Proceedings of the VLDB Endowment

Volume 16, No. 1 – September 2022

Editors in Chief:

Georgia Koutrika and Jun Yang

Associate Editors:

Alkis Simitsis, Amol Deshpande, Angela Bonifati, Ashwin Machanavajjhala, Badrish Chandramouli, Boris Glavic, Ce Zhang, Cyrus Shahabi, Dan Olteanu, Eric Lo, Evaggelia Pitoura, Evimaria Terzi, Gustavo Alonso, Helen (Zi) Huang, Hong Cheng, Kenneth Ross, Khuzaima Daudjee, Kyuseok Shim, Letizia Tanca, Lucian Popa, Magdalena Balazinska, Meihui Zhang, Neoklis Polyzotis, Nesime Tatbul, Nikos Mamoulis, Rachel Pottinger, Wenjie Zhang, Wolfgang Gatterbauer, Wook-Shin Han, Xiaokui Xiao, Yannis Velegrakis, Yanyan Shen, Yi Chen, Yongxin Tong, Zhifeng Bao

Publication Editors:

Manos Athanassoulis, Kostas Stefanidis, Ju Fan

PVLDB – Proceedings of the VLDB Endowment

Volume 16, No. 1, September 2022.

All papers published in this issue will be presented at the 49th International Conference on Very Large Data Bases, Vancouver, Canada, 2023.

Copyright 2022 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 16, Number 1, September 2022

Pages i – vi and 1 - 126

ISSN 2150-8097

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

TABLE OF CONTENTS

Front Matter

Copyright Notice i
Table of Contents ii
PVLDB Organization and Review Board – Vol. 16 iii

Research Papers

C5: Cloned Concurrency Control That Always Keeps Up 1
Jeffrey Helt, Abhinav Sharma, Daniel J Abadi, Wyatt Lloyd, Jose Faleiro

The Case for Distributed Shared-Memory Databases with RDMA-Enabled Memory Disaggregation 15
Ruihong Wang, Jianguo Wang, Stratos Idreos, Tamer Özsu, Walid G Aref

FlexChain: An Elastic Disaggregated Blockchain 23
Chenyuan Wu, Mohammad Javad Amiri, Jared Asch, Heena Nagda, Qizhen Zhang, Boon Thau Loo

MiCS: Near-linear Scaling for Training Gigantic Model on Public Cloud 37
Zhen Zhang, Shuai Zheng, Yida Wang, Justin Chiu, George Karypis, Trishul A Chilimbi, Mu Li, Xin Jin

Privacy-preserving Cooperative Online Matching over Spatial Crowdsourcing Platforms 51
Yi Yang, Yurong Cheng, Ye Yuan, Guoren Wang, Lei Chen, Yongjiao Sun

Coresets over Multiple Tables for Feature-rich and Data-efficient Machine Learning 64
Jiayi Wang, Chengliang Chai, Nan Tang, Jiabin Liu, Guoliang Li

STARRY: Multi-master Transaction Processing on Semi-leader Architecture 77
Zihao Zhang, Huiqi Hu, Xuan Zhou, Jiang Wang

SIFTER: Space-Efficient Value Iteration for Finite-Horizon MDPs 90
Konstantinos Skitsas, Ioannis G Papageorgiou, Mohammad Sadegh Talebi, Vasiliki Kantere, Michael Katehakis, Panagiotis Karras

TreeLine: An Update-In-Place Key-Value Store for Modern Storage 99
Geoffrey Yu, Markos Markakis, Andreas Kipf, Per-ake Larson, Umar Farooq Minhas, Tim Kraska

DPXPlain: Privately Explaining Aggregate Query Answers 113
Yuchao Tao, Amir Gilad, Ashwin Machanavajjhala, Sudeepa Roy

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 16

Editors in Chief of PVLDB
Georgia Koutrika (Athena Research Center)
Jun Yang (Duke University)

