Proceedings of the VLDB Endowment

Volume 17, No. 9 – May 2024

Editors in Chief:

Meihui Zhang and Cyrus Shahabi

Associate Editors:

Alkis Polyzotis, Amol Deshpande, Angela Bonifati, Ashraf Aboulnaga, Ashwin Machanavajjhala, Beng Chin Ooi, Boris Glavic, Ce Zhang, Divy Agrawal, Eric Lo, Fatma Ozcan, Guoliang Li, Jeffrey Xu Yu, Jian Pei, Jianliang Xu, Johannes Gehrke, K. Selçuk Candan, Kyuseok Shim, Li Xiong, Magdalena Balazinska, Matthias Boehm, Melanie Herschel, Michael Böhlen, Nikos Mamoulis, Pinar Tozun, Rachel Pottinger, Sharad Mehrotra, Surajit Chaudhuri, Tamer Özsu, Tien Tuan Anh Dinh, Walid Aref, Wei Wang, Xiaokui Xiao, Yanyan Shen, Yongxin Tong, Zi Huang

Publication Editors:

Ju Fan, Yang Cao, Xiaoou Ding

PVLDB – Proceedings of the VLDB Endowment

Volume 17, No. 9, May 2024.

All papers published in this issue will be presented at the 50th International Conference on Very Large Data Bases, Guangzhou, China, 2024.

Copyright 2024 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 17, Number 9, May 2024

Pages i – vii and 2123 - 2404

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. 17 iv

Research Papers

CIVET: Exploring Compact Index for Variable-Length Subsequence Matching on Time Series 2123
Haoran Xiong, Hang Zhang, Zeyu Wang, Zhenying He, Peng Wang, X. Sean Wang

Spatialyze: A Geospatial Video Analytics System with Spatial-Aware Optimizations 2136
Chanwut Kittivorawong, Yongming Ge, Yousef Helal, Alvin Cheung

Optimal Matrix Sketching over Sliding Windows 2149
Hanyan Yin, Dongxie Wen, Jiajun Li, Zhewei Wei, Xiao Zhang, Zengfeng Huang, Feifei Li

Window Function Expression: Let the Self-join Enter 2162
Radim Baca

SplitDF: Splitting Dataframes for Memory-Efficient Data Analysis 2175
Aarati Kakaraparthy, Jignesh M. Patel

Sampling Methods for Inner Product Sketching 2185
Majid Daliri, Juliana Freire, Christopher Musco, Aécio Santos, Haoxiang Zhang

DIDS: Double Indices and Double Summarizations for Fast Similarity Search 2198
Han Hu, Jiye Qiu, Hongzhi Wang, Bin Liang, Songling Zou

Improving Graph Compression for Efficient Resource-Constrained Graph Analytics 2212
Qian Xu, Juan Yang, Feng Zhang, Zheng Chen, Jiawei Guan, Kang Chen, Ju Fan, Youren Shen, Ke Yang, Yu Zhang, Xiaoyong Du

Efficient Unsupervised Community Search with Pre-trained Graph Transformer 2227
Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, Ying Zhang

DET-LSH: A Locality-Sensitive Hashing Scheme with Dynamic Encoding Tree for Approximate Nearest Neighbor Search 2241
Jiuqi Wei, Botao Peng, Xiaodong Lee, Themis Palpanas

BIRD: Efficient Approximation of Bidirectional Hidden Personalized PageRank 2255
Haoyu Liu, Siqiang Luo

GENTI: GPU-powered Walk-based Subgraph Extraction for Scalable Representation Learning on Dynamic Graphs 2269
Zihao Yu, Ningyi Liao, Siqiang Luo

ArcheType: A Novel Framework for Open-Source Column Type Annotation using Large Language Models 2279
Benjamin Feuer, Yurong Liu, Chinmay Hegde, Juliana Freire

