



Proceedings of the VLDB Endowment

Volume 19, No. 11 – July 2026

Editors in Chief:

Yanlei Diao and Xiaokui Xiao

Associate Editors:

Alkis Polyzotis, Alkis Simitsis, Andrew Pavlo, Angela Bonifati, Anh Dinh, Antonis Deligiannakis, Ashraf Aboulnaga, Badrish Chandramouli, Bin Cui, Bogdan Cautis, Ce Zhang, Elena Ferrari, Evaggelia Pitoura, Fatemeh Nargesian, Guoliang Li, Hong Cheng, Immanuel Trummer, Jana Giceva, Jennie Rogers, Jian Pei, Jianliang Xu, Karima Echihabi, Katja Hose, Khuzaima Daudjee, Kunsoo Park, Kyuseok Shim, Laks Lakshmanan, Lei Cao, Lei Chen, Li Xiong, Mauro Sozio, Meihui Zhang, Melanie Herschel, Michael Böhlen, Nan Tang, Niv Dayan, Norman May, Raul Castro Fernandez, Raymond Chi-Wing Wong, Reynold Cheng, Ruben Mayer, Semih Salihoglu, Shimin Chen, Sibor Wang, Sourav S Bhowmick, Spyros Blanas, Steven Whang, Sudipto Das, Umar Farooq Minhas, Wei Wang, Wook-Shin Han, Yannis Velegrakis, Yanyan Shen, Yi Chen, Yuncheng Wu, Zhifeng Bao, Zi Huang

Publication Editors:

Andrea Mauri, Giovanni Simonini, Siqiang Luo, Stefan Halfpap, Subhadeep Sarkar

PVLDB – Proceedings of the VLDB Endowment

Volume 19, No. 11, July 2026.

All papers published in this issue will be presented at the 52nd International Conference on Very Large Data Bases, Boston, MA, USA, 2026.

Copyright 2026 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 19, Number 11, July 2026

Pages i – xii and 2936 - 3833

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. 19	vi

Research Papers

EcoTable: Cost-effective Table Integration in Data Lakes for Natural Language Queries	2936
<i>Yuhui Wang, Jinqi Liu, Chengliang Chai, Hangyu Zhao, Yuhao Deng, Yuyu Luo, Xin Tang, Ye Yuan, Guoren Wang, Fengjin Wang, Lei Cao</i>	
Unified Static–Dynamic Pruning for Efficient LLM Inference.....	2950
<i>Jinhyeok Kim, Yejoon Lee, Jaeyoung Do</i>	
Craw: A Unified and Efficient Querying Framework for Large-Scale Video Datasets.....	2964
<i>Ziqi Zhou, Hanjian Jiang, Zihao Zeng, Xupuzhe Shao, Zichen Xu</i>	
CrocSort: Resource-Efficient, Skew-Resilient Parallel External Merge Sort.....	2978
<i>Riki Otaki, Charles Benello, Fuheng Zhao, Aaron Elmore, Goetz Graefe</i>	
Bolt-on, Verifiable Provenance for LLM-Powered Data Processing	2992
<i>Yiming Lin, Sepanta Zeighami, Aditya Parameswaran</i>	
The Data World Is Not Flat: Efficient Factorized Execution for Relational Systems	3006
<i>Stefan Lehner, Thomas Neumann</i>	
I-Rex: An Interactive Debugger for SQL	3020
<i>Yihao Hu, Zian Chen, Zhiming Leong, Sharan Sokhi, Zachary Zheng, Alex Chao, Kristin Stephens-Martinez, Sudeepa Roy, Jun Yang</i>	
Rethinking Learned Index and LSM-tree Integration	3034
<i>Guangxun Zhao, Yongjie Zhu, Charles Jaranilla, Seehwan Yoo, Jongmoo Choi</i>	
Featurized-Decomposition Join: Low-Cost Semantic Joins with Guarantees	3048
<i>Sepanta Zeighami, Shreya Shankar, Aditya Parameswaran</i>	
Multi-Objective Agentic Rewrites for Unstructured Data Processing.....	3062
<i>Lindsey Wei, Shreya Shankar, Sepanta Zeighami, Yeounoh Chung, Fatma Ozcan, Aditya Parameswaran</i>	
Remora: Scale-out Deterministic Execution for Smart Contracts.....	3076
<i>Zhengqing Liu, Alberto Sonnino, Igor Zablotchi, Eleftherios Kokoris Kogias, Marios Kogias</i>	
QBAT: Model-based Query Budget Autotuner for Clustering-based Approximate Nearest Neighbor Search	3091
<i>Jonghyun Bae, Tae Jun Ham, Alan Li, Supawit Chockchawat, Yannis Papakonstantinou</i>	
Error-bounded Point Cloud Compression Using Truncated Octahedron Quantization.....	3105
<i>Youyuan Liu, Longtao Zhang, Ruoyu Li, Bo Jiang, Taolue Yang, Kai Zhao, Sheng Di, Eduard Dragut, Sian Jin</i>	
Noisy Interactive Graph Search: An Uncertainty-Based Approach with Online Modeling of Latent Expertise and Difficulty.....	3119

Han Linghu, Qianhao Cong, Liang Feng, Lei Chen, Jing Tang

GAS: A Lightweight Framework for Filtered Search over Wide-table Vectors 3133
Ziyan He, Yuxiang Wang, Yu Sun, Zijie Ma, Hui Li, Qian Tao, Yu Li, Yongxin Tong