Associate Editors of PVLDB
Alkis Simitsis (Athena Research Center)
Amol Deshpande (University of Maryland at College Park)
Angela Bonifati (Lyon 1 University)
Ashwin Machanavajjhala (Duke University/Tumult Labs)
Badrish Chandramouli (Microsoft Research)
Boris Glavic (Illinois Institute of Technology)
Ce Zhang (ETH Zurich)
Cyrus Shahabi (University of Southern California)
Dan Olteanu (University of Zurich)
Eric Lo (The Chinese University of Hong Kong)
Evaggelia Pitoura (University of Ioannina)
Evimaria Terzi (Boston University)
Gustavo Alonso (ETH Zurich)
Helen (Zi) Huang (University of Queensland)
Hong Cheng (The Chinese University of Hong Kong)
Kenneth Ross (Columbia University)
Khuzaima Daudjee (University of Waterloo)
Kyuseok Shim (Seoul National University)
Letizia Tanca (Politecnico di Milano)
Lucian Popa (IBM Research - Almaden)
Magdalena Balazinska (University of Washington)
Meihui Zhang (Beijing Institute of Technology)
Neoklis Polyzotis (Databricks)
Nesime Tatbul (Intel Labs and MIT)
Nikos Mamoulis (University of Ioannina)
Rachel Pottinger (University of British Columbia)
Wenjie Zhang (University of New South Wales)
Wolfgang Gatterbauer (Northeastern University)
Wook-Shin Han (Pohang University of Science and Technology)
Xiaokui Xiao (National University of Singapore)
Yannis Velegrakis (University of Trento and Utrecht University)
Yanyan Shen (Shanghai Jiao Tong University)
Yi Chen (New Jersey Institute of Technology)
Yongxin Tong (Beihang University)
Zhifeng Bao (RMIT University)

Publication Editors
Manos Athanassoulis (Boston University)
Kostas Stefanidis (Tampere University)
Ju Fan (Renmin University of China)

PVLDB Managing Editor
Wolfgang Lehner (Dresden University of Technology)

PVLDB Advisory Board
Vanessa Braganholo (Universidade Federal Fluminense)
Sourav S Bhowmick (Nanyang Technological University)
Chris Jermaine (Rice University)
Peter Triantafillou (University of Warwick)
Xin Luna Dong (Facebook)
Fatma Ozcan (Google)
Lei Chen (Hong Kong University of S&T)
Juliana Freire (New York University)
Graham Cormode (University of Warwick)
Divesh Srivastava (AT&T Labs-Research)
Wolfgang Lehner (Dresden University of Technology)
Felix Naumann (HPI)
Xuemin Lin (University of New South Wales)
Georgia Koutrika (Athena Research Center)
Jun Yang (Duke University)

