



Proceedings of the VLDB Endowment

Volume 18, No. 11 – July 2025

Editors in Chief:

Themis Palpanas and Nesime Tatbul

Associate Editors:

Walid G. Aref, Manos Athanassoulis, Carsten Binnig, Spyros Blanas, Matthias Boehm, Angela Bonifati, K. Selcuk Candan, Lei Cao, Raul Castro Fernandez, Lei Chen, Shimin Chen, Yi Chen, Reynold Cheng, Alvin Cheung, Sudipto Das, Niv Dayan, Antonis Deligiannakis, Jens Dittrich, Xin Luna Dong, Karima Echihiabi, Alan Fekete, Avrilia Floratou, Jana Giceva, Katja Hose, H. V. Jagadish, Panos Kalnis, Georgia Koutrika, Eric Lo, Nikos Mamoulis, Stefan Manegold, Ioana Manolescu, Norman May, Umar Farooq Minhas, Fatemeh Nargesian, Beng Chin Ooi, Fatma Ozcan, Tamer Ozsu, Tilmann Rabl, Mirek Riedewald, Jennie Rogers, Alkis Simitsis, Letizia Tanca, Nan Tang, Yuanyuan Tian, Yongxin Tong, Pinar Tozun, Yannis Velegrakis, Matthias Weidlich, Steven E. Whang, Raymond Chi-Wing Wong

Publication Editors:

Xiaoou Ding, Subhadeep Sarkar, Giovanni Simonini

PVLDB – Proceedings of the VLDB Endowment

Volume 18, No. 11, July 2025.

All papers published in this issue will be presented at the 51st International Conference on Very Large Data Bases, London, United Kingdom, 2025.

Copyright 2025 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 18, Number 11, July 2025

Pages i – xi and 3695-4762

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. 18	vii

Research Papers

Doctopus: Budget-aware Structural Table Extraction from Unstructured Documents.....	3695
Chengliang Chai, Jiajun Li, Yuhao Deng, Yuanhao Zhong, Ye Yuan, Guoren Wang, Lei Cao	
S ³ AND: Efficient Subgraph Similarity Search Under Aggregated Neighbor Difference Semantics.....	3708
Qi Wen, Yutong Ye, Xiang Lian, Mingsong Chen	
Lighter-X: An Efficient and Plug-and-play Strategy for Graph-based Recommendation through Decoupled Propagation.....	3721
Yanping Zheng, Zhewei Wei, Frank De Hoo, Xu Chen, Hongteng Xu, Yuhang Ye, Jiadeng Huang	
Not Small Enough? SegPQ: A Learned Approach to Compress Product Quantization Codebooks.....	3730
Qiyu Liu, Yanlin Qi, Siyuan Han, Jingshu Peng, Jin Li, Lei Chen	
The Accuracy of Cardinality Estimators: Unraveling the Evaluation Result Conundrum	3744
Nazanin Rashedi, Guido Moerkotte	
LogLite: Lightweight Plug-and-Play Streaming Log Compression.....	3757
Benzhao Tang, Shiyu Yang, Zhitao Shen, Wenjie Zhang, Xuemin Lin, Zhihong Tian	
Enabling Efficient Attack Investigation via Human-in-the-Loop Security Analysis	3771
Saimon Tsegai, Xinyu Yang, Haoyuan Liu, Peng Gao	
Shifting Transaction Isolation on Graphs: From Systems to Data.....	3784
Wenzhi Fu, Yang Cao	
Fast Graph Vector Search via Hardware Acceleration and Delayed-Synchronization Traversal	3797
Wenqi Jiang, Hang Hu, Torsten Hoeﬂer, Gustavo Alonso	
Fremer: Lightweight and Effective Frequency Transformer for Workload Forecasting in Cloud Services	3812
Hengyu Ye, Jiadong Chen, Xiao He, Fuxin Jiang, Tieying Zhang, Jianjun Chen, Xiaofeng Gao	
TabulaX: Leveraging Large Language Models for Multi-Class Table Transformations	3826
Arash Dargahi Nobari, Davood Rafiei	
Bonspiel: Low Tail Latency Transactions in Geo-Distributed Databases	3840
Fan Cui, Eric Lo, Srijan Srivastava, Ziliang Lai	
Efficient Graph Data Access for Out-of-Memory GPU Streaming Graph Processing.....	3854
Qiang Wang, Yongze Yan, Hongshi Tan, Cheng Chen, Cheng Zhao, Jiaming Tian, Jiaxin Jiang, Xiaoliang Cong, Yanfeng Zhang, Ge Yu, Weng-Fai Wong, Bingsheng He	
Extensible and Robust Evaluation of Similarity Queries	3868
Daniel Schmitt, Thomas Hütter, Nikolaus Augsten	

