# Proceedings of the VLDB Endowment

Volume 15, No. 9 – May 2022

Editors in Chief:

**Fatma Özcan, Juliana Freire and Xuemin Lin**

Associate Editors:

**Arun Kumar, Azza Abouzied, Beng Chin Ooi, Boris Glavic, Dan Suciu,**
**Divyakant Agrawal, Eugene Wu, Georgia Koutrika, Ioana Manolescu,**
**Jeffrey Xu Yu, Julia Stoyanovich, Jun Yang, K. Selçuk Candan,**
**Khuzaima Daudjee, Laure Berti-Equille, Lei Chen, Mohamed Mokbel,**
**Neoklis Polyzotis, Paolo Papotti, Peter Boncz, Sebastian Schelter,**
**Sourav S Bhowmick, Surajit Chaudhuri, Themis Palpanas, Vanessa Braganholo,**
**Viktor Leis, Wang-Chiew Tan, Wenjie Zhang, Wook-Shin Han, Xiaofang Zhou**

Publication Editors:

**Lijun Chang and Xin Cao**

PVLDB – Proceedings of the VLDB Endowment

Volume 15, No. 9, May 2022.

All papers published in this issue will be presented at the 48th International Conference on Very Large Data Bases, Sydney, Australia, 2022.

# 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 15, Number 9, May 2022

Pages i – viii and 1739 - 1977

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. 15 ..... iv


**Research Papers**


Scalable Byzantine Fault Tolerance via Partial Decentralization ..... 1739
*Balaji Arun, Binoy Ravindran*

Efficient and Error-bounded Spatiotemporal Quantile Monitoring in Edge Computing Environments 1753
*Huan Li, Lanjing Yi, Bo Tang, Hua Lu, Christian S Jensen*

HDPView: Differentially Private Materialized View for Exploring High Dimensional Relational Data .. 1766
*Fumiyuki Kato, Tsubasa Takahashi, Shun Takagi, Yang Cao, Seng Pei Liew, Masatoshi Yoshikawa*

Anomaly Detection in Time Series: A Comprehensive Evaluation ..... 1779
*Sebastian Schmidl, Phillip Wenig, Thorsten Papenbrock*

Guided Exploration of Data Summaries ..... 1798
*Brit Youngmann, Sihem Amer-yahia, Aurélien Personnaz*

Facilitating Database Tuning with Hyper-Parameter Optimization: A Comprehensive Experimental Evaluation ..... 1808
*Xinyi Zhang, Zhuo Chang, Yang Li, Hong Wu, Jian Tan, Feifei Li, Bin Cui*

Efficient Secure and Verifiable Location-Based Skyline Queries over Encrypted Data ..... 1822
*Zuan Wang, Xiaofeng Ding, Hai Jin, Pan Zhou*

AB-tree: Index for Concurrent Random Sampling and Updates ..... 1835
*Zhuoyue Zhao, Dong Xie, Feifei Li*

On Repairing Timestamps for Regular Interval Time Series ..... 1848
*Chenguang Fang, Shaoxu Song, Yinan Mei*

Towards Event Prediction in Temporal Graphs ..... 1861
*Wenfei Fan, Ruochun Jin, Ping Lu, Chao Tian, Ruiqi Xu*

Decentralized Crowdsourcing for Human Intelligence Tasks with Efficient On-Chain Cost ..... 1875
*Yihuai Liang, Yan Li, Byeong-seok Shin*

Towards Distributed Bitruss Decomposition on Bipartite Graphs ..... 1889
*Yue Wang, Ruiqi Xu, Xun Jian, Alexander Zhou, Lei Chen*

Generalized Supervised Meta-blocking ..... 1902
*Luca Gagliardelli, George Papadakis, Giovanni Simonini, Sonia Bergamaschi, Themis Palpanas*

Your Read is Our Priority in Flash Storage ..... 1911
*Mijin An, Soojun Im, Dawoon Jung, Sang Won Lee*

New Wine in an Old Bottle: Data-aware Hash Functions for Bloom Filters ..... 1924
*Arindam Bhattacharya, Chathur Gudesa, Amitabha Bagchi, Srikanta Bedathur*

SANCUS: Staleness-Aware Communication-Avoiding Full-Graph Decentralized Training in Large-Scale Graph Neural Networks ..... 1937
*Jingshu Peng, Zhao Chen, Yingxia Shao, Yanyan Shen, Lei Chen, Jiannong Cao*