Fast and Private Max-Sum Diversification..... 3146
Ron Zadacario, Tova Milo

Auto-Fill: Learning to Predict Missing Values Accurately with Specialist Language Models 3160
Yurong Liu, Yeye He, Haoyu Dong, Junjie Xing, Shi Han, Dongmei Zhang, Surajit Chaudhuri

Worst-Case Optimal BGPs on Temporal Graphs 3174
Diego Arroyuelo, Aidan Hogan, Gonzalo Navarro, Juan Reutter

PipeLens: Identifying Interventions for Resolving Malfunctioning Data Science Pipelines 3188
Jahid Hasan, Stanley Jiang, Tejendra Singh, Sainyam Galhotra, Romila Pradhan, Divesh Srivastava

How Much Can RocksDB Chew? Achieving Near-Zero Write Stalls with Sustainable RocksDB 3202
Hojin Shin, Yongmin Lee, Seehwan Yoo, Jongmoo Choi

Interoperable ACID Transactions for Open Table Formats..... 3216
Tobias Goetz, Daniel Ritter, Muhammad El-Hindi, Jana Giceva

SEMA: A High-performance System for LLM-based Semantic Query Processing..... 3231
Kangkang Qi, Dongyang Xie, Wenbo Li, Hao Zhang, Yuanyuan Zhu, Jeffrey Xu Yu, Kangfei Zhao

PAIL: Efficient kNN Search on Set-Valued Attributes 3245
Daniel Schmitt, Thomas Hütter, Nikolaus Augsten

Maximum Defective Biclique Search in Large Bipartite Graphs..... 3259
Donghang Cui, Ronghua Li, Qiangqiang Dai, Hongchao Qin, Guoren Wang

UniTG: A Unified System for Efficient and Seamless Textual Graph Learning..... 3273
Meng Zhang, Zhisheng Ye, Qiyu Liu, Jingshu Peng, Tianwei Zhang

Fugue: Online Elasticity for Distributed Stateful Stream Processing..... 3288
Yuqiu Zhang, Yunhao Mao, Hans-Arno Jacobsen

Fides: Secure and Scalable Asynchronous DAG Consensus via Trusted Components..... 3302
Shaokang Xie, Dakai Kang, Hanzheng Lyu, Jianyu Niu, Mohammad Sadoghi

Efficient and Secure Range Counting over Distributed Geographic Data with Query Range Protection .. 3317
Haoxin Yang, Pinghui Wang, Zhe Hou, Tian Zhou, Guangmingzi Yang, Zehua Lei, Rundong Li, Yutong Song, Yongyuan Peng, Fangming Dong, Xiaohong Guan

Local Shapley: Model-Induced Locality and Optimal Reuse in Data Valuation..... 3330
Xuan Yang, Hsi-Wen Chen, Ming-Syan Chen, Jian Pei

Nav-Index: A High-Performance, Adaptive Index for Shortest Path Queries in RDBMS..... 3344
Maximilian Reif, Thomas Neumann

CONDA: A Connectivity-Aware Dynamic Index for Approximate Nearest Neighbor Search over Evolving Data 3357
Darae Lee, Min-Soo Kim

DeepPrep: An LLM-Powered Agentic System for Autonomous Data Preparation	3371
<i>Meihao Fan, Ju Fan, Yuxin Zhang, Shaolei Zhang, Xiaoyong Du, Jie Song, Peng Li, Fuxin Jiang, Tieying Zhang, Jianjun Chen</i>	
KAFY: An Extensible and Scalable Transformers-Based System for Trajectory Data Analysis	3385
<i>Youssef Hussein, Mohamed Mokbel</i>	
SciTables : A Dataset and Evaluation Framework for Complex Table-to-Text Generation	3400
<i>Mehrnoush Alizade, Tengrui Kong, Suman Kalyan Maity</i>	
GPU-Accelerated ANNS: Quantized for Speed, Built for Change	3413
<i>Hunter Mccoy, Zikun Wang, Prashant Pandey</i>	
Efficient GPU-Accelerated Adaptive Minimum Cost Seed Selection	3427
<i>Gongyao Guo, Chen Feng, Yiran Li, Jieming Shi</i>	
Detecting Data-Type-Related Logic Bugs in Relational DBMSs via Compatible Database Construction...	3440
<i>Jiansen Song, Wensheng Dou, Yingying Zheng, Yu Gao, Quanqing Xu, Ziyu Cui, Xudong Xie, Hongtao Zhou, Jiaying Zhou, Jun Wei, Wei Wang</i>	
GPU-Native Approximate Nearest Neighbor Search with IVF-RaBitQ: Fast Index Build and Search	3454
<i>Jifan Shi, Jianyang Gao, James Xia, Tamás Béla Fehér, Cheng Long</i>	
BBC: Improving Large- k Approximate Nearest Neighbor Search with a Bucket-based Result Collector .	3468
<i>Ziqi Yin, Gao Cong, Kai Zeng, Jinwei Zhu, Bin Cui</i>	
Resource-Efficient FirmCore Decomposition on Billion-scale Multilayer Graphs	3483
<i>Cheng Huang, Davide Mottin, Ira Assent</i>	
Breaking Structural Isolation: Scalable Graph Clustering via Community-Aware Sampling and Structural Entropy	3497
<i>Jingyun Zhang, Hao Peng, Jianxin Li, Angsheng Li, Philip S. Yu</i>	
ConRAD: Conformal Risk-Aware Neural Databases	3511
<i>Sonia Horchidan, Fabian Zeiher, Xiangyu Shi, Vasiliki Kalavri, Henrik Boström, Ioannis Kontoyiannis, Paris Carbone</i>	
Window Function Optimization: Co-Evaluation and Other Techniques.....	3525
<i>Daniel Lindner, Felix Naumann, Alberto Lerner</i>	
Can we trust LLM Self-Explanations for Entity Resolution?	3538
<i>Tommaso Teofili, Donatella Firmani, Nick Koudas, Paolo Merialdo, Divesh Srivastava</i>	
Computing Why-Provenance for Property Graph Queries	3552
<i>Koumudi Ganepola, Maxime Jakubowski, Katja Hose</i>	
Orca: Flexible Quorums Meet Dynamic Quorums.....	3565
<i>Kevin Dharmawan, Prajwal Annigeri, Mohammad Amiri</i>	
One Pass to Parse Them All: Fused Parallel CSV Processing.....	3579
<i>Simon Ellmann, Thomas Neumann</i>	
Efficient GPU-Accelerated Local Subgraph Counting	3592
<i>Qiao He, Yiran Li, Man Lung Yiu, Jieming Shi</i>	