Trajectory Similarity Measurement: An Efficiency Perspective 2293
Yanchuan Chang, Egemen Tanin, Gao Cong, Christian S. Jensen, Jianzhong Qi

BYO: A Unified Framework for Benchmarking Large-Scale Graph Containers 2307
Brian Wheatman, Xiaojun Dong, Zheqi Shen, Laxman Dhulipala, Jakub Łącki, Prashant Pandey, Helen Xu

Secure and Verifiable Data Collaboration with Low-Cost Zero-Knowledge Proofs 2321
Yizheng Zhu, Yuncheng Wu, Zhaojing Luo, Beng Chin Ooi, Xiaokui Xiao

Rashnu: Data-Dependent Order-Fairness 2335
Heena Nagda, Shubhendra Pal Singhal, Mohammad Javad Amiri, Boon Thau Loo

Sparcle: Boosting the Accuracy of Data Cleaning Systems through Spatial Awareness 2349
Yuchuan Huang, Mohamed F. Mokbel

TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods 2363
Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Linn Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, Bin Yang

FSM: A Fine-grained Splitting and Merging Framework for Dual-balanced Graph Partition 2378
Chengjun Liu, Zhuo Peng, Weiguo Zheng, Lei Zou

Efficient and Reliable Estimation of Knowledge Graph Accuracy 2392
Stefano Marchesin, Gianmaria Silvello

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 17

Editors in Chief of PVLDB
Meihui Zhang (Beijing Institute of Technology)
Cyrus Shahabi (University of Southern California)

Associate Editors of PVLDB
Alkis Polyzotis (Databricks)
Amol Deshpande (University of Maryland at College Park)
Angela Bonifati (Lyon 1 University)
Ashraf Aboulnaga (Qatar Computing Research Institute, HBKU)
Ashwin Machanavajjhala (Duke)
Beng Chin Ooi (NUS)
Boris Glavic (Illinois Institute of Technology)
Ce Zhang (ETH)
Divy Agrawal (University of California, Santa Barbara)
Eric Lo (Chinese University of Hong Kong)
Fatma Ozcan (Google)
Guoliang Li (Tsinghua University)
Jeffrey Xu Yu (Chinese University of Hong Kong)
Jian Pei (Simon Fraser University)
Jianliang Xu (Hong Kong Baptist University)
Johannes Gehrke (Microsoft)
K. Selçuk Candan (Arizona State University)
Kyuseok Shim (Seoul National University)
Li Xiong (Emory University)
Magdalena Balazinska (UW)
Matthias Boehm (Technische Universität Berlin)
Melanie Herschel (Universität Stuttgart)
Michael Böhlen (University of Zurich)
Nikos Mamoulis (University of Ioannina)
Pinar Tozun (IT University of Copenhagen)
Rachel Pottinger (Univ. of British Columbia)
Sharad Mehrotra (U.C. Irvine)
Surajit Chaudhuri (Microsoft)
Tamer Özsu (University of Waterloo)
Tien Tuan Anh Dinh (Deakin University)
Walid Aref (Purdue University)
Wei Wang (ByteDance)
Xiaokui Xiao (National University of Singapore)
Yanyan Shen (Shanghai Jiao Tong University)
Yongxin Tong (Beihang University)
Zi Huang (University of Queensland)

Publication Editors
Ju Fan (Renmin University of China)
Yang Cao (Tokyo Institute of Technology)
Xiaoou Ding (Harbin Institute of Technology)

PVLDB Managing Editor
Wolfgang Lehner (Dresden University of Technology)

PVLDB Advisory Board
Vanessa Braganholo (Universidade Federal Fluminense)
Sourav S Bhowmick (Nanyang Technological University)
Torsten Grust (University of Tuebingen)
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)
Felix Naumann (HPI)
Georgia Koutrika (Athena Research Center)
Jun Yang (Duke University)
Meihui Zhang (Beijing Institute of Technology)
Cyrus Shahabi (University of Southern California)
Nesime Tatbul (Intel Labs and MIT)
Themis Palpanas (Universite Paris Cite)