CORE: a COmplex event Recognition Engine ..... 1951
*Marco Bucchi, Alejandro Grez, Andres F Quintana, Cristian Riveros, Stijn Vansummeren*

TAOBench: An End-to-End Benchmark for Social Networking Workloads ..... 1965
*Audrey Cheng, Xiao Shi, Aaron N Kabcenell, Shilpa Lawande, Hamza Qadeer, Jason Chan, Harrison Tin, Ryan Zhao, Peter Bailis, Mahesh Balakrishnan, Nathan Bronson, Natacha Crooks, Ion Stoica*

## PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 15


**Editors in Chief of PVLDB**
Fatma Ozcan (Google)
Juliana Freire (New York University)
Xuemin Lin (University of New South Wales)

**Associate Editors of PVLDB**
Arun Kumar (University of California, San Diego)
Azza Abouzied (NYU Abu Dhabi)
Beng Chin Ooi (NUS)
Boris Glavic (Illinois Institute of Technology)
Dan Suciu (University of Washington)
Divyakant Agrawal (University of California, Santa Barbara)
Eugene Wu (Columbia University)
Georgia Koutrika (ATHENA)
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)
Jeffrey Xu Yu (Chinese University of Hong Kong)
Julia Stoyanovich (New York University)
Jun Yang (Duke University)
K. Seçuk Candan (Arizona State University)
Khuzaima Daudjee (University of Waterloo)
Laks Lakshmanan (The University of British Columbia)
Laure Berti-Equille (IRD)
Lei Chen (Hong Kong University of Science and Technology)
Mohamed Mokbel (University of Minnesota, Twin Cities)
Neoklis Polyzotis (Google)
Paolo Papotti
Peter Boncz (CWI)
Sebastian Schelter (University of Amsterdam)
Sharad Mehrotra (U.C. Irvine)
Sourav S Bhowmick (Nanyang Technological University)
Surajit Chaudhuri (Microsoft Research)
Themis Palpanas (University of Paris)
Vanessa Braganholo (Fluminense Federal University)
Viktor Leis (Friedrich Schiller University Jena)
Wang-Chiew Tan (Megagon Labs)
Wenjie Zhang (University of New South Wales)
Wook-Shin Han (POSTECH)
Xiaofang Zhou (Hong Kong University of Science and Technology)

**Publication Editors**
Lijun Chang (University of Sydney)
Xin Cao (University of New South Wales)

**PVLDB Managing Editor**
Wolfgang Lehner (Dresden University of Technology)

**PVLDB Advisory Committee**
Felix Naumann (HPI)
Juliana Freire (New York University)
Xuemin Lin (U of New South Wales)
Georgia Koutrika (Athena Research Center)
Jun Yang (Duke University)
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)
Graham Cormode (University of Warwick)
Divesh Srivastava (AT&T Labs-Research)
Wolfgang Lehner (TU Dresden)