QDBO: A Real-time Quantum-augmented Database System Optimizer	3606
<i>Hanwen Liu, Abhishek Kumar, Federico Spedalieri, Ibrahim Sabek</i>	
Storing and Indexing Multiple Tables by Interesting Orderings: For Efficient Joins, Groupings, and Updates in Relational Databases	3621
<i>Wenhui Lyu, Goetz Graefe</i>	
Accelerating String-Heavy Queries with LLM Token Tables	3635
<i>Tobias Schmidt, Nicolas Schmitt, Thomas Neumann, Andreas Kipf</i>	
MFTune: An Efficient Multi-fidelity Framework for Spark SQL Configuration Tuning.....	3649
<i>Beicheng Xu, Lingching Tung, Yuchen Wang, Yupeng Lu, Bin Cui</i>	
CGIF: Combining Proximity Graphs and Inverted Files for Efficient Filtered Vector Search over Arbitrary Predicates.....	3663
<i>Jiarui Luo, Chaoji Zuo, Dong Deng</i>	
Rhyme Native: Efficient Code Generation for Structured and Semi-Structured Workloads.....	3676
<i>Ran Guo, Tiark Rompf</i>	
RAGPerf: An End-to-End Benchmarking Framework for Retrieval-Augmented Generation Systems.....	3689
<i>Shaobo Li, Yirui Zhou, Yuan Xu, Kevin Chen, Daniel Waddington, Swaminathan Sundararaman, Hubertus Franke, Jian Huang</i>	
Matryoshka: Uncovering Relevant Features in Data Lakes to Enhance Machine Learning Applications .	3704
<i>Fedor Turchenko, Runjie Zhang, Binger Chen, Matthias Boehm, Babak Salimi, Amir Shaikhha, Ziawasch Abedjan</i>	
Dial: A Knowledge-Grounded Dialect-Specific NL2SQL System	3718
<i>Xiang Zhang, Le Zhou, Hongming Xu, Wei Zhou, Xuanhe Zhou, Guoliang Li, Yuyu Luo, Changdong Liu, Guorun Chen, Jiang Liao, Fan Wu</i>	
Developing and Benchmarking Verification Algorithms to Improve Text-to-SQL Generation.....	3732
<i>Tarfah Alrashed, Madhup Sukoon, David Karger, Natasha Noy</i>	
MosaicJoin: Compact Semantic Sketches for Value-Level Join Discovery	3745
<i>Grace Fan, Eden Wu, Majid Daliri, Juliana Freire</i>	
Bespoke OLAP: Synthesizing Workload-Specific One-size-fits-one Database Engines	3759
<i>Johannes Wehrstein, Timo Eckmann, Matthias Jasny, Carsten Binnig</i>	
Benchmarking the Full Pipeline of Materialized-View-Based Query Rewriting	3772
<i>Xinjie Hu, Zhengjie Miao</i>	
Tidehunter: Large-Value Storage With Minimal Data Relocation	3786
<i>Andrey Chursin, Lefteris Kokoris-Kogias, Alex Orlov, Alberto Sonnino, Igor Zablotchi</i>	
stratum: A System Infrastructure for Massive Agent-Centric ML Workloads.....	3799
<i>Arnab Phani, Elias Strauss, Sebastian Schelter</i>	
Lower-Bound Distance Queries under Partial Information	3807
<i>Swastik Biswas, Sohrab Namazi Nia, Jeess Augustine, Suraj Shetiya, Senjuti Basu Roy, Gautam Das</i>	
ANNiE: A Learned Query Cost Estimator for Graph-Based Approximate Nearest Neighbor Search	3820
<i>Zeyu Wang, Manos Chatzakis, Qitong Wang, Themis Palpanas, Peng Wang, Wei Wang</i>	

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 19

Editors in Chief of PVLDB

Yanlei Diao (Ecole Polytechnique and Amazon Web Services)
Xiaokui Xiao (National University of Singapore)