PBench: Workload Synthesizer with Real Statistics for Cloud Analytics Benchmarking	3883
Yan Zhou, Chunwei Liu, Bhuvan Urgaonkar, Zhengle Wang, Magnus Mueller, Chao Zhang, Songyue Zhang, Pascal Pfeil, Dominik Horn, Zhengchun Liu, Davide Pagano, Tim Kraska, Samuel Madden, Ju Fan	
Accelerating Subgraph Matching through Fine-grained and Powerful Equivalences	3896
Yujie Lu, Zhijie Zhang, Weiguo Zheng, Lei Zou	
How to Optimize SQL Queries? A Comparison Between Split, Holistic, and Hybrid Approaches	3910
Luca Gretscher, Jens Dittrich	
Diva: Dynamic Range Filter for Var-Length Keys and Queries	3923
Navid Eslami, Ioana Bercea, Niv Dayan	
Approximate 2-hop neighborhoods on incremental graphs: An efficient lazy approach	3937
Luca Becchetti, Andrea Clementi, Luciano Gualà, Luca Pepè Sciarria, Alessandro Straziota, Matteo Stromieri	
Cracking Vector Search Indexes.....	3951
Vasilis Mageirakos, Bowen Wu, Gustavo Alonso	
DobLIX: A Dual-Objective Learned Index for Log-Structured Merge Trees	3965
Alireza Heidari, Amirhossein Ahmadi, Wei Zhang	
CoLA: Model Collaboration for Log-based Anomaly Detection	3979
Xuhang Zhu, Xiu Tang, Sai Wu, Jichen Li, Haobo Wang, Chang Yao, Quanqing Xu, Gang Chen	
Towards Designing Future-Proof Data Processing Systems.....	3988
Michael Jungmair, Jana Giceva	
Advancing Fact Attribution for Query Answering: Aggregate Queries and Novel Algorithms	3996
Omer Abramovich, Daniel Deutch, Nave Frost, Ahmet Kara, Dan Olteanu	
What If: Causal Analysis with Graph Databases.....	4009
Amedeo Pachera, Mattia Palmiotto, Angela Bonifati, Andrea Mauri	
AnyBlox: A Framework for Self-Decoding Datasets.....	4017
Mateusz Gienieccko, Maximilian Kuschewski, Thomas Neumann, Viktor Leis, Jana Giceva	
Federated and Balanced Clustering for High-dimensional Data.....	4032
Yushuai Ji, Shengkun Zhu, Shixun Huang, Zepeng Liu, Sheng Wang, Zhiyong Peng	
Relational Data Models for Genetic VCF data	4045
Mohamed Sabri Hafidi, Ozan Kahramanoğulları, Anton Dignös, Johann Gamper	
BURST: Rendering Clustering Techniques Suitable for Evolving Streams.....	4054
Apostolos Giannoulidis, Anastasios Gounaris, John Paparrizos	
Environmental Footprints of Query Processing: A Vision for Sustainable Database Architectures.....	4064
Michail Bachras, Hans-Arno Jacobsen	
Semantic Integrity Constraints: Declarative Guardrails for AI-Augmented Data Processing Systems.....	4073
Alexander Lee, Justin Chan, Michael Fu, Nicolas Kim, Akshay Mehta, Deepti Raghavan, Ugur Cetintemel	
RICH: Real-time Identification of negative Cycles for High-efficiency Arbitrage.....	4081

Bingqiao Luo, Jiaxin Jiang, Yuhang Chen, Junyi Hou, Cheng Jun Tey, Ziyang Qiu, Bingsheng He, Spencer Xiao, Dominic Ong, Wee Howe Ang

Balancing Privacy and Utility in Correlated Data: A Study of Bayesian Differential Privacy 4090
Martin Lange, Patricia Guerra-Balboa, Javier Parra-Arnau, Thorsten Strufe

Enhancing Transaction Processing through Indirection Skipping 4104
Riki Otaki, Jun Hyuk Chang, Aaron Elmore, Goetz Graefe

UniClean: A Scalable Data Cleaning Solution for Mixed Errors based on Unified Cleaners and Optimized Cleaning Workflow 4117
Xiaoou Ding, Zekai Qian, Hongzhi Wang, Siying Chen, Yafeng Tang, Hongbin Su, Huan Hu, Chen Wang

ShaRP: Explaining Rankings and Preferences with Shapley Values 4131
Venetia Pliatsika, Joao Fonseca, Kateryna Akhynko, Ivan Shevchenko, Julia Stoyanovich

SQLStorm: Taking Database Benchmarking into the LLM Era 4144
Tobias Schmidt, Leis Viktor, Peter Boncz, Thomas Neumann

Suna: Scalable Causal Confounder Discovery over Relational Data 4158
Jiaxiang Liu, Siyuan Xia, Daniel Alabi, Eugene Wu

Semantic Operators and Their Optimization: Towards AI-Based Data Analytics with Accuracy Guarantees 4171
Liana Patel, Siddharth Jha, Melissa Pan, Harshit Gupta, Parth Asawa, Carlos Guestrin, Matei Zaharia

Improving DBMS Scheduling Decisions with Accurate Performance Prediction on Concurrent Queries 4185
Ziniu Wu, Markos Markakis, Chunwei Liu, Peter Chen, Balakrishnan Narayanaswamy, Tim Kraska, Samuel Madden

