



Proceedings of the VLDB Endowment

Volume 13, No. 5 – January 2020

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. 5, January 2020.

PVLDB is indexed in Scopus (Elsevier) as well as covered by the following Clarivate Analytics services:

- Science Citation Index Expanded (also known as SciSearch®)
- Journal Citation Reports/Science Edition, and
- Current Contents®/Engineering Computing and Technology

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

Copyright 2020 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 5, January 2020
Pages i – vii and 588 - 782
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	iv

Research Papers

M.B.R.O.S.I.A: Providing Performant Virtual Resiliency for Distributed Applications	588
<i>Jonathan Goldstein, Ahmed S. Abdelhamid, Mike Barnett, Sebastian Burckhardt, Badrish Chandramouli, Darren Gehring, Niel I Lebeck, Christopher Meiklejohn, Umar Farooq Minhas, Ryan Newton, Rahee Peshawaria, Tal Zaccai, Irene Zhang</i>	
Efficient Shortest Path Index Maintenance on Dynamic Road Networks with Theoretical Guarantees	602
<i>Dian Ouyang, Long Yuan, Lu Qin, Lijun Chang, Ying Zhang, Xuemin Lin</i>	
ParPaRaw: Massively Parallel Parsing of Delimiter-Separated Raw Data	616
<i>Elias Stehle, Hans-Arno Jacobsen</i>	
Opportunities for Optimism in Contended Main-Memory Multicore Transactions.....	629
<i>Yihe Huang, William L Qian, Eddie Kohler, Barbara Liskov, Liuba Shrira</i>	
PM-LSH: A Fast and Accurate LSH Framework for High-Dimensional Approximate NN Search	643
<i>Bolong Zheng, Xi Zhao, Lianggui Weng, Hung Nguyen Quoc Viet, Hang Liu, Christian S Jensen</i>	
Hunting multiple bumps in graphs.....	656
<i>Yahui Sun, Jun Luo, Theodoros Lappas, Xiaokui Xiao, Bin Cui</i>	
Homogeneous Network Embedding for Massive Graphs via Reweighted Personalized PageRank .	670
<i>Renchi Yang, Jieming Shi, Xiaokui Xiao, Yin Yang, Sourav S Bhowmick</i>	
Pattern Functional Dependencies for Data Cleaning	684
<i>Abdulhakim Qahtan, Nan Tang, Mourad Ouzzani, Yang Cao, Michael Stonebraker</i>	
MEGA: Multi-View Semi-Supervised Clustering of Hypergraphs	698
<i>Joyce J Whang, Rundong Du, Sangwon Jung, Geon Lee, Barry Drake, Qingqing Liu, Seonggoo Kang, Haesun Park</i>	
MDedup: Duplicate Detection with Matching Dependencies.....	712
<i>Ioannis Koumarelas, Thorsten Papenbrock, Felix Naumann</i>	
Programmable View Update Strategies on Relations	726
<i>Van-Dang Tran, Hiroyuki Kato, Zhenjiang Hu</i>	
Amber: A Debuggable Dataflow System Based on the Actor Model	740
<i>Avinash Kumar, Zuozhi Wang, Shengquan Ni, Chen Li</i>	
Dynamic Speculative Optimizations for SQL Compilation in Apache Spark.....	754
<i>Filippo Schiavio, Daniele Bonetta, Walter Binder</i>	

Mind the Gap: An Experimental Evaluation of Imputation of Missing Values Techniques in Time Series 768
Mourad Khayati, Alberto Lerner, Zakhar Tymchenko, Philippe Cudre-Mauroux

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 13

Editors in Chief of PVLDB

Magdalena Balazinska (University of Washington, USA)
Xiaofang Zhou (University of Queensland, Australia)

Associate Editors of PVLDB

Azza Abouzied (New York University Abu Dhabi, UAE)
Amr El Abbadi (University of California, Santa Barbara, USA)
Phil Bernstein (Microsoft Research, USA)
Xin Luna Dong (Amazon, USA)
Zi (Helen) Huang (University of Queensland, Australia)
Nick Koudas (University of Toronto, Canada)
Georgia Koutrika (Athena Research Center, Greece)
Guoliang Li (Tsinghua University, China)
Alexandra Meliou (University of Massachusetts, Amherst, USA)
Felix Naumann (Hasso Plattner Institute, University of Potsdam, Germany)
Dan Olteanu (University of Oxford, United Kingdom)
M. Tamer Özsu (University of Waterloo, Canada)
Aditya Parameswaran (University of California, Berkeley, USA)
Andy Pavlo (Carnegie Mellon University, USA)
Xiaokui Xiao (National University of Singapore, Singapore)
Jeffrey Xu Yu (The Chinese University of Hong Kong, China)

Meihui Zhang (Beijing Institute of Technology, China)
Jingren Zhou (Alibaba Group, China)

Publication Editors

Hiroaki Shiokawa (University of Tsukuba, Japan)
Sen Wang (University of Queensland, Australia)

PVLDB Managing Editor

Wolfgang Lehner (TU Dresden, Germany)