Associate Editors of PVLDB

Alkis Polyzotis (Databricks)
Alkis Simitsis (Athena Research Center)
Andrew Pavlo (Carnegie Mellon University)
Angela Bonifati (Univ. of Lyon)
Anh Dinh (Deakin University)
Antonis Deligiannakis (Technical University of Crete)
Ashraf Aboulnaga (University of Texas at Arlington)
Badrish Chandramouli (Microsoft Research)
Bin Cui (Peking University)
Bogdan Cautis (University of Paris-Saclay)
Ce Zhang (University of Chicago)
Elena Ferrari (University of Insubria, Varese)
Evaggelia Pitoura (Univ. of Ioannina)
Fatemeh Nargesian (University of Rochester)
Guoliang Li (Tsinghua University)
Hong Cheng (Chinese University of Hong Kong)
Immanuel Trummer (Cornell University)
Jana Giceva (TU Munich)
Jennie Rogers (Northwestern University)
Jian Pei (Duke University)
Jianliang Xu (Hong Kong Baptist University)
Karima Echiabi (Mohammed VI Polytechnic University)
Katja Hose (TU Wien)
Khuzaima Daudjee (University of Waterloo)
Kunsoo Park (Seoul National University)
Kyuseok Shim (Seoul National University)
Laks Lakshmanan (The University of British Columbia)
Lei Cao (University of Arizona)
Lei Chen (Hong Kong University of Science and Technology)
Li Xiong (Emory University)
Mauro Sozio (Luiss University of Rome)
Meihui Zhang (Beijing Institute of Technology)
Melanie Herschel (Nanyang Technological University)
Michael Böhlen (University of Zurich)
Nan Tang (HKUST (GZ))
Niv Dayan (University of Toronto)
Norman May (SAP SE)
Raul Castro Fernandez (The University of Chicago)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
Reynold Cheng (The University of Hong Kong, China)
Ruben Mayer (University of Bayreuth)
Semih Salihoglu (University of Waterloo)
Shimin Chen (Chinese Academy of Sciences)
Sibo Wang (The Chinese University of Hong Kong)
Sourav S Bhowmick (Nanyang Technological University)
Spyros Blanas (The Ohio State University)
Steven Whang (KAIST)

Sudipto Das (Amazon Web Services)
Umar Farooq Minhas (Apple)
Wei Wang (ByteDance)
Wook-Shin Han (POSTECH)
Yannis Velegrakis (Utrecht University)
Yanyan Shen (Shanghai Jiao Tong University)
Yi Chen (NJIT)
Yuncheng Wu (Renmin University of China)
Zhifeng Bao (The University of Queensland)
Zi Huang (University of Queensland)

Publication Editors

Andrea Mauri (Lyon 1 University)
Giovanni Simonini (University of Modena and Reggio Emilia)
Siqiang Luo (Nanyang Technological University)
Stefan Halfpap (TU Berlin)
Subhadeep Sarkar (Brandeis University)

PVLDB Managing Editor

Jun Yang (Duke University)

PVLDB Advisory Board

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

Review Board

Abhishek Roy (Microsoft)
Ahmed Aly (Google)
Ahmed Eldawy (University of California, Riverside)
Ahmed Mahmood (Google)
Alberto Lerner (Computing Flows GmbH)
Alberto Sonnino (MystenLabs & University College London)
Alexander Artikis (University of Piraeus)
Alexander Böhm (SAP SE)
Alexander Krause (TU Dresden)
Amarnadh Sai Eluri (Google)
Amelie Marian (Rutgers University)
Amine Mhedhbi (Polytechnique Montréal)
Amir Shaikhha (University of Edinburgh)
Amr Magdy (University of California Riverside)
Andrea Mauri (Lyon 1 University)
Andreas Züfle (Emory University)
Antonio Boffa (EPFL)
Arash Termehchy (Oregon State University)
Arijit Khan (Bowling Green State University)
Aristides Gionis (KTH Royal Institute of Technology)
Arkaprava Saha (University of Tsukuba)
Arnab Phani (TU Berlin)
Arnd Christian König (Microsoft)
Avinash Kumar (Google)
Baihua Zheng (Singapore Management University)
Berthold Reinwald (IBM Research - Almaden)
Bin Yang (Aalborg University)
Bo Tang (Southern University of Science and Technology)
Bo Zhao (Aalto University)
Bobbi W Yogatama (NVIDIA)
Bojan Karlaš (Harvard University)
Boris Glavic (University of Illinois Chicago)
Brian Kroth (Microsoft)
Brit Youngmann (Technion)
Calisto Zuzarte (IBM)
Carlo Curino (Microsoft)
Chang Ge (University of Minnesota)
Chao Zhang (University of Waterloo)
Chen Zhang (The Hong Kong Polytechnic University)
Cheng Long (Nanyang Technological University)
Chenghao Lyu (Amazon)
Chengkai Li (The University of Texas at Arlington)
Chengliang Chai (Beijing Institute of Technology)
Chenhao Ma (The Chinese University of Hong Kong, Shenzhen)
Christan Grant (University of Florida)
Christos Doukeridis (University of Piraeus)
Christos Koutras (New York University)
Chrysanthi Kosyfaki (The University of Hong Kong)
Chuan Lei (Oracle)
Chunwei Liu (MIT CSAIL)
Cindy Chen (University of Massachusetts Lowell)
Cong Yue (National University of Singapore)
Daniel Kang (UIUC)
Daniel Ritter (SAP)
Davood Rafiei (University of Alberta)
Demetris Zeinalipour (University of Cyprus)
Dimitrios Skoutas (Athena Research Center)
Dixin Tang (The University of Texas, Austin)
Domagoj Vrgoč (PUC)
Dominik Durner (CedarDB)
Donatella Firmani (Sapienza University)
Donato Tiano (Università degli Studi di Modena e Reggio Emilia)
Dong Deng (Alibaba Cloud Computing)
Dong Xie (Penn State University)
Dumitrel Loghin (OKX)
Eduard Dragut (Temple Univ.)
Egemen Tanin (University of Melbourne)
El Kindi Rezig (University of Utah)
Eleni Tzirita Zacharatou (Hasso Plattner Institute (HPI))
Eliaana Pastor (Politecnico di Torino)
Elisa Bertino (Purdue University)
Enzo Veltri (Università della Basilicata)
Ergute Bao (Mohamed bin Zayed University of Artificial Intelligence)
Eser Kandogan (Megagon Labs)
Essam Mansour (Concordia University)
Farnoush Banaei-Kashani (University of Colorado Denver)
Felix M Schuhknecht (Johannes Gutenberg University Mainz)
Florian Hahn (University of Twente)
Florin Rusu (UC Merced)
Francesco Gullo (University of L'Aquila)
Fuad Jamour (Amazon Web Services)
Gao Cong (Nanyang Technological University)
Garima Gaur (Inria Saclay, Ecole Polytechnique)
Gengrui Zhang (Concordia University)
George Christodoulou (TU Delft)
George Papadakis (University of Athens)
George Papastefanatos (ATHENA Research Center)
Georgios Siachamis (Inria)
Georgios R Theodorakis (Nvidia)
Giorgio Vinciguerra (Università di Pisa)
Giuseppina Andresini (University of Bari Aldo Moro)
Goce Trajcevski (Iowa State University)
Grigorios Loukides (King's College London)
Guido Moerkotte (University of Mannheim)
Guillaume Lachaud (Ecole Polytechnique)
Haibo Hu (Hong Kong Polytechnic University)
Haipeng Dai (Nanjing University)
Hannes Voigt (Neo4j)
Hantian Zhang (Google)
Hanzhi Wang (University of Copenhagen)
Haralampos Gavriilidis (BIFOLD & Technische Universität Berlin)
Haridimos Kondylakis (FORTH-ICS & Computer Science Department, University of Crete)
Harry Kai-Ho Chan (The University of Sheffield)
Hazar Harmouch (University of Amsterdam)
Herald Kllapi (Google Inc)
Hongzhi Wang (Harbin Institute of Technology)
Hua Lu (Aalborg University)
Huanchen Zhang (Tsinghua University)
Huiping Chen (University of Birmingham)
Hyungsoo Jung (Seoul National University)
Hyunjoon Kim (Hanyang University)
Ibrahim Sabek (University of Southern California)