Review Board

Abolfazl Asudeh (University of Illinois at Chicago)
Alexander J Ratner (University of Washington)
Alexandra Meliou (University of Massachusetts Amherst)
Amelie Marian (Rutgers University)
Amir Gilad (Duke University)
Amir Shaikhha (University of Edinburgh)
Amrita Roy Chowdhury (University of Wisconsin-Madison)
Anastasios Kementsietsidis (Google Research)
Andrew Crotty (Carnegie Mellon University)
Anna Fariha (Microsoft)
Anton Dignös (Free University of Bozen-Bolzano)
Antonios Deligiannakis (Technical University of Crete)
Arijit Khan (Nanyang Technological University)
Ashraf Aboulnaga (Qatar Computing Research Institute, HBKU)
Asterios Katsifodimos (Delft University of Technology)
Baihua Zheng (Singapore Management University)
Bin Cui (Peking University)
Bingsheng He (National University of Singapore)
Binhang Yuan (ETH Zurich)
Bojan Karlas (ETH Zurich)
Bolin Ding (Data Analytics and Intelligence Lab, Alibaba Group)
Bolong Zheng (Huazhong University of Science and Technology)
Bongki Moon (Seoul National University)
Botong Huang (Alibaba)
Brad Glasbergen (University of Waterloo)
Brandon Haynes (Microsoft Gray Systems Lab)
Cedric Renggli (ETH Zurich)
Chao Zhang (Lyon 1 University)
Chen Li (UC Irvine)
Chengfei Liu (Swinburne University of Technology)
Chengkai Li (The University of Texas at Arlington)
Chengliang Chai (Tsinghua University)
Chong Wang (Amazon)
Cristian Riveros (PUC Chile)
Daichi Amagata (Osaka University)
Dan Kifer (Pennsylvania State University)
Daniel Kang (Stanford University)
Diego Calvanese (Free University of Bozen)
Dimitrios Skoutas (Athena Research Center)
Dimitris Sacharidis (ULB)
Dirk Habich (TU Dresden)
Dong Deng (Rutgers University - New Brunswick)
Dong Wen (University of New South Wales)
Dong Xie (Pennsylvania State University)
Dongxiang Zhang (Zhejiang University)
Dumitrel Loghin (National University of Singapore)
Elena Ferrari (University of Insubria, Varese)
Eleni Tzirita Zacharatou (IT University of Copenhagen)
Essam Mansour (Concordia University)
Faisal Nawab (University of California at Irvine)
Fan Zhang (Guangzhou University)
Fatemeh Nargesian (University of Rochester)
Fei Chiang (McMaster University)
Floris Geerts (University of Antwerp)
Gao Cong (Nanyang Technological University)
George Fakas (Uppsala University)
George Fletcher (Eindhoven University of Technology the Netherlands)
George Papadakis (University of Athens)
George Papastefanatos (Athena Research Center)
Giovanni Simonini (University of Modena and Reggio Emilia)
Graham Cormode (University of Warwick)
Guna Prasaad (Meta Platforms Inc.)
Guoliang Li (Tsinghua University)
Guoren Wang (Beijing Institute of Technology)
Haibo Hu (Hong Kong Polytechnic University)
Hannes Voigt (Neo4j)
Haridimos Kondylakis (FORTH-ICS)
Holger Pirk (Imperial College)
Huanchen Zhang (Tsinghua University)
Ibrahim Sabek (MIT)
Immanuel Trummer (Cornell University)
Ingo Müller (Google)
James Cheng (The Chinese University of Hong Kong)
Jeffrey Xu Yu (The Chinese University of Hong Kong)
Jens Teubner (TU Dortmund University)
Jia Yu (Washington State University)
Jian Lou (Xidian University)
Jianguo Wang (Purdue University)
Jianliang Xu (Hong Kong Baptist University)
Jianxin Li (Deakin University)
Jiawei Jiang (ETH Zurich)
Jieming Shi (Hong Kong Polytechnic University)
Jinfei Liu (Zhejiang University)
Jing Tang (Hong Kong University of Science and Technology)
John Liagouris (Boston University)
John Paparrizos (University of Chicago)
Joseph Near (University of Vermont)
Junhao Gan (University of Melbourne)
K. Selcuk Candan (Arizona State University)
Kai Wang (University of New South Wales)
Karima Echihabi (Mohammed VI Polytechnic University)
Kartik Nayak (Duke University)
Katja Hose (Aalborg University)
Kexin Rong (Stanford University)
Kun Qian (Amazon)
Kunsoo Park (Seoul National University)
Kyriakos Mouratidis (Singapore Management University)
Laks Lakshmanan (University of British Columbia)
Laurel Orr (Stanford University)
Lei Cao (MIT)
Lei Chen (Hong Kong University of Science and Technology)
Lei Li (Hong Kong University of Science and Technology, Guang Zhou)
Lijun Chang (The University of Sydney)
Lin Ma (Carnegie Mellon University)
Long Yuan (Nanjing University of Science and Technology)
Lu Qin (UTS)
Lucas Lersch (Amazon Web Services)
Lukasz Golab (University of Waterloo)
Matteo Interlandi (Microsoft)
Matteo Lissandrini (Aalborg University)
Matthias Renz (University of Kiel)
Matthias Weidlich (Humboldt University of Berlin)
Michael Abebe (University of Waterloo)

