgia Tech)
Manos Athanassoulis (Boston University)
Zhifeng Bao (RMIT University)
Ilaria Bartolini (University of Bologna)
Leilani Battle (University of Maryland)
Kaustubh Beedkar (TU Berlin)
Arnab Bhattacharya (IIT Kanpur)
Sourav S Bhowmick (Nanyang Technological University)
Carsten Binnig (TU Darmstadt)
Spyros Blanas (The Ohio State University)
Matthias Boehm (Graz University of Technology)
Alexander Böhm (SAP SE)
Michael Böhlen (University of Zürich)
Peter Boncz (Centrum Wiskunde & Informatica)
Angela Bonifati (Lyon 1 University)
Philippe Bonnet (IT University of Copenhagen)
Renata Borovica-Gajic (University of Melbourne)
Huiping Cao (New Mexico State University)
Lei Cao (MIT)
Lijun Chang (The University of Sydney)
Surajit Chaudhuri (Microsoft Research)
Lei Chen (Hong Kong University of Science and Technology)
Hong Cheng (The Chinese University of Hong Kong)
Reynold Cheng (The University of Hong Kong)
Fei Chiang (McMaster University)
Xu Chu (Georgia Tech)
Bobbie Cochrane (IBM)
Gao Cong (Nanyang Technological University)
Brian Cooper (Google)
Natacha Crooks (University of Texas at Austin)
Andrew Crotty (Brown University)
Bin Cui (Peking University)
Sudipto Das (Microsoft Research)
Akash Das Sarma (Facebook)
Khuzaima Daudjee (University of Waterloo)
Niv Dayan (Harvard University)
Dong Deng (Rutgers University)
Bailu Ding (Microsoft Research)
Bolin Ding (Alibaba Group)
Jens Dittrich (Saarland University)
Harish Doraiswamy (New York University)
Eduard C. Dragut (Temple University)
Curtis Dyreson (Utah State University)
Mohamed Y. Eltabakh (Teradata Labs)
Jose M. Faleiro (University of California, Berkeley)
Ju Fan (Renmin University of China)
Raul Castro Fernandez (MIT)
Avrilia Floratou (Microsoft Research)
Avigdor Gal (Technion)
Alex Galakatos (Brown University)
Johann Gamper (Free University of Bozen-Bolzano)
Jing Gao (University at Buffalo)
Yunjun Gao (Zhejiang University)
Tingjian Ge (University of Massachusetts, Lowell)
Floris Geerts (University of Antwerp)
Johannes Gehrke (Microsoft Research)
Jonathan Goldstein (Microsoft Research)
Torsten Grust (University of Tübingen)
Wook-Shin Han (POSTECH)
Takahiro Hara (Osaka University)
Oktie Hassanzadeh (IBM Research)
Michael Hay (Colgate University)
Xi He (University of Waterloo)
Melanie Herschel (University of Stuttgart)
Katja Hose (Aalborg University)
Wen Hua (The University of Queensland)
Xin Huang (Hong Kong Baptist University)
Yan Huang (University of North Texas)
Seung-won Hwang (Yonsei University)
Christopher Jermaine (Rice University)
Ruoming Jin (Kent State University)
Eser Kandogan (IBM Research - Almaden)
Murat Kantarcioglu (University of Texas at Dallas)
Verena Kantere (University of Ottawa)
Pinar Karagoz (Middle East Technical University)
Manos Karpathiotakis (Facebook)
Batya Kenig (University of Washington)
Oliver Kennedy (University at Buffalo)
Arijit Khan (Nanyang Technological University)
Daniel Kifer (Pennsylvania State University)
Hideaki Kimura (Oracle)
Sanjay Krishnan (University of Chicago)
Arun Kumar (University of California, San Diego)
Chuan Lei (IBM Research - Almaden)
Viktor Leis (Technical University of Munich)
Ulf Leser (Humboldt-Universität zu Berlin)
Chengkai Li (The University of Texas at Arlington)
Feifei Li (University of Utah)
Rong-Hua Li (Beijing Institute of Technology)
Sebastian Link (The University of Auckland)
Chengfei Liu (Swinburne University of Technology)
Hua Lu (Aalborg University)
Jiaheng Lu (University of Helsinki)
Wei Lu (Renmin University of China)
Shuai Ma (Beihang University)
Nikos Mamoulis (University of Ioannina)
Ioana Manolescu (INRIA)
Essam Mansour (Concordia University, Canada)
Ryan Marcus (Brandeis University)
Sergey Melnik (Google)
Mohamed Mokbel (Qatar Computing Research Institute)
Mirella Moura Moro (Universidade Federal de Minas Gerais)
Davide Mottin (Aarhus University)
Parth Nagarkar (New Mexico State University)
Faisal Nawab (University of California, Santa Cruz)
Thomas Neumann (Technical University of Munich)
Milos Nikolic (The University of Edinburgh)