**Review Board**

Abolfazl Asudeh (University of Michifan)
Aécio Santos (New York University)
Ahmed Eldawy (University of California, Riverside)
Alexander Hall (RelationalAI)
Alexander J Ratner (University of Washington)
Aline Bessa (New York University)
Alkis Simitsis (Athena Research Center)
Altigran da Silva (Universidade Federal do Amazonas)
AnHai Doan (University of Wisconsin-Madison)
Anna Fariha (Microsoft)
Anton Dignös (Free University of Bozen-Bolzano)
Antonio Cavalcante Araujo Neto (University of Alberta)
Arijit Khan (Nanyang Technological University)
Arvind Arasu (Microsoft)
Babak Salimi (University of California, San Diego)
Bailu Ding (Microsoft Research)
Bertram Ludaescher (University of Illinois)
Bolong Zheng (Huazhong University of Science and Technology)
Brandon Haynes (Gray Systems Lab, Microsoft)
Byron Choi (Hong Kong Baptist University)
Carlo Curino (Microsoft -- GSL)
Carlos Scheidegger (The University of Arizona)
Carsten Binnig (TU Darmstadt)
Ce Zhang (ETH)
Cheng Long (Nanyang Technological University)
Chengfei Liu (Swinburne University of Technology)
Chuan Lei (Instacart)
Chunbin Lin (Amazon AWS)
Curtis Dyreson (Utah State University)
Dan Kifer (Pennsylva State University)
Dana M Van Aken (Carnegie Mellon University)
Daniel Deutch (Tel Aviv University)
Daniel Oliveira (UFF, Brazil)
David Koop (Northern Illinois University)
Davide Mottin (Aarhus University)
Dong Xie (Penn State University)
Eduardo Ogasawara (CEFET-RJ)
Eleni Tzirita Zacharatou (TU Berlin)
Fabio Porto (LNCC)
Faisal Nawab (University of California at Irvine)
Fan Zhang (Guangzhou University)
Fatemeh Nargesian (University of Rochester)
Fei Chiang (McMaster University)
Florin Rusu (UC Merced)
Floris Geerts (University of Antwerp)
Fotis Psallidas (Microsoft)
George Fletcher (Eindhoven University of Technology)
George Papadakis (University of Athens)
Gerhard Weikum (Max-Planck-Institut fur Informatik)
Germain Forestier (University of Haute Alsace)
Guoliang Li (Tsinghua University)
Haipeng Dai (Nanjing University)
Harish Doraiswamy (Microsoft Research India)
Heiko Mueller (DeepReason.ai)
Herodotos Herodotou (Cyprus University of Technology)
Holger Pirk (Imperial College)
Hongzhi Yin (The University of Queensland)
Huiping Cao (New Mexico State University)
Immanuel Trummer (Cornell)
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)
Ippokratis Pandis (Amazon)
Ishtiyaque Ahmad (University of California, Santa Barbara)
Jae-Gil Lee (KAIST)
Jana Giceva (TU Munich)
Jeffrey Xu Yu (Chinese University of Hong Kong)
Jens Teubner (TU Dortmund University)
Jia Zou (Arizona State University)
Jian Pei (Simon Fraser University)
Jianguo Wang (Purdue University)
Jiannan Wang (Simon Fraser University)
Jianxin Li (Deakin University)
Jianye Yang (Central South University)
Jiwon Seo (Hanyang University)
Johannes Gehrke (Microsoft)
Jorge Arnulfo Quiane Ruiz (TU Berlin)
Joseph Near (University of Vermont)
Junhu Wang (Griffith University)
Kaiping Zheng (National University of Singapore)
Kangfei Zhao (The Chinese University of Hong Kong)
Karima Echihabi (Mohammed VI Polytechnic University)
Katja Hose (Aalborg University)
Kenneth A Ross (Columbia University)
Kostas Zoumpatianos (Snowflake Computing)
Lei Zou (Peking University)
Leopoldo Bertossi (Universidad Adolfo Ibanez)
Li Xiong (Emory University)
Lianke Qin (University of California, Santa Barbara)
Lijun Chang (The University of Sydney)
Lin Ma (Carnegie Mellon University)
Long Yuan (Nanjing University of Science and Technology)
Lu Qin (UTS)
Luciano Barbosa (Universidade Federal de Pernambuco)
Marcelo Arenas (Universidad Catolica & IMFD)
Maria Luisa Sapino (U. Torino)
Matteo Lissandrini (Aalborg University)
Matthias Boehm (Graz University of Technology)
Matthias Renz (University of Kiel)
Max Heimel (Snowflake)
Maximilian Schleich (University of Washington)
Meihui Zhang (Beijing Institute of Technology)
Melanie Herschel (Universität Stuttgart)
Michael Abebe (University of Waterloo)
Min Xie (Instacart)
Mirella M Moro (Universidade Federal de Minas Gerais)
Mohamed Sarwat (Arizona State University)
Mohammad Dashti (MongoDB)
Mohammad Javad Amiri (University of Pennsylvania)
Mohammad Sadoghi (University of California, Davis)
Muhammad Aamir Cheema (Monash University)