Practical and Accurate Local Edge Differentially Private Graph Algorithms 4199
Pranay Mundra, Charalampos Papamanthou, Julian Shun, Quanquan Liu

Continuous Publication of Weighted Graphs with Local Differential Privacy 4214
Wen Xu, Pengpeng Qiao, Shang Liu, Zhirun Zheng, Yang Cao, Zhetao Li

TxnSails: Achieving Serializable Transaction Scheduling with Self-Adaptive Isolation Level Selection... 4227
Qiyu Zhuang, Wei Lu, Shuang Liu, Yuxing Chen, Xinyue Shi, Zhanhao Zhao, Yipeng Sun, Anqun Pan, Xiaoyong Du

No Cap, This Memory Slaps: Breaking Through the Memory Wall of Transactional Database Systems with Processing-in-Memory 4241
Hyoungjoo Kim, Yiwei Zhao, Andrew Pavlo, Phillip B. Gibbons

GraphCSR: A Degree-Equalized CSR Format for Large-scale Graph Processing 4255
Xinbiao Gan, Tiejun Li, Chunye Gong, Dongsheng Li, Dezun Dong, Jie Liu, Kai Lu

Effective and Efficient Attributed Hypergraph Embedding on Nodes and Hyperedges 4269
Yiran Li, Gongyao Guo, Chen Feng, Jieming Shi

Subgraph Matching: A New Decomposition Based Approach 4282
Qiyang Li, Jeffrey Yu, Zongyan He

SSD-iq: Uncovering the Hidden Side of SSD Performance 4295
Gabriel Haas, Bohyun Lee, Philippe Bonnet, Viktor Leis

Faster Convergence in Mini-batch Graph Neural Networks Training with Pseudo Full Neighborhood Compensation.....	4309
Qiqi Zhou, Yanyan Shen, Lei Chen	
TreeCat: Standalone Catalog Engine for Large Data Systems.....	4323
Keonwoo Oh, Pooja Nilangekar, Amol Deshpande	
Select Edges Wisely: Monotonic Path Aware Graph Layout Optimization for Disk-based ANN Search....	4337
Ziyang Yue, Bolong Zheng, Ling Xu, Kanru Xu, Shuhao Zhang, Yajuan Du, Yunjun Gao, Xiaofang Zhou, Christian Jensen	
Powerful GPUs or Fast Interconnects: Analyzing Relational Workloads on Modern GPUs.....	4350
Marko Kabić, Bowen Wu, Jonas Dann, Gustavo Alonso	
TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection.....	4364
Qinghua Liu, Seunghak Lee, Paparrizos John	
Time-Series Clustering: A Comprehensive Study of Data Mining, Machine Learning, and Deep Learning Methods.....	4380
John Paparrizos, Sai Prasanna Teja Reddy Bogireddy	
Beyond Compression: A Comprehensive Evaluation of Lossless Floating-Point Compression.....	4396
Kaisei Hishida, Chunwei Liu, John Paparrizos, Aaron Elmore	
ThriftLLM: On Cost-Effective Selection of Large Language Models for Classification Queries.....	4410
Keke Huang, Yimin Shi, Dujian Ding, Yifei Li, Yang Fei, Laks Lakshmanan, Xiaokui Xiao	
Sphinx: A Succinct Perfect Hash Index for x86.....	4424
Sajad Faghfoor Maghrebi, Niv Dayan	
NaviX: A Native Vector Index Design for Graph DBMSs With Robust Predicate-Agnostic Search Performance.....	4438
Gaurav Sehgal, Semih Salihoglu	
DIM-SUM: Dynamic IMputation for Smart Utility Management.....	4451
Ryan Hildebrant, Rahul Bhope, Sharad Mehrotra, Christopher Tull, Nalini Venkatasubramanian	
Robust Recursive Query Parallelism in Graph Database Management Systems.....	4465
Anurag Chakraborty, Semih Salihoglu	
OasisDB: An Oblivious and Scalable System for Relational Data.....	4478
Haseeb Ahmed, Nachiket Rao, Abdelkarim Kati, Florian Kerschbaum, Sujayya Maiyya	
CEDAR: A System for Cost-Efficient Data-Driven Claim Verification.....	4492
Tharushi Jayasekara, Immanuel Trummer	
Benchmarking Adaptive Multidimensional Indices.....	4505
Konstantinos Lampropoulos, Fatemeh Zardbani, Nikos Mamoulis, Panagiotis Karras	
Scaling GPU-Accelerated Databases beyond GPU Memory Size.....	4518
Yinan Li, Bailu Ding, Ziyun Wei, Lukas Maas, Momin Al-Ghosien, Spyros Blanas, Nicolas Bruno, Carlo Curino, Matteo Interlandi, Craig Peeper, Kaushik Rajan, Surajit Chaudhuri, Johannes Gehrke	
PAR2QO: Parametric Penalty-Aware Robust Query Optimization.....	4532

Haibo Xiu, Yang Li, Qianyu Yang, Pankaj Agarwal, Jun Yang

LIMAO: A Framework for Lifelong Modular Learned Query Optimization..... 4546
Qihan Zhang, Shaolin Xie, Ibrahim Sabek