Michael H Boehlen (University of Zurich)
Michael Hay (Colgate University/Tumult Labs)
Michael Mathioudakis (University of Helsinki)
Michal Friedman (ETH)
Milos Nikolic (University of Edinburgh)
Mirek Riedewald (Northeastern University)
Mohamed Sharaf (United Arab Emirates University)
Mohammad Sadoghi (University of California, Davis)
Mostafa Milani (The University of Western Ontario)
Nick Koudas (University of Toronto)
Nikolaos Tziavelis (Northeastern University)
Nikolay Yakovets (Eindhoven University of Technology)
Ning Wang (Beijing Jiaotong University)
Oliver A Kennedy (University at Buffalo, SUNY)
Panagiotis Bouros (Johannes Gutenberg University Mainz)
Panos Vassiliadis (University of Ioannina)
Paolo Papotti (EURECOM, France)
Periklis Andritsos (University of Toronto)
Prashanth Menon (Databricks)
Raghav Kaushik (Microsoft)
Rainer Gemulla (University of Mannheim)
Raul Castro Fernandez (University of Chicago)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
Renata Borovica-Gajic (University of Melbourne)
Reynold Cheng (The University of Hong Kong, China)
Riccardo Torlone (Roma Tre University)
Ronghua Li (Beijing Institute of Technology)
Ryan C Marcus (MIT)
Ryan Stutsman (University of Utah)
Sai Wu (Zhejiang Univ)
Sairam Gurajada (Apple)
Sebastian Link (University of Auckland)
Senjuti Basu Roy (New Jersey Institute of Technology)
Seokki Lee (University of Cincinnati)
Shaoxu Song (Tsinghua University)
Shiyu Yang (Guangzhou University)
Shuai Ma (Beihang University)
Sibo Wang (The Chinese University of Hong Kong)
Siqiang Luo (Nanyang Technological University)
Sourav S Bhowmick (Nanyang Technological University)
Spyros Blanas (The Ohio State University)
Srikanta Bedathur (IIT Delhi)
Stefania Dumbrava (ENSIIE)
Stefano Paraboschi (Universita' degli Studi di Bergamo)
Steffen Zeuch (DFKI Berlin)
Steven E Whang (KAIST)
Stijn Vansummeren (Hasselt University)
Sudipto Das (Amazon Web Services)
Tarique Siddiqui (Microsoft Research)
Theodore Dalamagas (Athena Research Center)
Thomas Neumann (TU Munich)
Tian Li (Carnegie Mellon University)
Tianhao Wang (University of Virginia)
Tianzheng Wang (Simon Fraser University)
Tien Tuan Anh Dinh (Singapore University of Technology and Design)
Torben Bach Pedersen (Aalborg University)
Utku Sirin (Harvard University)
Vasiliki Kalavri (Boston University)
Vassilios S Verykios (Hellenic Open University)
Walid G Aref (Purdue)
Wang-Chiew Tan (Facebook AI)
Weiguo Zheng (Fudan University)
Wendy Hui Wang (Stevens Institute of Technology)
Wentao Wu (Microsoft Research)
Xi He (University of Waterloo)
Xiang Lian (Kent State University)
Xiangmin Zhou (RMIT University)
Xiangyao Yu (University of Wisconsin-Madison)
Xiaochun Yang (Northeastern University)
Xiaofei Zhang (University of Memphis)
Xiaoyang Wang (University of New South Wales)
Xin Cao (University of New South Wales)
Xin Huang (Hong Kong Baptist University)
Yan Zhao (Aalborg University)
Yang Cao (Kyoto University)
Yao Lu (Microsoft Research)
Ye Yuan (Beijing Institute of Technology)
Yeye He (Microsoft Research)
Yi Yu (NII)
Yinghui Wu (Case Western Reserve University)
Yingxia Shao (BUPT)
Yixiang Fang (The Chinese University of Hong Kong, Shenzhen)
Yongluan Zhou (University of Copenhagen)
You Peng (University of New South Wales)
You Wu (Google)
Yufei Tao (The Chinese University of Hong Kong)
Yuncheng Wu (National University of Singapore)
Yuyu Luo (Tsinghua University)
Zeke Wang (Zhejiang University)
Zhiwei Zhang (Beijing Institute of Technology)
Zhongle Xie (Zhejiang University)
Zhuoyue Zhao (University at Buffalo - SUNY)
Ziawasch Abedjan (Leibniz University Hannover)
Zimu Zhou (Singapore Management University)

LETTER FROM THE EDITORS IN CHIEF

It is our pleasure to present the first issue of PVLDB's Volume 16. The Proceedings of the VLDB (PVLDB) present original research papers on a broad range of topics related to all aspects of data and information management, spanning from theoretical foundations, system architectures, models and techniques, to novel applications as well as large-scale deployment and evaluation. There are four equally important categories of papers in the research track: (a) regular research, (b) scalable data science, (c) experiment, analysis & benchmark, and (d) vision papers.

