



# Proceedings of the VLDB Endowment

Volume 16, No. 4 – December 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. 4, December 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](mailto:info@vldb.org).

Volume 16, Number 4, December 2022

Pages i – vii and 574 - 999

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 .....	iv

### Research Papers

Cache Me If You Can: Accuracy-Aware Inference Engine for Differentially Private Data Exploration.....	574
<i>Miti Mazmudar, Thomas Humphries, Jiaxiang Liu, Matthew Rafuse, Xi He</i>	
Range Search over Encrypted Multi-Attribute Data.....	587
<i>Francesca Falzon, Evangelia Anna Markatou, Zachary Espiritu, Roberto Tamassia</i>	
HEDA: Multi-Attribute Unbounded Aggregation over Homomorphically Encrypted Database .....	601
<i>Xuanle Ren, Le Su, Zhen Gu, Sheng Wang, Feifei Li, Yuan Xie, Song Bian, Chao Li, Fan Zhang</i>	
Density Personalized Group Query .....	615
<i>Chih-Ya Shen, Shao-Heng Ko, Guang-Siang Lee, Wang-Chien Lee, De-Nian Yang</i>	
Nezha: Deployable and High-Performance Consensus Using Synchronized Clocks.....	629
<i>Jinkun Geng, Anirudh Sivaraman, Balaji Prabhakar, Mendel Rosenblum</i>	
Pantheon: Private Retrieval from Public Key-Value Store .....	643
<i>Ishtiyaque Ahmad, Divyakant Agrawal, Amr El Abbadi, Trinabh Gupta</i>	
Bayesian Sketches for Volume Estimation in Data Streams .....	657
<i>Francesco Da Dalt, Simon Scherrer, Adrian Perrig</i>	
Waffle: A Workload-Aware and Query-Sensitive Framework for Disk-Based Spatial Indexing.....	670
<i>Moin Hussain Moti, Panagiotis Simatis, Dimitris Papadias</i>	
Fast Algorithms for Denial Constraint Discovery.....	684
<i>Eduardo H. M. Pena, Fabio Porto, Felix Naumann</i>	
Toward Quantity-of-Interest Preserving Lossy Compression for Scientific Data.....	697
<i>Pu Jiao, Sheng Di, Hanqi Guo, Kai Zhao, Jiannan Tian, Dingwen Tao, Xin Liang, Franck Cappello</i>	
Scalable Graph Convolutional Network Training on Distributed-Memory Systems .....	711
<i>Gunduz Vehbi Demirci, Aparajita Haldar, Hakan Ferhatosmanoglu</i>	
Motiflets - Simple and Accurate Detection of Motifs in Time Series .....	725
<i>Patrick Schäfer, Ulf Leser</i>	
Can Foundation Models Wrangle Your Data? .....	738
<i>Avanika Narayan, Ines Chami, Laurel Orr, Christopher Ré</i>	
M2Bench: A Database Benchmark for Multi-Model Analytic Workloads .....	747
<i>Bogyong Kim, Kyoseung Koo, Undraa Enkhbat, Sohyun Kim, Juhun Kim, Bongki Moon</i>	
Parallelism-Optimizing Data Placement for Faster Data-Parallel Computations .....	760
<i>Nirvik Baruah, Peter Kraft, Fiodar Kazhamiaka, Peter Bailis, Matei Zaharia</i>	
SubStrat: A Subset-Based Optimization Strategy for Faster AutoML.....	772
<i>Teddy Lazebnik, Amit Somech, Abraham Itzhak Weinberg</i>	