QUEST: Query Optimization in Unstructured Document Analysis 4560
Zhaoze Sun, Chengliang Chai, Qiyang Deng, Kaisen Jin, Xinyu Guo, Han Han, Ye Yuan, Guoren Wang, Lei Cao

CENTS: A Flexible and Cost-Effective Framework for LLM-Based Table Understanding..... 4574
Guorui Xiao, Dong He, Jin Wang, Magdalena Balazinska

OmniMatch: Joinability Discovery in Data Products..... 4588
Christos Koutras, Jiani Zhang, Xiao Qin, Chuan Lei, Vassilis Ioannidis, Christos Faloutsos, George Karypis, Asterios Katsifodimos

Pistis: A Decentralized Knowledge Graph Platform Enabling Ownership-Preserving SPARQL Querying..... 4602
Enyuan Zhou, Song Guo, Zicong Hong, Christian Jensen, Yang Xiao, Jinwen Liang, Dalin Zhang

Selective Late Materialization in Modern Analytical Databases 4616
Yihao Liu, Shaoxuan Tang, Yulong Hui, Hangrui Zhou, Huanchen Zhang

The FastLanes File Format 4629
Azim Afroozeh, Peter Boncz

POLARIS: An Interactive and Scalable Data Infrastructure for Polar Science 4644
Yuchuan Huang, Ana Elena Uribe, Kareem Eldahshoury, Youssef Hussein, Grant Ogren, Mohamed Mokbel

Efficiently Joining Large Relations on Multi-GPU Systems 4653
Tobias Maltener, Ilin Tolovski, Tilmann Rabl

Stress-Testing ML Pipelines with Adversarial Data Corruption 4668
Jiongli Zhu, Geyang Xu, Felipe Lorenzi, Boris Glavic, Babak Salimi

PrivAGM: Secure Construction of Differentially Private Directed Attributed Graph Models on Decentralized Social Graphs..... 4682
Songlei Wang, Yifeng Zheng, Xiaohua Jia, Haibo Hu

OmniSQL: Synthesizing High-quality Text-to-SQL Data at Scale..... 4695
Haoyang Li, Shang Wu, Xiaokang Zhang, Xinmei Huang, Jing Zhang, Fuxin Jiang, Shuai Wang, Tieying Zhang, Jianjun Chen, Rui Shi, Hong Chen, Cuiping Li

Turbocharging Vector Databases using Modern SSDs..... 4710
Joobo Shim, Jaewon Oh, Hongchan Roh, Jaeyoung Do, Sang-Won Lee

SIEVE: Effective Filtered Vector Search with Collection of Indexes 4723
Zhaoheng Li, Silu Huang, Wei Ding, Yongjoo Park, Jianjun Chen

Enhancing Graph Edit Distance Computation: Stronger and Orientation-based ILP Formulations..... 4737
Andrea D'ascenzo, Julian Meffert, Petra Mutzel, Fabrizio Rossi

ParSEval: Plan-aware Test Database Generation for SQL Equivalence Evaluation 4750
Chunyu Chen, Zhengjie Miao, Yong Zhang, Jiannan Wang

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 18

Editors in Chief of PVLDB

Themis Palpanas (University Paris Cite)
Nesime Tatbul (Intel Labs and MIT)

Associate Editors of PVLDB

Walid G. Aref (Purdue University)
Manos Athanassoulis (Boston University)
Carsten Binnig (Technical University of Darmstadt)
Spyros Blanas (Ohio State University)
Matthias Boehm (Technical University of Berlin)
Angela Bonifati (University of Lille)
K. Selcuk Candan (Arizona State University)
Lei Cao (University of Arizona)
Raul Castro Fernandez (University of Chicago)
Lei Chen (Hong Kong University of Science and Technology)
Shimin Chen (Chinese Academy of Sciences)
Yi Chen (New Jersey Institute of Technology)
Reynold Cheng (University of Hong Kong)
Alvin Cheung, University of California (Berkeley)
Sudipto Das (Amazon Web Services)
Niv Dayan (University of Toronto)
Antonis Deligiannakis (Technical University of Crete)
Jens Dittrich (Saarland University)
Xin Luna Dong (Meta)
Karima Echiabi (Mohammed VI Polytechnic University)
Alan Fekete (University of Sydney)
Avrilia Floratou (Microsoft)
Jana Giceva (Technical University of Munich)
Katja Hose (Technical University of Vienna)
H. V. Jagadish (University of Michigan)
Panos Kalnis (King Abdullah University of Science and Technology)
Georgia Koutrika (Athena Research Center)
Eric Lo (Chinese University of Hong Kong)
Nikos Mamoulis (University of Ioannina)
Stefan Manegold (CWI)
Ioana Manolescu (Inria and Polytechnic Institute of Paris)
Norman May (SAP SE)
Umar Farooq Minhas (Apple)
Fatemeh Nargesian (University of Rochester)
Beng Chin Ooi (National University of Singapore)
Fatma Ozcan (Google)
Tamer Ozsu (University of Waterloo)
Tilman Rabl (Hasso Plattner Institute and University of Potsdam)
Mirek Riedewald (Northeastern University)
Jennie Rogers (Northwestern University)
Alkis Simitsis (Athena Research Center)
Letizia Tanca (Polytechnic University of Milan)
Nan Tang (Hong Kong University of Science and Technology (GZ))
Yuanyuan Tian (Microsoft)
Yongxin Tong (Beihang University)
Pinar Tozun (IT University of Copenhagen)