We have updated the list of PVLDB topics — which can be found on the PVLDB website — to reflect new research trends and better capture the broad range of topics now being tackled by our research community. Some of the additions include *Video Management and Analytics Systems*, *Explainable AI*, *Interactive Querying & Visualization for Large Data*, *NL Interfaces to Data*, and *Recommender Engines*.

PVLDB strives to give high-quality and constructive feedback in the form of reviews and meta-reviews. Submissions are carefully peer-reviewed by an expert board of Associate Editors and reviewers. Each paper is evaluated by at least three reviewers and an Associate Editor, who summarizes in a meta-review all reviews and the results of a three-week discussion phase during which the reviewers exchange their viewpoints and converge to a joint decision. Some submissions will enter a revision phase, where the authors are given three months to prepare a revised version for another round of review. Accepted papers are then published in the journal and ultimately presented at the following VLDB conference.

We have introduced some changes this year to further improve the review process. First, we increased the number of Associate Editors, to cover the broader range of PVLDB topics and to ensure better matching of expertise and a more balanced load. Recognizing that feedback is important for improving the whole process, we are providing an online form for authors to confidentially provide feedback on the quality of the reviews and meta-review; another form allows the reviewers and the Associate Editors to give kudos or raise concerns for other members of the review board. Some submissions may now undergo an additional formal shepherding phase, where a designated shepherd will work with the authors to ensure that the final version of the paper is acceptable. In addition to using Microsoft's Conference Management Toolkit (CMT) for managing submissions, reviews, and revisions, we rely on the services of the Toronto Paper Matching System (TPMS) to suggest review assignments, the ConfLict Of IntereSt DEtection & ManagemenT System (CLOSET) for detecting conflicts of interest, and iThenticate, for detecting plagiarism.

Regarding conflicts of interest (COIs), we now apply an updated policy with a stricter definition than in the past. In particular, a COI based on co-authorship is defined as a person X being a co-author of a paper with Y in the last 3 years, or of 4 (or more) papers in the last 10 years. With stricter COI rules, declaring COIs with a large review board requires considerable extra effort. To facilitate, we provide a searchable spreadsheet of COIs with the review board based on DBLP data. Authors can use this spreadsheet as a resource when declaring COIs during submission. This resource is not meant to be authoritative or cover all types of COIs, so authors must still do their due diligence to identify and declare COIs missing from the list. We will keep updating the spreadsheet throughout the year.

Authors are expected to make available supplemental materials such as code, data, and other implementation artifacts used to produce the results reported in the paper. When there are compelling reasons why this availability requirement cannot be met, authors must explain and can be exempted from the requirement. As part of the meta-reviews for accepted submissions, Associate Editors use a standard rubric to assess the availability of supplemental materials, ensuring their openness and permanence, as well as the readability of instructions for the reuse of the artifacts by other members of the community. All accepted papers that provide supplementary materials meeting the availability requirement are awarded an official ACM badge.

This first issue of PVLDB's Volume 16 includes ten papers, spanning the topics of database engines, distributed databases, blockchain, data management for ML, crowdsourcing, and privacy — with database engines and data management for ML as the two most popular topics in this issue, at four and three papers, respectively. Out of the ten papers, one was a straight accept, two were shepherded, and seven were accepted after revision. One paper is in the scalable data science category, and the rest are regular research papers.

We are very grateful to our board of associate editors and reviewers as well as our proceedings chairs who contribute to the success of PVLDB.

Georgia Koutrika and Jun Yang
Editors-in-Chief of PVLDB Vol. 16
Program Chairs for VLDB 2023