Review Board

Abolfazl Asudeh (University of Illinois Chicago)
Aditya Parameswaran (University of California, Berkeley)
Ahmed S. Abdelhamid (Purdue University)
Ahmed Eldawy (University of California, Riverside)
Ahmed El-Roby (Carleton University)
Ahmed Mahmood (Google)
Alberto Lerner (University of Fribourg, Switzerland)
Alexander Thomson (Google)
Amr Magdy (University of California Riverside)
Andreas Züfle (Emory University)
Angelos Christos Anadiotis (Oracle)
Anja Gruenheid (Microsoft)
Anthony Tung (National U. of Singapore)
Anton Dignös (Free University of Bozen-Bolzano, Italy)
Arijit Khan (Aalborg University)
Avrilia Floratou (Microsoft)
Baihua Zheng (Singapore Management University)
Bailu Ding (Microsoft Research)
Berthold Reinwald (IBM Research-Almaden)
Bin Yang (East China Normal University)
Bingsheng He (National University of Singapore)
Bolin Ding (Data Analytics and Intelligence Lab, Alibaba Group)
Brandon Haynes (Microsoft Gray Systems Lab)
Chao Zhang (University of Waterloo)
Cheng Long (Nanyang Technological University)
Chengfei Liu (Swinburne University of Technology)
Chengkai Li (The University of Texas at Arlington)
Chengliang Chai (Beijing Institute of Technology)
Chrysanthi Kosyfaki (University of Ioannina)
Chunwei Liu (MIT)
Cong Yan (Microsoft research)
Daisy Zhe Wang (University of Florida)
Dan Kifer (Pennsylva State Univ., USA)
Dan Lin (Vanderbilt University)
Daniel Kang (UIUC)
Demetrios Zeinalipour-Yazti (University of Cyprus)
Dimitris Papadias (HKUST)
Dong Deng (Rutgers Universituy - New Brunswick)
Dong Wen (University of New South Wales)
Dong Xie (Penn State University)
Dongxiang Zhang (Zhejiang University)
Dumitrel Loghin (National University of Singapore)
Egemen Tanin (University of Melbourne)
El Kindi Rezig (Massachusetts Institute of Technology)
Elena Ferrari (University of Insubria, Varese)
Eser Kandogan (Megagon Labs)
Essam Mansour (Concordia University)
Fan Zhang (Guangzhou University)
Fatemeh Nargesian (University of Rochester)
Fei Chiang (McMaster University)
Feng Zhang (Renmin University of China)
Florin Rusu (UC Merced)
Gabriel Ghinita (Hamad Bin Khalifa University)
Gao Cong (Nanyang Technological Univesity)
George Fakas (Uppsala University)
Haibo Hu (Hong Kong Polytechnic University)
Holger Pirk (Imperial College)
Hong Cheng (Chinese University of Hong Kong)
Hongzhi Wang (Harbin Institute of Technology)
Hua Lu (Roskilde University)
Huanchen Zhang (Tsinghua University)
Huiping Cao (New Mexico State University)
Ibrahim Sabek (MIT)
Ilaria Bartolini (University of Bologna)
Jana Giceva (TU Munich)
Jennie Rogers (Northwestern University)
Jia Zou (Arizona State University)
Jian Lou (Zhejiang University)
Jiangshan Yu (Monash University)
Jianguo Wang (Purdue University)
Jiannan Wang (Simon Fraser University)
Jianqiu Xu (Nanjing University of Aeronautics and Astronautics)
Jianxin Li (Deakin University)
Jieming Shi (The Hong Kong Polytechnic University)
Jin Wang (Megagon Labs)
Jinfei Liu (Zhejiang University)
Johes Bater (Tufts University)
John Liagouris (Boston University)
Jonathan Goldstein (Microsoft)
Ju Fan (Renmin University of China)
Juhua Hu (University of Washington)
Kai Wang (Shanghai Jiao Tong University)
Kangfei Zhao (Beijing Insitute of Technology)
Karima Echihabi (Mohammed VI Polytechnic University)
Katja Hose (TU Wien)
Khuzaima Daudjee (University of Waterloo)
Kyoungmin Kim (POSTECH)
Lawrence Benson (HPI, University of Potsdam)
Lei Chen (Hong Kong University of Science and Technology)
Lei Zou (Peking University)
Leong Hou U (University of Macau)
Lin Ma (University of Michigan)
Lingyang Chu (McMaster University)
Liyue Fan (UNC Charlotte)
Lu Chen (Zhejiang University)
Luigi Bellomarini (Banca d'Italia)
Madelon Hulsebos (University of Amsterdam)
Manolis Terrovitis (IMIS, Athena RC)
Marco Patella (University of Bologna)
Mario Nascimento (Northeastern University)
Matteo Lissandrini (Aalborg University)
Matthias Renz (University of Kiel)
Michael Hay (Colgate University & Tumult Labs)
mingjie tang (Ant Financial)
Mirek Riedewald (Northeastern University)
Mohamed S. Hassan (Google)
Mohamed Mokbel (University of Minnesota - Twin Cities)
Mohammad Javad Amiri (University of Pennsylvania)
Mostafa Milani (The University of Western Ontario)
Mourad OUZZANI (Qatar Computing Research Institute, HBKU)
Nesime Tatbul (Intel Labs and MIT)
Norman May (SAP SE)
Oliver A Kennedy (University at Buffalo, SUNY)