Yannis Velegrakis (Utrecht University)
Matthias Weidlich (Humboldt University of Berlin)
Steven E. Whang (Korea Advanced Institute of Science and Technology)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)

Publication Editors

Xiaoou Ding (Harbin Institute of Technology)
Subhadeep Sarkar (Brandeis University)
Giovanni Simonini (University of Modena and Reggio Emilia)

PVLDB Managing Editor

Jun Yang (Duke University)

PVLDB Advisory Board

Sourav S. Bhowmick (Nanyang Technological University)
Vanessa Braganholo (Universidade Federal Fluminense)
Lei Chen (Hong Kong University of Science and Technology)
Yanlei Diao (Ecole Polytechnique)
Xin Luna Dong (Meta)
Torsten Grust (University of Tuebingen)
Wolfgang Lehner (TU Dresden)
Alexandra Meliou (University of Massachusetts Amherst)
Felix Naumann (HPI)
Fatma Ozcan (Google)
Themis Palpanas (Universite Paris Cite)
Divesh Srivastava (AT&T Labs - Research)
Nesime Tatbul (Intel Labs and MIT)
Xiaokui Xiao (National University of Singapore)
Meihui Zhang (Beijing Institute of Technology)

Review Board

Ahmed S. Abdelhamid (Purdue University)
Ziawasch Abedjan (TU Berlin)
Ahmed Aly (Google)
Mohammad Javad Amiri (Stony Brook University)
Yael Amsterdamer (Bar-Ilan University)
Renzo Angles (Universidad de Talca)
Alexander Artikis (University of Piraeus)
Joy Arulraj (Georgia Tech)
Abolfazl Asudeh (University of Illinois Chicago)
Maurizio Atzori (University of Cagliari)
Nikolaus Augsten (University of Salzburg)
Zhifeng Bao (MIT University)
Ilaria Bartolini (University of Bologna)
Johes Bater (Tufts University)
Lawrence Benson (HPI and University of Potsdam)
Sonia Bergamaschi (University of Modena and Reggio Emilia)
Anna Bernasconi (Politecnico di Milano)
Arnab Bhattacharya (IIT Kanpur)
Alexander Boehm (SAP SE)
Paul Boniol (Universite de Paris)
Renata Borovica-Gajic (University of Melbourne)
Panagiotis Bouras (Johannes Gutenberg University Mainz)
Vanessa Braganholo (Fluminense Federal University)
Matteo Brucato (Microsoft Research)
Michael J. Cahill (University of Sydney)
Diego Calvanese (Free University of Bozen Bolzano)
Jesus Camacho-Rodriguez (Microsoft)
Helena Caminal (Google)
Huiping Cao (New Mexico State University)
Yang Cao (University of Edinburgh)
Zhao Cao (Huawei Technologies)
Zhichao Cao (Arizona State University)
Matteo Ceccarello (University of Padova)
Chengliang Chai (Beijing Institute of Technology)
Yunpeng Chai (Renmin University of China)
Harry Kai-Ho Chan (The University of Sheffield)
Tsz Nam Chan (Shenzhen University)
Subarna Chatterjee (Harvard University)
Cindy Chen (University of Massachusetts Lowell)
Lu Chen (Zhejiang University)
Hong Cheng (The Chinese University of Hong Kong)
Rada Chirkova (NC State University)
Theodoros Chondrogiannis (University of Konstanz)
Shihabur Chowdhury (Apple)
George Christodoulou (TU Delft)
Periklis Chrysogelos (Oracle)
Gao Cong (Nanyang Technological University)
Alex Conway (Cornell Tech)
Andrew Crotty (Northwestern University)
Bin Cui (Peking University)
Patrick Damme (TU Berlin)
Roshan Dathathri (Microsoft Research)
Jesse Davis (MongoDB)
Gagatay Demiralp (MIT)
Dong Deng (Rutgers University New Brunswick)
Laxman Dhulipala (University of Maryland, College Park)