Nikita Bhutani (Megagon Labs)
Oliver A Kennedy (University at Buffalo, SUNY)
Panos K. Chrysanthis (University of Pittsburgh)
Paolo Missier (Newcastle University)
Parth Nagarkar (NMSU)
Paul Groth (University of Amsterdam)
Peng CHENG (East China Normal University)
Peter Pietzuch (Imperial College London)
Pierangela Samarati (Universita delgi Studi di Milano)
Pinar Karagoz (METU, Turkey)
Pinar Tozun (IT University of Copenhagen)
Prithu Banerjee (UBC)
Raoni Lourenço (New York University)
Raul Castro Fernandez (UChicago)
Ravi Ramamurthy (Microsoft)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
Renata Borovica-Gajic (University of Melbourne)
Reynold Cheng (The University of Hong Kong)
Rui Mao (Shenzhen University)
Ruoming Jin (Kent State University)
Sai Wu (Zhejiang University)
Sainyam Galhotra (University of Chicago)
Sanjay Krishnan (University of Chicago)
Sanjib Kumar Das (Google)
Sayan Ranu (IIT Delhi)
Sebastian Link (University of Auckland)
Semih Salihoglu (University of Waterloo)
Senjuti Basu Roy (New Jersey Institute of Technology)
Sergey Melnik (Google)
Shantanu Sharma (New Jersey Institute of Technology)
Shaoxu Song (Tsinghua University)
Sheng Wang (New York University)
Shimin Chen (Chinese Academy of Sciences)
Shumo Chu (University of California, Santa Barbara)
Shweta Jain (University of Illinois, Urbana-Champaign)
Sibo Wang (The Chinese University of Hong Kong)
Srinivasan Keshav (University of Cambridge)
Steffen Zeuch (DFKI GmbH)
Steven E Whang (KAIST)
Subarna Chatterjee (Harvard University)
Sudip Roy (Google)
Supun C Nakandala (University of California, San Diego)
Tamer Özsu (University of Waterloo)
Tarique A Siddiqui (Microsoft Research)
Thomas Heinis (Imperial College)
Thomas Neumann (TUM)
Tianzheng Wang (Simon Fraser University)
Tien Tuan Anh Dinh (Singapore University of Technology and Design)
Tilmann Rabl (HPI, University of Potsdam)
Ting Yu (Qatar Computing Research Institute)
Torben Bach Pedersen (Aalborg University)
Torsten Grust (Universität Tübingen)
Umar Farooq Minhas (Microsoft Research)
Vasiliki Kalavri (Boston University)
Verena Kantere (National Technical University of Athens)
Victor Zakhary (Oracle)
Vivek Narasayya (Microsoft Research)
Vraj Shah (University of California, San Diego)
Walid G Aref (Purdue)
Wasay Abdul (Harvard)
Wei Wang (Hong Kong University of Science and Technology (Guangzhou))
Wei Lu (Renmin university of china)
Weiren Yu (University of Warwick)
Wen Hua (The University of Queensland)
Wolfgang Lehner (TU Dresden)
Xi He (University of Waterloo)
Xiang Lian (Kent State University)
Xiao Qin (IBM Research)
Xiaofei Zhang (University of Memphis)
Xiaokui Xiao (National University of Singapore)
Xiaolan Wang (Megagon Labs)
Xiaoyang Wang (Zhejiang Gongshang University)
Xin Huang (Hong Kong Baptist University)
Yael Amsterdamer (Bar-Ilan university)
Yanyan Shen (Shanghai Jiao Tong University)
Ye Yuan (Northeastern University)
Yeye He (Microsoft Research)
Yi Chen (NJIT)
Yi Lu (MIT)
Yikai Zhang (Chinese University of Hong Kong)
Yinan Li (Microsoft Research)
Ying Zhang (University of Technology Sydney)
Yongxin Tong (Beihang University)
Yuanyuan Zhu (Wuhan University)
Yue Wang (Shenzhen Institute of Computing Sciences, Shenzhen University)
Yufei Tao (Chinese University of Hong Kong)
Yuliang Li (Megagon Labs)
Yuncheng Wu (National University of Singapore)
Yunjun Gao (Zhejiang University)
Yuval Moskovitch (University of Michigan)
Zhifeng Bao (RMIT University)
Zhongle Xie (Zhejiang University)
Zi Huang (University of Queensland)
Ziawasch Abedjan (Leibniz Universität Hannover)
Zohar Karnin (Amazon)
Zsolt István (IT University of Copenhagen)

## LETTER FROM THE EDITORS IN CHIEF

We are very pleased to present the ninth issue of the Proceedings of the VLDB (PVLDB) Volume 15. PVLDB publishes research in the area of database and information system technology. Together with expert boards of associate editors and reviewers, submissions are carefully peer-reviewed, often entering a revision phase, then published in the journal and ultimately presented at the following VLDB conference. We are very grateful to all colleagues who contribute to the success of PVLDB.