Igor Zablotchi (Mysten Labs)
 Ilia Petrov (Reutlingen University)
 Ilias Azizi (Mohammed VI Polytechnic University)
 Ilie Sarpe (KTH Royal Institute of Technology)
 Indrakshi Ray (Colorado State University)
 Ioannis Demertzis (UCSC)
 Jaeyoung Do (Seoul National University)
 Jan Hidders (Birkbeck, University of London)
 Jeeta Ann Chacko (Technical University of Munich)
 Jeffrey Xu Yu (The Hong Kong University of Science and Technology (Guangzhou))
 Jelle Hellings (McMaster University)
 Jesús Camacho-Rodríguez (Microsoft)
 Jia Li (Hong Kong University of Science and Technology)
 Jia Zou (Arizona State University)
 Jialin Ding (Princeton University)
 Jianguo Wang (Purdue University)
 Jiannan Wang (Tsinghua University)
 Jianxin Li (Edith Cowan University)
 Jiawei Jiang (Wuhan University)
 Jieming Shi (The Hong Kong Polytechnic University)
 Jignesh Patel (Carnegie Mellon University)
 Jin Wang (Arizona State University)
 Jiwon Seo (Seoul National University)
 Johannes Pietrzyk (TU Dresden)
 Johes Bater (Tufts University)
 John Paparrizos (The Ohio State University, USA & Aristotle University of Thessaloniki, Greece)
 Jonathan Fürst (ZHAW Zurich University of Applied Sciences)
 Jonghyeok Park (Korea University)
 Joseph Near (University of Vermont)
 Ju Fan (Renmin University of China)
 Junghoon Kim (UNIST)
 Junhao Gan (University of Melbourne)
 K. Selçuk Candan (Arizona State University)
 Kaisong Huang (University of Calgary)
 Kai-Uwe Sattler (TU Ilmenau)
 Kanchan Chowdhury (Marquette University)
 Karim Benouaret (Univ Lyon)
 Katsiaryna Milylenka (Zalando SE)
 Kaustubh Beedkar (Indian Institute of Technology Delhi)
 Keke HUANG (The University of British Columbia)
 Kenneth A Ross (Columbia University)
 KI HYUN TAE (Samsung Research)
 Kostas Stefanidis (Tampere University)
 Kurt Stockinger (Zurich University of Applied Sciences)
 Kwanghyun Park (Yonsei University)
 Kyoungmin Kim (EPFL)
 Larissa Capobianco Shimomura (Hasselt University)
 Le Gruenwald (The University of Oklahoma)
 Lefteris Kokoris Kogias (MystenLabs)
 Lei Zou (Peking University)
 Leong Hou U (University of Macau)
 Leopoldo Bertossi (Carleton University)
 Liang Liang (EPFL)
 Lingyang Chu (McMaster University)
 Loredana Caruccio (University of Salerno)
 Lu Chen (Zhejiang University)
 Luca Gagliardelli (eCampus University)
 Lucas Braun (Oracle Labs)
 Lukasz Golab (University of Waterloo)
 Madhulika Mohanty (Inria & Institut Polytechnique de Paris)
 Magnus Mueller (AWS)
 Mahdi Esmailoghli (HU Berlin)
 Makoto Onizuka (Osaka University)
 Man Lung Yiu (Hong Kong Polytechnic University)
 Manuel Rigger (National University of Singapore)
 Marco Patella (University of Bologna)
 Markos Markakis (Massachusetts Institute of Technology)
 Martin Prammer (Carnegie Mellon University)
 Matteo Ceccarello (University of Padova)
 Matteo Interlandi (Microsoft)
 Matthaïos Olma (MongoDB)
 Matthew J Perron (Microsoft)
 Matthias Boehm (Technische Universität Berlin)
 Matthias Weidlich (Humboldt-Universität zu Berlin)
 Maximilian E Schüle (University of Bamberg)
 Meichun Hsu (Oracle America)
 Miao Qiao (The University of Auckland)
 Miao Yu (Meta)
 Michael J Cahill (University of Sydney)
 Michael Grossniklaus (University of Konstanz)
 Michael Gubanov (Florida State University)
 Michael J Mior (Rochester Institute of Technology)
 Mohamed Eltabakh (Qatar Foundation)
 Mohamed Mokbel (University of Minnesota - Twin Cities)
 Mohammad Sadoghi (University of California, Davis)
 Mohammad Javad Amiri (Stony Brook University)
 Mohammed J Saeed (Apple)
 Mostafa Milani (The University of Western Ontario)
 Nick Koudas (University of Toronto)
 Nikolay Yakovets (TU Eindhoven)
 Oktie Hassanzadeh (IBM Research)
 Oliver A Kennedy (University at Buffalo, SUNY)
 Oscar Romero (Universitat Politècnica de Catalunya)
 Panagiotis Bouros (Johannes Gutenberg University Mainz)
 Panagiotis Karras (University of Copenhagen)
 Paolo Garza (Politecnico di Torino)
 Paolo Merialdo (Università degli Studi Roma Tre)
 Paolo Missier (University of Birmingham)
 Paolo Papotti (EURECOM)
 Patrick Damme (Technische Universität Berlin)
 Patrick Schäfer (Humboldt-Universität zu Berlin)
 Paul Boniol (Inria, Ecole normale supérieure)
 Peizhi Wu (University of Pennsylvania)
 Peng Peng (Hunan University)
 Periklis Chrysogelos (Oracle)
 Peter M. Fischer (University of Augsburg)
 Pierangela Samarati (Università degli Studi di Milano)
 Prashant Pandey (Northeastern University)
 Protiva Rahman (University of Florida)
 Qingqing Ye (Hong Kong Polytechnic University)
 Qitong Wang (Harvard University)
 Qizhen Zhang (University of Toronto)
 Rada Chirkova (NC State University)
 Rainer Gemulla (Universität Mannheim)