Shimin Di (The Hong Kong University of Science and Technology)
Claudia Diamantini (Universita Politecnica delle Marche)
Anton Dignos (Free University of Bozen Bolzano)
Bailu Ding (Microsoft Research)
Bolin Ding (Alibaba Group)
Jialin Ding (Amazon Web Services)
Anh Dinh (Deakin University)
AnHai Doan (University of Wisconsin Madison)
Christos Doulkeridis (University of Piraeus)
Stefania Dumbrava (ENSIIE)
Ahmed Eldawy (University of California Riverside)
Mohamed Eltabakh (Qatar Foundation)
Venkatesh Emani (Microsoft)
Ju Fan (Renmin University of China)
Zhiwei Fan (Meta)
Yixiang Fang (The Chinese University of Hong Kong)
Anna Fariha (University of Utah)
Ziqiang Feng (Google)
Hakan Ferhatosmanoglu (University of Warwick and Amazon Web Services)
Elena Ferrari (University of Insubria)
Donatella Firmani (Sapienza University)
Peter M. Fischer (University of Augsburg)
George Fletcher (Eindhoven University of Technology)
Juliana Freire (New York University)
Sainyam Galhotra (Cornell University)
Johann Gamper (Free University of Bozen Bolzano)
Yunjun Gao (Zhejiang University)
Paolo Garza (Politecnico di Torino)
Chang Ge (University of Minnesota)
Tingjian Ge (University of Massachusetts Lowell)
Rainer Gemulla (Universitat Mannheim)
Nikos Giatrakos (Technical University of Crete)
Aristides Gionis (KTH Royal Institute of Technology)
Boris Glavic (Illinois Institute of Technology)
Lukasz Golab (University of Waterloo)
Jonathan Goldstein (Microsoft)
Sven Groppe (Universitat zu Lubeck)
Michael Grossniklaus (University of Konstanz)
Anja Gruenheid (Microsoft)
Le Gruenwald (The University of Oklahoma)
Vincenzo Gulisano (Chalmers University of Technology)
Rihan Hai (TU Delft)
Wook-Shin Han (POSTECH)
Mohamed S. Hassan (Oracle)
Oktie Hassanzadeh (IBM Research)
Wenjia He (University of Michigan)
Xi He (University of Waterloo)
Yeye He (Microsoft Research)
Meichun Hsu (Oracle)
Haibo Hu (The Hong Kong Polytechnic University)
Xiao Hu (University of Waterloo)
Qiang Huang (National University of Singapore)
Xin Huang (Hong Kong Baptist University)
Yan Huang (University of North Texas)
Zi Helen Huang (University of Queensland)
Madelon Hulsebos (University of California Berkeley)
Romain Ilbert (Huawei Paris Research Center)
Matteo Interlandi (Microsoft)

Ekaterini Ioanou (Tilburg University)
 Gabriela Jacques-Silva (Facebook)
 Fuad Jamour (Amazon Web Services)
 Soren Keiser Jensen (Aalborg University)
 Peiquan Jin (University of Science and Technology of China)
 Alekh Jindal (SmartApps)
 Hyungsoo Jung (Seoul National University)
 Vasiliki Kalavri (Boston University)
 Vana Kalogeraki (Athens University of Economics and Business)
 Eser Kandogan (Megagon Labs)
 Daniel Kang (UIUC)
 Zoi Kaoudi (IT University of Copenhagen)
 Pinar Karagoz (Middle East Technical University (METU))
 Bojan Karlas (Harvard University)
 Asterios Katsifodimos (TU Delft)
 Oliver A. Kennedy (University at Buffalo SUNY)
 Arijit Khan (Aalborg University)
 Guy Khazma (University of Toronto)
 Haridimos Kondylakis (FORTH-ICS)
 Arnd Christian Konig (Microsoft)
 Chrysanthi Kosyfaki (The University of Hong Kong)
 Nick Koudas (University of Toronto)
 Paraschos Koutris (University of Wisconsin Madison)
 Mayuresh Kunjir (Amazon Web Services)
 Alexandros Labrinidis (University of Pittsburgh)
 Wolfgang Lehner (TU Dresden)
 Chuan Lei (Amazon Web Services)
 Viktor Leis (TU Munich)
 Alberto Lerner (University of Fribourg)
 Ulf Leser (Humboldt-Universitat zu Berlin)
 Guoliang Li (Tsinghua University)
 Jia Li (The Hong Kong University of Science and Technology (GZ))
 Jianxin Li (Deakin University)
 Tian Li (Carnegie Mellon University)
 Tianyu Li (MIT)
 Yinan Li (Microsoft Research)
 Yuchen Li (Singapore Management University)
 Xiang Lian (Kent State University)
 Shen Liang (Universite Paris Cite)
 Michele Linardi (CYU)
 Matteo Lissandrini (University of Verona)
 Chunwei Liu (MIT)
 Jinfei Liu (Zhejiang University)
 Xueli Liu (Tianjin University)
 Cheng Long (Nanyang Technological University)
 Baotong Lu (Microsoft Research)
 Jiaheng Lu (University of Helsinki)
 Siqiang Luo (Nanyang Technological University)
 Yuyu Luo (The Hong Kong University of Science and Technology (GZ))
 Manisha Luthra (TU Darmstadt)
 Joana M. F. da Trindade (MIT)
 Chenhao Ma (The Chinese University of Hong Kong)
 Lin Ma (University of Michigan)
 Amr Magdy (University of California Riverside)
 Ahmed Mahmood (Google)
 Sujaya Maiyya (University of Waterloo)
 Neha Makhija (Northeastern University)
 Silviu Maniu (Universite Grenoble Alpes)
 Essam Mansour (Concordia University)
 Ryan Marcus (University of Pennsylvania)
 Amelie Marian (Rutgers University)
 Davide Martinenghi (Politecnico di Milano)
 Venkata Vamsikrishna Meduri (IBM Research - Almaden)
 Sharad Mehrotra (University of California Irvine)
 Alexandra Meliou (University of Massachusetts Amherst)
 Paolo Merialdo (Universita degli Studi Roma Tre)
 Amine Mhedhbi (Polytechnique Montreal)
 Xiaoye Miao (Zhejiang University)
 Sebastian Michel (RPTU Kaiserslautern Landau)
 Katsiaryna Mirylenka (IBM Research Zurich)
 Madhulika Mohanty (Inria Saclay)
 Mohamed Mokbel (University of Minnesota Twin Cities)
 Mirella M. Moro (Universidade Federal de Minas Gerais)
 Davide Mottin (Aarhus University)
 Kyriakos Mouratidis (Singapore Management University)
 Ingo Müller (Google)
 Balakrishnan Narayanaswamy (Amazon)
 Mario Nascimento (Northeastern University)
 Parimarjan Negi (MIT)
 Quoc Viet Hung Nguyen (Griffith University)
 Milos Nikolic (University of Edinburgh)
 Matthaïos Olma (MongoDB)
 Prashant Pandey (University of Utah)
 George Papadakis (University of Athens)
 Dimitris Papadias (The Hong Kong University of Science and Technology)
 Odysseas Papapetrou (TU Eindhoven)
 John Paparrizos (The Ohio State University)
 George Papastefanatos (ATHENA Research Center)
 Stefano Paraboschi (Universita degli Studi di Bergamo)
 Aditya Parameswaran (University of California Berkeley)
 Yongjoo Park (UIUC)
 Eliana Pastor (Politecnico di Torino)
 Jignesh Patel (Carnegie Mellon University)
 Marco Patella (University of Bologna)
 Torben Bach Pedersen (Aalborg University)
 Botao Peng (Chinese Academy of Sciences)
 Peng Peng (Hunan University)
 Matthew J. Perron (MIT)
 Ilia Petrov (Reutlingen University)
 Holger Pirk (Imperial College)
 Stefan Plantikow (Neo4j)
 Orestis Polychroniou (Amazon)
 Danica Porobic (Oracle)
 Abdulhakim Qahtan (Utrecht University)
 Abdul Quamar (Google)
 Weixiong Rao (Tongji University)
 Berthold Reinwald (IBM Research Almaden)
 El Kindi Rezig (MIT)
 Daniel Ritter (SAP)
 Oscar Romero (Universitat Politecnica de Catalunya)
 Kexin Rong (Georgia Institute of Technology)
 Abhishek Roy (Snowflake)