MultiBiSage: A Web-Scale Recommendation System Using Multiple Bipartite Graphs at Pinterest .....	781
<i>Saket Gurukar, Nikil Pancha, Andrew Zhai, Eric Kim, Samson Hu, Srinivasan Parthasarathy, Charles Rosenberg, Jure Leskovec</i>	
TokenJoin: Efficient Filtering for Set Similarity Join with Maximum Weighted Bipartite Matching.....	790
<i>Alexandros Zeakis, Dimitrios Skoutas, Dimitris Sacharidis, Odysseas Papapetrou, Manolis Koubarakis</i>	
Quasi-stable Coloring for Graph Compression: Approximating Max-Flow, Linear Programs, and Centrality .	803
<i>Moe Kayali, Dan Suciu</i>	
Multi-Analyst Differential Privacy for Online Query Answering.....	816
<i>David Pujol, Albert Sun, Brandon Fain, Ashwin Machanavajhala</i>	
Excalibur: A Virtual Machine for Adaptive Fine-grained JIT-Compiled Query Execution based on VOILA .....	829
<i>Tim Gubner, Peter Boncz</i>	
Differentially Oblivious Relational Database Operators.....	842
<i>Lianke Qin, Rajesh Jayaram, Elaine Shi, Zhao Song, Danyang Zhuo, Shumo Chu</i>	
Keep CALM and CRDT On.....	856
<i>Shadaj Laddad, Conor Power, Mae Milano, Alvin Cheung, Natacha Crooks, Joseph M. Hellerstein</i>	
MQH: Locality Sensitive Hashing on Multi-level Quantization Errors for Point-to-Hyperplane Distances .....	864
<i>Kejing Lu, Yoshiharu Ishikawa, Chuan Xiao</i>	
The LDBC Social Network Benchmark: Business Intelligence Workload.....	877
<i>Gábor Szárnyas, Jack Waudby, Benjamin A. Steer, Dávid Szakállas, Altan Birler, Mingxi Wu, Yuchen Zhang, Peter Boncz</i>	
Making Cache Monotonic and Consistent.....	891
<i>Shuai An, Yang Cao</i>	
SkinnerMT: Parallelizing for Efficiency and Robustness in Adaptive Query Processing on Multicore Platforms .....	905
<i>Ziyun Wei, Immanuel Trummer</i>	
On Efficient Approximate Queries over Machine Learning Models.....	918
<i>Dujian Ding, Sihem Amer-Yahia, Laks Lakshmanan</i>	
Integrating Data Lake Tables .....	932
<i>Aamod Khatiwada, Roei Shraga, Wolfgang Gatterbauer, Renée J. Miller</i>	
PIM-tree: A Skew-resistant Index for Processing-in-Memory .....	946
<i>Hongbo Kang, Yiwei Zhao, Guy E. Blelloch, Laxman Dhulipala, Yan Gu, Charles Mcguffey, Phillip B. Gibbons</i>	
Web Record Extraction with Invariants .....	959
<i>Zhijia Chen, Weiyi Meng, Eduard Dragut</i>	
A Deep Generative Model for Trajectory Modeling and Utilization .....	973
<i>Yong Wang, Guoliang Li, Kaiyu Li, Haitao Yuan</i>	
L2chain: Towards High-performance, Confidential and Secure Layer-2 Blockchain Solution for Decentralized Applications.....	986
<i>Zihuan Xu, Lei Chen</i>	

## 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 Zacharathou (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)  
 Weiguang 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 fourth issue of PVLDB (Proceedings of the VLDB) Volume 16. PVLDB presents 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.

This issue includes 33 papers, spanning the topics of *Data Mining and Analytics*; *Data Privacy and Security*; *Database Engines*; *Database Performance and Manageability*; *Distributed Database Systems*; *Graph and Network Data*; *Information Integration and Data Quality*; *Languages*; *Machine Learning, AI, and Databases*; *Novel Database Architectures*; *Specialized and Domain-Specific Data Management*; *Text and Semi-Structured Data*; and *User Interfaces* (these cover all top-level topics of interest defined by our Call for Contributions, except *Provenance and Workflows*). The most popular topics in this issue are: *Database Engines* (9 papers); *Machine Learning, AI, and Databases* (8); *Data Privacy and Security* (7); and *Information Integration and Data Quality* (7).

Out of the 33 papers, two were accepted with shepherding, 25 were accepted after revision, and six were accepted after revision plus shepherding. Two papers are in the scalable data science category, two are in the experiment, analysis & benchmark category, two are in the vision category, and the rest are regular research papers.

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.

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.

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