PVLDB Advisory Committee

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

Review Board

Ziawasch Abedjan (TU Berlin, Germany)
Ashraf Aboulnaga (Qatar Computing Research Institute, Qatar)
Pelin Angin (Middle East Technical University, Turkey)
Arvind Arasu (Microsoft Research, USA)
Joy Arulraj (Georgia Tech, USA)
Manos Athanassoulis (Boston University, USA)
Zhifeng Bao (RMIT University, Australia)
Ilenia Bartolini (University of Bologna, Italy)
Leilani Battle (University of Maryland, USA)
Kaustubh Beedkar (TU Berlin, Germany)
Arnab Bhattacharya (IIT Kanpur, India)
Sourav S Bhowmick (Nanyang Technological University, Singapore)
Carsten Binnig (TU Darmstadt, Germany)
Spyros Blanas (The Ohio State University, USA)
Matthias Boehm (Graz University of Technology, Austria)
Alexander Böhm (SAP SE, Germany)
Michael Böhlen (University of Zürich, Switzerland)
Peter Boncz (Centrum Wiskunde & Informatica, Netherlands)
Angela Bonifati (Lyon 1 University, France)
Philippe Bonnet (IT University of Copenhagen, Denmark)
Renata Borovica-Gajic (University of Melbourne, Australia)
Huiping Cao (New Mexico State University, USA)
Lei Cao (MIT, USA)
Lijun Chang (The University of Sydney, Australia)
Surajit Chaudhuri (Microsoft Research, USA)
Lei Chen (Hong Kong University of Science and Technology, China)
Hong Cheng (The Chinese University of Hong Kong, China)
Reynold Cheng (The University of Hong Kong, China)
Fei Chiang (McMaster University, Canada)
Xu Chu (Georgia Tech, USA)
Bobbie Cochrane (IBM, USA)
Gao Cong (Nanyang Technological University, Singapore)
Brian Cooper (Google, USA)
Natacha Crooks (University of Texas at Austin, USA)
Andrew Crotty (Brown University, USA)
Bin Cui (Peking University, China)
Sudipto Das (Amazon Web Service, USA)
Akash Das Sarma (Facebook, USA)
Khuzaima Daudjee (University of Waterloo, Canada)
Niv Dayan (Harvard University, USA)
Dong Deng (Rutgers University, USA)
Bailu Ding (Microsoft Research, USA)
Bolin Ding (Alibaba Group, China)
Jens Dittrich (Saarland University, Germany)
Harish Doraiswamy (New York University, USA)
Eduard C. Dragut (Temple University, USA)
Curtis Dyreson (Utah State University, USA)
Mohamed Y. Eltabakh (Teradata Labs, USA)

Jose M. Faleiro (Microsoft Research, USA)
Ju Fan (Renmin University of China, China)
Raul Castro Fernandez (The University of Chicago, USA)
Avrilia Floratou (Microsoft Research, USA)
Avigdor Gal (Technion, Israel)
Alex Galakatos (Brown University, USA)
Johann Gamper (Free University of Bozen-Bolzano, Italy)
Jing Gao (University at Buffalo, USA)
Yunjun Gao (Zhejiang University, China)
Tingjian Ge (University of Massachusetts, Lowell, USA)
Floris Geerts (University of Antwerp, Belgium)
Johannes Gehrke (Microsoft Research, USA)
Jonathan Goldstein (Microsoft Research, USA)
Torsten Grust (University of Tübingen, Germany)
Wook-Shin Han (POSTECH, South Korea)
Takahiro Hara (Osaka University, Japan)
Oktie Hassanzadeh (IBM Research, USA)
Michael Hay (Colgate University, USA)
Xi He (University of Waterloo, Canada)
Melanie Herschel (University of Stuttgart, Germany)
Katja Hose (Aalborg University, Denmark)
Wen Hua (The University of Queensland, Australia)
Xin Huang (Hong Kong Baptist University, China)
Yan Huang (University of North Texas, USA)
Seung-won Hwang (Yonsei University, South Korea)
Christopher Jermaine (Rice University, USA)
Ruoming Jin (Kent State University, USA)
Eser Kandogan (Megagon Labs, USA)
Murat Kantarcioglu (University of Texas at Dallas, USA)
Verena Kantere (University of Ottawa, Canada)
Pinar Karagoz (Middle East Technical University, Turkey)
Manos Karpathiotakis (Facebook, United Kingdom)
Batya Kenig (University of Washington, USA)
Oliver Kennedy (University at Buffalo, USA)
Arijit Khan (Nanyang Technological University, Singapore)
Daniel Kifer (Pennsylvania State University, USA)
Hideaki Kimura (Oracle, USA)
Sanjay Krishnan (University of Chicago, USA)
Arun Kumar (University of California, San Diego, USA)
Chuan Lei (IBM Research - Almaden, USA)
Viktor Leis (Technical University of Munich, Germany)
Ulf Leser (Humboldt-Universität zu Berlin, Germany)
Chengkai Li (The University of Texas at Arlington, USA)
Feifei Li (University of Utah, USA)
Rong-Hua Li (Beijing Institute of Technology, China)
Sebastian Link (The University of Auckland, New Zealand)
Chengfei Liu (Swinburne University of Technology, Australia)
Hua Lu (Aalborg University, Denmark)
Jiaheng Lu (University of Helsinki, Finland)
Wei Lu (Renmin University of China, China)
Shuai Ma (Beihang University, China)
Nikos Mamoulis (University of Ioannina, Greece)
Ioana Manolescu (INRIA, France)
Essam Mansour (Concordia University, Canada)

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

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 fifth issue of the thirteenth volume of PVLDB. There are fourteen 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.

The papers in this issue truly exemplify the breadth of the data management field. Several papers shine a new light on an old problem, including view updates, query compilation, concurrency control, data parsing, data cleaning, and nearest neighbour search. Other papers tackle more recent problems in particular graph processing, with one paper focusing on road networks. Finally, one paper applies failure handling techniques from data processing to distributed application programming. Such breadth of technical problems as well as the breadth of solutions spanning the hardware-software stack and going from theoretical to systems-oriented are an important quality of the database community and its technical contributions.

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