Florin Rusu (University of California Merced)
 Sourav S. Bhowmick (Nanyang Technological University)
 Ibrahim Sabek (University of Southern California)
 Mohammad Sadoghi (University of California Davis)
 Semih Salihoglu (University of Waterloo)
 Maria Luisa Sapino (University of Torino)
 Subhadeep Sarkar (Brandeis University)
 Kai-Uwe Sattler (TU Ilmenau)
 Patrick Schafer (Humboldt-Universitat zu Berlin)
 Felix M. Schuhknecht (Johannes Gutenberg University Mainz)
 Maximilian E. Schule (University of Bamberg)
 Malte Schwarzkopf (Brown University)
 Rathijit Sen (Microsoft)
 Jiwon Seo (Seoul National University)
 Juan Sequeda (data.world)
 Marco Serafini (University of Massachusetts Amherst)
 Amir Shaikhha (University of Edinburgh)
 Shantanu Sharma (New Jersey Institute of Technology)
 Yanyan Shen (Shanghai Jiao Tong University)
 Jieming Shi (The Hong Kong Polytechnic University)
 Roei Shraga (WPI)
 Tarique Siddiqui (Microsoft Research)
 Giovanni Simonini (University of Modena and Reggio Emilia)
 Utku Sirin (Harvard University)
 Spiros Skiadopoulos (University of the Peloponnese)
 Dimitrios Skoutas (Athena Research Center)
 Shaoxu Song (Tsinghua University)
 Divesh Srivastava (AT&T Chief Data Office)
 Chrysoula Stathakopoulou (Chainlink Labs)
 Kostas Stefanidis (Tampere University)
 Kurt Stockinger (ZHAW Zurich University of Applied Sciences)
 Uta Storl (University of Hagen)
 Shixuan Sun (Shanghai Jiao Tong University)
 Ki Hyun Tae (KAIST)
 Dixin Tang (University of Texas Austin)
 Jing Tang (The Hong Kong University of Science and Technology (GZ))
 Mingjie Tang (Sichuan University)
 Bo Tang (Southern University of Science and Technology)
 Egemen Tanin (University of Melbourne)
 Ernest Teniente (Universitat Politcnica de Catalunya)
 Arash Termehchy (Oregon State University)
 Jens Teubner (TU Dortmund)
 Riccardo Torlone (Roma Tre University)
 Goce Trajcevski (Iowa State University)
 Immanuel Trummer (Cornell University)
 Eleni Tzirita Zacharitou (IT University of Copenhagen)
 Katerina Tzompanaki (CY Cergy Paris University)
 Leong Hou U (University of Macau)
 Alexander van Renen (UTN)
 Genoveva Vargas-Solar (CNRS LIRIS)
 Nalini Venkatasubramanian (University of California Irvine)
 Hannes Voigt (Neo4j)
 Hongzhi Wang (Harbin Institute of Technology)
 Ning Wang (Beijing Jiaotong University)
 Qitong Wang (Universite Paris Cite)
 Sibo Wang (The Chinese University of Hong Kong)
 Tianzheng Wang (Simon Fraser University)
 Yifan Wang (University of Florida)
 Sai Wu (Zhejiang University)
 Yinghui Wu (Case Western Reserve University)
 Yuncheng Wu (Renmin University of China)
 Xiaokui Xiao (National University of Singapore)
 Jianliang Xu (Hong Kong Baptist University)
 Jianqiu Xu (Nanjing University of Aeronautics and Astronautics)
 Nikolay Yakovets (TU Eindhoven)
 Xiao Yan (Centre for Perceptual and Interactive Intelligence (CPII))
 Hongzhi Yin (The University of Queensland)
 Man Lung Yiu (The Hong Kong Polytechnic University)
 Brit Youngmann (Technion)
 Jeffrey Xu Yu (The Chinese University of Hong Kong)
 Xiaohui Yu (York University)
 Yi Yu (NII)
 Ye Yuan (Beijing Institute of Technology)
 Cong Yue (National University of Singapore)
 Demetrios Zeinalipour-Yazti (University of Cyprus)
 Yuxiang Zeng (Beihang University)
 Steffen Zeuch (TU Berlin)
 Chao Zhang (University of Waterloo)
 Chen Zhang (The Hong Kong Polytechnic University)
 Huanchen Zhang (Tsinghua University)
 Meihui Zhang (Beijing Institute of Technology)
 Minjia Zhang (Microsoft AI and Research)
 Qizhen Zhang (University of Toronto)
 Xiaofei Zhang (University of Memphis)
 Yanfeng Zhang (Northeastern University)
 Bo Zhao (Aalto University)
 Zhuoyue Zhao (University at Buffalo)
 Bolong Zheng (Huazhong University of Science and Technology)
 Kaiping Zheng (National University of Singapore)
 Jingren Zhou (Alibaba Group)
 Xuan Zhou (East China Normal University)
 Yongluan Zhou (University of Copenhagen)
 Yiwen Zhu (Microsoft)
 Jia Zou (Arizona State University)
 Lei Zou (Peking University)
 Kostas Zoumpatianos (Snowflake)
 Andreas Zufle (Emory University)