Rana Alotaibi (King Abdulaziz City for Science and Technology (KACST))
 Rana Shahout (Harvard)
 Renata Borovica-Gajic (University of Melbourne)
 Renchi Yang (Hong Kong Baptist University)
 Ria Borrromeo (University of Philippines)
 Riccardo Torlone (Roma Tre University)
 Rihan Hai (TU Delft)
 Ritesh Sarkhel (Amazon)
 Roe Shraga (WPI)
 Romila Pradhan (Purdue University)
 Ruiyuan Li (Chongqing University)
 Ryan Marcus (University of Pennsylvania)
 Ryan McKenna (University of Massachusetts Amherst)
 Sai Wu (Zhejiang University)
 Sara Foresti (Universita' degli Studi di Milano)
 Sarah Kleest-Meißner (Hasselt University)
 Sebastian Link (University of Auckland)
 Sebastian Michel (RPTU Kaiserslautern-Landau)
 Sepanta Zeighami (University of California Berkeley)
 Shantanu Sharma (New Jersey Institute of Technology)
 Shaoxu Song (Tsinghua University)
 Sharad Mehrotra (U.C. Irvine)
 Shen Liang (Université Paris Cité)
 Shihabur Chowdhury (Apple)
 Shimin Di (The Hong Kong University of Science and Technology)
 Silviu Maniu (Université Grenoble Alpes)
 Siqiang Luo (Nanyang Technological University)
 Søren Keiser Jensen (Aalborg University)
 Soror Sahri (Université Paris Cité)
 Stavros Maroulis (ATHENA Research Center)
 Stefania Dumbrava (ENSIIE, INRIA, IRIF, Télécom-SudParis)
 Stefano Marchesin (Università di Padova)
 Stefano Paraboschi (Universita' degli Studi di Bergamo)
 Steffen Zeuch (TU Berlin)
 Surajit Chaudhuri (Microsoft)
 Suyash Gupta (University of Oregon)
 Sven Groppe (Universität zu Lübeck/Institute of Information Systems (IFIS))
 Tanzima Hashem (Bangladesh University of Engineering and Technology)
 Tarique Ashraf Siddiqui (Microsoft Corporation)
 Tenindra Abeywickrama (RIKEN Center for Computational Science)
 Theodoros Chondrogiannis (Norwegian University of Science and Technology (NTNU))
 Thomas Neumann (TUM)
 Tianyi Li (Aalborg University)
 Tianzheng Wang (Simon Fraser University)
 Tieying Zhang (Bytedance)
 Ting Yu (MBZUAI)
 Tingjian Ge (University of Massachusetts, Lowell)
 Tong Chen (The University of Queensland)
 Torben Bach Pedersen (Aalborg University)
 Tristan Allard (Univ Rennes, CNRS, IRISA)
 Tsz Nam Chan (Shenzhen University)
 Ulf Leser (Humboldt-Universität zu Berlin)
 Utku Sirin (Harvard University)
 Venkata Vamsikrishna Meduri (IBM Research-Almaden)
 Verena Kantere (University of Ottawa)
 Vikram Nathan (Amazon, Inc.)
 Viktor Leis (Technische Universität München)
 Vivek Narasayya (Microsoft)
 Vivek Shah (Samsung)
 Wang-Chien Lee (Pennsylvania State University, USA)
 Wei Dong (Nanyang Technological University)
 WEI LU (Renmin University of China)
 Weiguo Zheng (Fudan University)
 Wenqi FAN (The Hong Kong Polytechnic University)
 Wentao Wu (Microsoft Research)
 Wentao Zhang (Peking University)
 William Schultz (MongoDB)
 Xiang Lian (Kent State University)
 Xiangyao Yu (University of Wisconsin-Madison)
 Xiangyu Ke (Zhejiang University)
 Xiao Qin (Snowflake)
 Xiao Yan (Wuhan University)
 Xiaofei Zhang (University of Memphis)
 Xiaohui Yu (York University)
 Xiaou Ding (Harbin Institute of Technology)
 Xiaoying Wang (Microsoft)
 Xike Xie (University of Science and Technology of China)
 Xin Huang (Hong Kong Baptist University)
 Xuhao Chen (MSU)
 Xun Jian (Northwestern Polytechnical University)
 Xupeng Miao (Purdue University)
 Yael Amsterdamer (Bar-Ilan University)
 Yan Huang (University of North Texas)
 Yang Cao (Institute of Science Tokyo)
 Yang Wang (The Ohio State University)
 YAO LU (NUS)
 Yesdaulet Izenov (Nazarbayev University)
 Yeye He (Microsoft Research)
 Yifan Wang (University of Hawaii)
 Yihao Ang (National University of Singapore)
 Yin Yang (Hamad bin Khalifa University)
 Yinan Li (Microsoft Research)
 Yinghui Wu (Case Western Reserve University)
 Yingtai Xiao (TikTok)
 Yiwen Zhu (Microsoft)
 Yixiang Fang (School of Data Science, The Chinese University of Hong Kong, Shenzhen)
 Yongjoo Park (UIUC)
 Yongluan Zhou (University of Copenhagen)
 Yongxin Tong (Beihang University)
 Yu Yang (City University of Hong Kong)
 Yuan Qiu (Southeast University)
 Yujie Hui (Google)
 Yuke Wang (Rice University)
 Yuyu Luo (HKUST (GZ))
 Zeyu Ding (Binghamton University)
 Zhao Cao (Remin University of China)
 ZHAOJING LUO (Beijing Institute of Technology)
 Zhengjie Miao (Simon Fraser University)
 Zhewei Wei (Renmin University of China)
 ZHIWEI FAN (Meta)
 Zhongle Xie (Zhejiang University)
 Zhuoyue Zhao (University at Buffalo)
 Zoi Kaoudi (IT University of Copenhagen)