Beng Chin Ooi (National University of Singapore)
Ismail Oukid (SAP SE)
Mourad Ouzzani (Qatar Computing Research Institute)
Themis Palpanas (Paris Descartes University)
George Papadakis (University of Athens)
Olga Papaemmanouil (Brandeis University)
Thorsten Papenbrock (Hasso Plattner Institute)
Paolo Papotti (EURECOM)
Stefano Paraboschi (Universita' degli Studi di Bergamo)
Yongjoo Park (University of Michigan)
Jignesh M. Patel (University of Wisconsin-Madison)
Peter Pietzuch (Imperial College London)
Holger Pirk (Imperial College London)
Fábio Porto (National Laboratory for Scientific Computing (LNCC), Brazil)
Dan R. K. Ports (Microsoft Research)
Lu Qin (University of Technology Sydney)
Abdul H. Quamar (IBM Research - Almaden)
Tilmann Rabl (TU Berlin)
Karthik Ramachandra (Microsoft Research)
Maya Ramanath (IIT Delhi)
Berthold Reinwald (IBM Research)
Theodoros Rekatsinas (University of Wisconsin-Madison)
Uwe Roehm (The University of Sydney)
Jennie Rogers (Northwestern University)
Florin Rusu (University of California, Merced)
Diptikalyan Saha (IBM Research AI India)
Ken Salem (University of Waterloo)
Semih Salihoglu (University of Waterloo)
Maria Luisa Sapino (University of Torino)
A. Erdem Sariyuce (University at Buffalo)
Mohamed Sarwat (Arizona State University)
Maximilian Schleich (University of Oxford)
Mohamed Sharaf (University of Queensland)
Yanyan Shen (Shanghai Jiao Tong University)
Kyuseok Shim (Seoul National University)
Prashant Shiralkar (Amazon)
Alkis Simitsis (Hewlett Packard Enterprise)
Kostas Stefanidis (Tampere University)
Rebecca Taft (Cockroach Labs)
Nan Tang (Qatar Computing Research Institute)
Yufei Tao (The Chinese University of Hong Kong)
Jens Teubner (TU Dortmund)
Andreas Thor (University of Applied Sciences for Telecommunications Leipzig)
Yongxin Tong (Beihang University)
Anthony K. H. Tung (National University of Singapore)
Yannis Velegrakis (Utrecht University)
Stratis Viglas (University of Edinburgh)
Daisy Zhe Wang (University of Florida)
Guoren Wang (Beijing Institute of Technology)
Jiannan Wang (Simon Fraser University)
Junhu Wang (Griffith University)
Sibo Wang (The Chinese University of Hong Kong)
Eugene Wu (Columbia University)
Yingjun Wu (IBM Research - Almaden)
Yinglong Xia (Huawei Research America)
Chuan Xiao (Osaka University)
Yanghua Xiao (Fudan University)
Li Xiong (Emory University)
Jianliang Xu (Hong Kong Baptist University)
Xiaochun Yang (Northeastern University, China)
Junjie Yao (East China Normal University)
Hongzhi Yin (The University of Queensland)
Man Lung Yiu (Hong Kong Polytechnic University)
Haruo Yokota (Tokyo Institute of Technology)
Masatoshi Yoshikawa (Kyoto University)
Xiangyao Yu (MIT)
Demetrios Zeinalipour-Yazti (University of Cyprus)
Baihua Zheng (Singapore Management University)
Rui Zhang (University of Melbourne)
Wenjie Zhang (University of New South Wales)
Xiaofei Zhang (The University of Memphis)
Ying Zhang (University of Technology Sydney)
Yuanyuan Zhu (Wuhan University)
Lei Zou (Peking University)
Kostas Zoumpatianos (Harvard University)

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 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.

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 13 consists of 186 expert researchers, and reviewing is coordinated by 18 Associate Editors. Review Board members provide timely (within a 3-week deadline) high-quality reviews, and participate actively in online discussions led by the Associate Editors for each paper. When needed, the Associate Editors together with the Editors-in-Chief solicit additional reviews from external experts. We give special thanks here to those additional reviewers who in most cases need to complete their expert reviews on a very short notice.

Most of the accepted papers go through a revision process that requires a second round of reviews after the authors have addressed an initial set of issues and concerns raised by the reviewers during the first round. Some papers are further accepted with shepherding, which means that one of the reviewers works with the authors to address a final set of comments.

This is the fourth issue of the thirteenth volume of PVLDB. There are twelve accepted papers in this issue. These papers will be presented at the 46th International Conference on Very Large Data Bases (VLDB 2020), to be held in Tokyo, Japan during August 31 to September 4, 2020.

All these twelve papers have gone through the revision step following a rigorous reviewing process before they are accepted. There are four papers which cover database indexing and optimization techniques for new hardware and computing platforms, including two papers on persistent memory. New database join algorithms in the context of stream and sample-based query processing are presented in two separate papers. Graph databases remain to be a hot topic, with two papers on this topic included in this issue. We start to see more research work on developing new database technologies that can support video query processing using neural networks, as discussed in two papers here. Another two very interesting papers discuss learning individual fairness measures from data and automatic mining of factual mistakes from Wikipedia respectively.

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.

Magdalena Balazinska and Xiaofang Zhou

PVLDB Volume 13 Editors in Chief