LETTER FROM THE EDITORS IN CHIEF

We are pleased to present the eleventh (July 2025) issue of PVLDB Volume 18, which contains 81 papers from all 4 categories: Vision (4), Scalable Data Science (6), Experiment, Analysis & Benchmark (10), and Regular Research (61).

These papers cover a diverse collection of primary subject areas, including: Database Engines (19); Data Mining and Analytics (8); Database Performance and Manageability (2); Data Privacy and Security (4); Distributed Database Systems (1); Graph and Network Data (11); Information Integration and Data Quality (9); Languages (2); Machine Learning, AI, and Databases (6); Novel Database Architectures (10); Provenance and Workflows (1); Specialized and Domain-Specific Data Management (3); Text and Semi-Structured Data (2); and User Interfaces (3).

We thank our review board as well as proceedings chairs for their hard work in the creation of another high-quality PVLDB issue.

This is the largest issue for PVLDB Volume 18 to date, including the last batch of accepted research papers that will be presented at VLDB 2025 – The 51st International Conference on Very Large Databases to take place in London, UK in September.

It is with great pleasure that we welcome the data management community to **VLDB 2025 in London, UK**. The last time VLDB was hosted in the UK was in Edinburgh in 1999. Now, 26 years later, we have the privilege of bringing VLDB to London for the very first time. London is a truly global city, rich in history, diverse in culture, and vibrant with innovation, making it a perfectly fitting location for the exchange of ideas and advances that shape our field.

VLDB 2025 will take place at an iconic location in the heart of London: attendees will have the chance to explore both VLDB's cutting-edge research offerings and one of the world's most dynamic cities. We are sincerely grateful to everyone who has contributed to organizing this year's conference, from the programme and organizing committees to the many volunteers whose efforts make such an event possible. We look forward to a stimulating, memorable, and inspiring edition of VLDB. It is our privilege to share with you what we think will be a special moment in the history of this flagship conference.

Peter Pietzuch and Peter Triantafillou
General Chairs for VLDB 2025

Themis Palpanas and Nesime Tatbul
Editors-in-Chief of PVLDB Vol. 18
Program Chairs for VLDB 2025