Light Load Review Board

Abdulkhakim Qahtan (Utrecht University)
Abolfazl Asudeh (University of Illinois Chicago)
Ahmed El-Roby (Carleton University)
Alberto Abelló (Universitat Politècnica de Catalunya)
Alex Conway (Cornell Tech)
Alexander Erben (Technical University of Munich)
Alexandros Labrinidis (University of Pittsburgh)
Anupam Sanghi (TU Darmstadt)
Ariful Azad (Texas A&M University)
Babak Salimi (University of California at San Diego)
Baotong Lu (Microsoft Research)
Bettina Kemme (McGill University)
C. Seshadhri (UCSC)
Curtis Dyreson (Utah State University)
Danfeng Zhang (Duke University)
Daniel Kocher (University of Salzburg)
Davide Mottin (Aarhus University)
Fabian Panse (University of Augsburg)
Felix Naumann (Hasso Plattner Institute, University of Potsdam)
Haitao Yuan (Nanyang Technological University)
Hangdong Zhao (Microsoft)
Hideyuki Kawashima (Keio University)
Hongyan Chang (NUS)
Hongzhi Yin (The University of Queensland)
Ingo Müller (Google)
Jiaheng Lu (University of Helsinki)
Jinho Lee (Seoul National University)
Juan Sequeda (data.world)
Katia Papakonstantinou (Athens University of Economics and Business)
Kexin Rong (Georgia Institute of Technology)
Konstantinos Mamouras (Rice University)
Kyriakos Mouratidis (Singapore Management University)
Lipeng Wan (Georgia State University)
Manolis Terrovitis (Athena Research Center)
Matteo Brucato (Microsoft Research)
Mayuresh Kunjir (Amazon AWS)
Michele Linardi (CYU)
Mo Sha (Alibaba Group)
Murat Demirbas (MongoDB Research)
Myeongjae Jeon (POSTECH)
Nikos Giatrakos (Technical University of Crete)
Odysseas Papapetrou (TU Eindhoven)
Panagiotis Liakos (Athens University of Economics and Business)
Panagiotis Liakos (University of Athens)
Panos Vassiliadis (University of Ioannina)
Paris Carbone (KTH Royal Institute of Technology)
Paul Groth (University of Amsterdam)
Pierre Bourhis (CNRS)
Quang-Trung Ta (National University of Singapore)
Rathijit Sen (Microsoft)
Remy Wang (University of California, Los Angeles)
Renzo Angles (Universidad de Talca)
Seokki Lee (University of Cincinnati)
Sergi Nadal (Universitat Politècnica de Catalunya)
Sivaprasad Sudhir (MIT)
Song Jiang (University of Texas, Arlington)
Sonia Bergamaschi (Università di Modena e Reggio Emilia)
Subarna Chatterjee (Datastax)
Subhadeep Sarkar (Brandeis University)
Sujaya Maiyya (University of Waterloo)
Supawit Chockchowwat (University of Illinois Urbana-Champaign)
Theodore Johnson (AT&T Labs - Research)
Thomas Fahringer (University of Innsbruck)
Thorsten Papenbrock (Philipps University of Marburg)
Vassilis J. Tsotras (UC Riverside)
Venkatesh Emani (Microsoft)
Wenjia He (University of Michigan)
Wolfgang Gatterbauer (Northeastern University)
Xingguang Chen (National University of Singapore)
Yogesh Simmhan (Indian Institute of Science (IISc), Bangalore)
Yuxin Tang (Rice University)
Zhe Jiang (University of Florida)
Ziawasch Abedjan (BIFOLD/TU Berlin)