This issue includes eighteen papers covering a broad range of topics, including blockchains, privacy-preserving analytics, exploratory data analysis, spatial and temporal analytics, systems, graph mining, anomaly detection, indexing, event recognition, crowdsourcing, graph neural networks.

Arun and Ravindran propose a new approach for designing Byzantine Fault-Tolerant (BFT) consensus protocols that aims to improve their efficiency. Liang et al. proposed the use of blockchains to support cryptocurrency payments and serve as a trustworthy judge to resolve disputes between a worker and a requester in a decentralized setting. They provide formal definitions for decentralized human intelligence tasks and prove that their construction prevents false-reporting and free-riding.

Li et al. address the problem of quantile computation for spatiotemporal data in IoT applications; they propose an approach that is efficient and supports bounded errors. Fang et al. formalize the problem of timestamp repair and propose exact and approximate algorithms to address the problem; their approach is implemented as a function available in the open-source Apache IoTDB.

Kato et al. introduce new techniques that enable the exploration of high-dimensional sensitive data while preserving privacy. Wang et al. consider the problem of supporting secure and verifiable location-based skyline queries over encrypted data and propose an efficient approach to support these queries.

Zhao et al. introduce AB-tree, a new index that supports highly concurrent random sampling and update operations to support approximate queries. Bhattacharya et al. introduce Projection Hash Bloom Filter (PHBF), a space-efficient partitioned Bloom filter which uses random projection based data-aware hash functions. They provide both a theoretical analysis and experimental evaluation of their approach.

To avoid read stalls in DBMS buffers, An et al. propose a fused read and write as a new storage interface.

Bucchi et al. introduce CEQL, a functional query language designed to recognize complex events and present its implementation in the COmplex event Recognition Engine (CORE). They show, through an experimental evaluation, that CORE outperforms existing systems by one to five orders of magnitude in throughput on different query workloads. Fan et al. address the problem of event prediction; they propose the use of temporal association rules to enrich event prediction models with temporal conditions.

Wang et al. propose distributed algorithms for bitruss decomposition and experimentally show that their methods are able to handle graphs with 30 trillion butterflies in 2.5 hours, while existing parallel methods under shared-memory model fail to scale to such large graphs. Peng et al. consider the problem of how to efficiently train graph neural networks (GNNs) and propose SANCUS, a staleness-aware communication-avoiding decentralized GNN system. Cheng et al. present TAOBench, an open-source benchmark that captures the social graph workload at Meta and that accurately simulates the production request patterns of an online social network. They demonstrate the usefulness of the benchmark through an evaluation of distributed data stores.

Schmidl et al. carry out a detailed evaluation of 71 anomaly detection algorithms over 976 datasets; besides an overview of the techniques, they experimentally compare their effectiveness, efficiency, and robustness. Zhang et al. perform a comprehensive evaluation of database configuration tuning techniques and argue that hyper-parameter optimization algorithms, developed for machine learning applications, can be used to further enhance the database configuration tuning.

Youngmann et al. study the problem of data summarization in the context of exploratory data analysis. Instead of producing a single summary, they adopt a divide-and-conquer approach to produce a connected set of summaries. They propose EDA4Sum which aims to find a sequence of summaries whose cumulated utility is maximized.

Inspired by a real-world application to deduplicate a legacy customer database containing ~7.5 million entries, Gagliardelli et al. propose to reduce the computational cost of entity resolution by extending the idea of supervised meta-blocking. Their experimental evaluation demonstrates that their approach outperforms prior state of the art and that it is also scalable.

The following papers received the PVLDB Artifacts Available badge and made available the code/data used in the results they reported: Gagliardelli et al.; An et al.; Peng et al.; Bucchi et al.; Cheng et al., Fang et al., Arun and Ravindran; Li et al.; Schmidl et al.; Youngmann et al.; Zhang et al.; Zhao et al.;

All papers will be presented at the 2022 Conference on Very Large Databases (VLDB 2022) in Sydney, Australia. We hope you enjoy reading this issue and look forward to seeing you in Sydney!

Fatma Özcan, Juliana Freire and Xuemin Lin
Editors-in-Chief of PVLDB Volume 15
Program Chairs for VLDB 2022