Panagiotis Bouros (Johannes Gutenberg University Mainz)
Papotti Paolo (EURECOM)
Patrick Damme (Technische Universität Berlin)
Peng Peng (Hunan University)
Philippe Bonnet (IT Univ Copenhagen, Denmark)
Pinar Karagoz (METU, Turkey)
Prashant Pandey (University of Utah)
Primal Pappachan (Penn State University)
Qichen Wang (Hong Kong Baptist University)
Qing Liu (Zhejiang University)
Qun Chen (Northwestern Polytechnical University)
Renata Borovica-Gajic (University of Melbourne)
Rihan Hai (TU Delft)
Ritesh Ahuja (Oracle Labs)
Roger Zimmermann (NUS)
Ronghua Li (Beijing Institute of Technology)
Sai Wu (Zhejiang Univ)
Sanjay Krishnan (UChicago)
Senjuti Basu Roy (NJIT)
Seokki Lee (University of Cincinnati)
Shantanu Sharma (New Jersey Institute of Technology)
Shaofeng Cai (National University of Singapore)
Shaoxu Song (Tsinghua University)
Shuai Ma (Beihang University)
Shuang Hao (Beijing Jiaotong University)
Sibo Wang (The Chinese University of Hong Kong)
Stefania Dumbrava (ENSIIE)
Stefano Paraboschi (Universita' degli Studi di Bergamo)
Sujaya Maiyya (University of Waterloo)
Tarique Siddiqui (Microsoft Research)
Thanaa Ghanem (Metro State University)
Thang Dinh (VCU)
Themis Palpanas (Universite Paris Cite)
Thomas Neumann (TUM)
Tianhao Wang (University of Virginia)
Tianzheng Wang (Simon Fraser University)
Tieying Zhang (Bytedance)
Tristan Allard (Univ Rennes, CNRS, IRISA)
Umar Farooq Minhas (Apple)
Utku Sirin (Harvard University)
Viktor Leis (Technische Universität München)
Vincenzo Gulisano (Chalmers University of Technology)
Vraj Shah (IBM Research)
Wang-Chien Lee (Pennsylvania State University, USA)
WEI LU (Renmin University of China)
Wei Wang (Hong Kong University of Science and Technology (Guangzhou))
Wei-Shinn Ku (Auburn University)
Wenchao Zhou (Alibaba Group)
Wendy Hui Wang (Stevens Institute of Technology)
Xiang Lian (Kent State University)
Xiang Zhao (National University of Defence Technology)
Xiangyao Yu (University of Wisconsin-Madison)
Xiao Hu (Duke University)
Xiao Hu (University of Waterloo)
Xiaochun Yang (Northeastern University)
Xiaofang Zhou (The Hong Kong University of Science and Technology)
Xiaofei Zhang (University of Memphis)
Xiaohui Yu (York University)
Xiaoli Wang (Xiamen University)
Xin Huang (Hong Kong Baptist University)
Xin Wang (Tianjin University)
Xingquan Zhu (Florida Atlantic University)
Yanfeng Zhang (Northeastern University)
Yang Cao (Tokyo Institute of Technology)
Yannis Chronis (Google)
Yao Lu (Microsoft Research)
Ye Yuan (Beijing Institute of Technology)
Yeye He (Microsoft Research)
Ying Zhang (University of Technology Sydney)
Yingxia Shao (BUPT)
Yu Yang (City University of Hong Kong)
Yuhao Zhang (University of California, San Diego)
Yuncheng Wu (National University of Singapore)
Yunjun Gao (Zhejiang University)
Yuval Moskovitch (Ben Gurion University)
Yuxiang Zeng (Beihang University)
Zhaojing Luo (National University of Singapore)
Zhengjie Miao (Duke University)
Zhichao Cao (Arizona State University)
Zhifeng Bao (RMIT University)
Zhiwei Zhang (Beijing Institute of Technology)
Zhongle Xie (Zhejiang University)
Zhuoyue Zhao (University at Buffalo - SUNY)
Ziawasch Abedjan (Leibniz Universität Hannover)
Ziliang Lai (Chinese University of Hong Kong)
Zimu Zhou (City University of Hong Kong)