LETTER FROM THE EDITORS IN CHIEF

We are pleased to present Issue 11 of Volume 19 of the Proceedings of the VLDB Endowment (PVLDB), comprising 65 papers that reflect the breadth and continued evolution of data-management research. The contributions advance the foundations of database systems while addressing emerging demands created by vector data, large language models, specialized hardware, distributed platforms, and increasingly complex data-processing applications.

By primary subject area, this issue includes 14 papers on Database Engines, 11 on Machine Learning, AI, and Databases, 6 on Graph and Network Data, and 5 each on Information Integration and Data Quality, Specialized and Domain-Specific Data Management, and Novel Database Architectures. It also features 4 papers on Data Privacy and Security; 3 each on Provenance and Workflows, Database Performance and Manageability, Distributed Database Systems, and Text and Semi-Structured Data; 2 on User Interfaces; and 1 on Data Mining and Analytics.

A substantial group of papers advances the foundations and performance of database engines. These contributions introduce factorized execution techniques for relational queries, resource-efficient parallel sorting, improved storage and indexing for large values and multiple tables, and optimizations for window functions and materialized-view-based query rewriting. Other works address sustainable write performance in RocksDB, efficient code generation across structured and semi-structured workloads, workload-specific synthesis of OLAP engines, and systematic detection of logic bugs in relational systems. Together, these papers demonstrate continued progress in making database engines faster, more adaptable, and more dependable across diverse workloads.

Vector search constitutes another prominent theme, spanning indexing, cost estimation, tuning, and hardware acceleration. The papers develop techniques for large-(k) nearest-neighbor search, filtered search over vectors with structured attributes, and efficient processing of arbitrary predicates. They also propose connectivity-aware indexes for evolving data, learned estimators for graph-based search costs, automated configuration tuning, and GPU-native methods that accelerate both index construction and query execution. Collectively, these contributions address the growing need for vector-search systems that remain efficient as datasets, filters, workloads, and underlying hardware evolve.

The continuing interaction between data management and AI is reflected throughout the issue. Several papers improve the execution of AI workloads through efficient LLM inference, high-performance semantic query processing, semantic joins, and infrastructure for large-scale agent-centric applications. Other contributions investigate LLM-assisted processing of unstructured data, specialist models for predicting missing values, textual graph learning, data valuation, and knowledge-grounded translation from natural language to dialect-specific SQL. Complementary work develops benchmarks and verification techniques for semantic query processing engines, retrieval-augmented generation, and text-to-SQL systems, highlighting the importance of evaluating not only model accuracy but also system performance, reliability, and deployability.

Closely related contributions address data integration, preparation, and discovery. These works use natural-language interfaces and agentic methods to integrate heterogeneous tables and automate data preparation, while other papers discover useful features across large data lakes. Research on semantic sketches supports value-level join discovery, and risk-aware neural database techniques provide calibrated answers when querying incomplete knowledge graphs. Together, these papers illustrate how modern data systems are moving beyond syntactic matching toward semantic, adaptive, and uncertainty-aware methods for making heterogeneous data usable.

Graph and network data form another important thread. The contributions include scalable decomposition and clustering of large and multilayer graphs, interactive search under uncertainty, defective-biclique discovery, shortest-path indexing, and worst-case optimal processing of temporal graph patterns. Other papers exploit GPUs for local subgraph counting and influence-based seed selection, support textual graph learning, and compute provenance for property-graph queries. These works jointly advance graph algorithms, query processing, explainability, and hardware-aware analytics for increasingly large and complex networks.

The issue also presents new approaches to distributed systems, emerging architectures, and specialized data processing. The papers support interoperable transactions across open table formats, online elasticity for stateful stream processing, and quorum protocols that adapt to changing system conditions. Further contributions explore quantum-augmented database optimization, transformer-based trajectory analytics, video querying, point-cloud compression, and efficient processing of CSV files and string-intensive workloads. These systems demonstrate how database research continues to expand toward new hardware, data modalities, and deployment environments.

Finally, several papers address security, privacy, provenance, and system trustworthiness. The contributions include deterministic execution for scalable smart contracts, secure and fair asynchronous consensus, privacy-preserving

diversification, and protected range counting over distributed geographic data. Other works provide verifiable provenance for LLM-powered data processing, interactive debugging of SQL, and methods for diagnosing and repairing malfunctioning data-science pipelines. Research on explanations for entity resolution and verification of text-to-SQL outputs further emphasizes the need to make automated data systems understandable, auditable, and reliable.

We thank the authors, reviewers, meta-reviewers, and members of the PVLDB editorial board whose work made this issue possible. We hope readers will find the breadth of ideas in this issue valuable and that these contributions will stimulate further research across both established and emerging areas of data management.

Sincerely,

Bogdan Cautis, Anh Dinh, Ferrari Elena, and Helen Huang
Associate Editors of PVLDB Vol. 19

Yanlei Diao and Xiaokui Xiao
Editors-in-Chief of PVLDB Vol. 19
Program Chairs for VLDB 2026