LETTER FROM THE EDITORS IN CHIEF

It is our pleasure to present the ninth issue of Volume 17 of PVLDB (Proceedings of the VLDB). PVLDB is dedicated to showcasing original research papers that encompass a wide spectrum of subjects within the area of data and information management. Our coverage spans from fundamental theoretical principles and cutting-edge system architectures to innovative models, techniques, novel applications, and the comprehensive assessment and deployment of large-scale solutions. In our research track, we feature four equally significant categories of papers: (a) regular research, (b) scalable data science (SDS), (c) experiment, analysis & benchmark (EA&B), and (d) vision papers.

The ninth issue of PVLDB's Volume 17 includes 21 papers, spanning the topics of Graph and Network Data; Database Engines; Information Integration and Data Quality; Data Mining and Analytics; Novel Database Architectures; Spatial and temporal databases; Machine Learning, AI, and Databases; Data Privacy and Security; and Distributed Database Systems. Three topics stood out, the most popular ones in this issue: Graph and Network Data (6 papers), Database Engines (3 papers), and Information Integration and Data Quality (3 papers).

Out of the 21 papers, three are in the experiment, analysis & benchmark category, one is in the scalable data science category, and the rest are regular research papers. Eight papers were accepted after a shepherding round, one paper was a direct accept, and the others were accepted after revision.

PVLDB is committed to providing valuable and constructive feedback through a rigorous review process. All submissions undergo meticulous peer review by a team of accomplished Associate Editors and dedicated reviewers. Each paper receives comprehensive evaluation from a minimum of three reviewers, along with the oversight of an Associate Editor. During a three-week discussion phase, reviewers engage in a thorough exchange of perspectives, ultimately converging on a consensus, which is summarized in a meta-review. Some submissions may proceed to a revision phase, affording authors a three-month window to refine their work for subsequent review cycles.

This issue is the result of all the work put in by the authors as well as the great commitment and effort of our associate editors and reviewers as well as our proceedings chairs.

Meihui Zhang and Cyrus Shahabi
Editors-in-Chief of PVLDB Vol. 17
Program Chairs for VLDB 2024