



# Proceedings of the VLDB Endowment

Volume 15, No. 3 – November 2021

Editors in Chief:

**Juliana Freire and Xuemin Lin**

Associate Editors:

**Arun Kumar, Azza Abouzied, Beng Chin Ooi, Boris Glavic, Dan Suciu,  
Divyakant Agrawal, Eugene Wu, Fatma Ozcan, 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. 3, November 2021.

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

## **Copyright 2021 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 15, Number 3, November 2021

Pages i – vii and 388 - 751

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

Enabling SQL-based Training Data Debugging for Federated Learning.....	388
<i>Yejia Liu, Weiyuan Wu, Lampros Flokas, Jiannan Wang, Eugene Wu</i>	
Leveraging Query Logs and Machine Learning for Parametric Query Optimization.....	401
<i>Kapil Vaidya, Anshuman Dutt, Vivek Narasayya, Surajit Chaudhuri</i>	
Pre-training Summarization Models of Structured Datasets for Cardinality Estimation.....	414
<i>Yao Lu, Srikanth Kandula, Arnd Christian König, Surajit Chaudhuri</i>	
xFraud: Explainable Fraud Transaction Detection.....	427
<i>Susie Xi Rao, Shuai Zhang, Zhichao Han, Zitao Zhang, Wei Min, Zhiyao Chen, Yinan Shan, Yang Zhao, Ce Zhang</i>	
Subgraph Matching over Graph Federation .....	437
<i>Ye Yuan, DeLong Ma, Zhenyu Wen, Zhiwei Zhang, Guoren Wang</i>	
Provenance-based Data Skipping .....	451
<i>Xing Niu, Boris Glavic, Ziyu Liu, Pengyuan Li, Dieter Gawlick, Vasudha Krishnaswamy, Zhen Hua Liu, Danica Porobic</i>	
Deep Transfer Learning for Multi-source Entity Linkage via Domain Adaptation .....	465
<i>Di Jin, Bunyamin Sisman, Hao Wei, Xin Luna Dong, Danai Koutra</i>	
An Experimental Evaluation and Investigation of Waves of Misery in R-trees.....	478
<i>Lu Xing, Eric Lee, Tong An, Bo-cheng Chu, Ahmed Mahmood, Ahmed M. Aly, Jianguo Wang, Walid G. Aref</i>	
PRUC : P-Regions with User-Defined Constraint .....	491
<i>Yongyi Liu, Ahmed R. Mahmood, Amr Magdy, Sergio Rey</i>	
Points-of-Interest Relationship Inference with Spatial-enriched Graph Neural Networks .....	504
<i>Yile Chen, Xiucheng Li, Gao Cong, Cheng Long, Zhifeng Bao, Shang Liu, Wanli Gu, Fuzheng Zhang</i>	
SAFE: A Share-and-Aggregate Bandwidth Exploration Framework for Kernel Density Visualization ...	513
<i>Tsz Nam Chan, Pak Lon Ip, Leong Hou U, Byron Choi, Jianliang Xu</i>	
The next 50 Years in Database Indexing or: The Case for Automatically Generated Index Structures .....	527
<i>Jens Dittrich, Joris Nix, Christian Schön</i>	
DARLING: Data-Aware Load Shedding in Complex Event Processing Systems .....	541
<i>Koral Chapnik, Ilya Kolchinsky, Assaf Schuster</i>	

Rearchitecting In-Memory Object Stores for Low Latency.....	555
<i>Danyang Zhuo, Kaiyuan Zhang, Zhuohan Li, Siyuan Zhuang, Stephanie Wang, Ang Chen, Ion Stoica</i>	
MT-Teql: Evaluating and Augmenting Neural NLIDB on Real-world Linguistic and Schema Variations .....	569
<i>Pingchuan Ma, Shuai Wang</i>	
Theoretically and Practically Efficient Parallel Nucleus Decomposition .....	583
<i>Jessica Shi, Laxman Dhulipala, Julian Shun</i>	
APEX: A High-Performance Learned Index on Persistent Memory.....	597
<i>Baotong Lu, Jialin Ding, Eric Lo, Umar Farooq Minhas, Tianzheng Wang</i>	
Unsupervised Time Series Outlier Detection with Diversity-Driven Convolutional Ensembles .....	611
<i>David Campos, Tung Kieu, Chenjuan Guo, Feiteng Huang, Kai Zheng, Bin Yang, Christian S. Jensen</i>	
Efficient and Effective Data Imputation with Influence Functions .....	624
<i>Xiaoye Miao, Yangyang Wu, Lu Chen, Yunjun Gao, Jun Wang, Jianwei Yin</i>	
Parallel Training of Knowledge Graph Embedding Models: A Comparison of Techniques.....	633
<i>Adrian Kochsiek, Rainer Gemulla</i>	
Detecting Layout Templates in Complex Multiregion Files.....	646
<i>Gerardo Vitagliano, Lan Jiang, Felix Naumann</i>	
What Is the Price for Joining Securely? Benchmarking Equi-Joins in Trusted Execution Environments .....	659
<i>Kajetan Maliszewski, Jorge-Arnulfo Quiane-Ruiz, Jonas Traub, Volker Markl</i>	
Efficient Temporal Pattern Mining in Big Time Series Using Mutual Information.....	673
<i>Van Long Ho, Nguyen Ho, Torben Bach Pedersen</i>	
Efficient Label-Constrained Shortest Path Queries on Road Networks: A Tree Decomposition Approach .....	686
<i>Junhua Zhang, Long Yuan, Wentao Li, Lu Qin, Ying Zhang</i>	
Ember: No-Code Context Enrichment via Similarity-Based Keyless Joins .....	699
<i>Sahaana Suri, Ihab F Ilyas, Christopher Re, Theodoros Rekatsinas</i>	
Incremental Partitioning for Efficient Spatial Data Analytics .....	713
<i>Tin Vu, Ahmed Eldawy, Vagelis Hristidis, Vassilis J. Tsotras</i>	
Lux: Always-on Visualization Recommendations for Exploratory Dataframe Workflows.....	727
<i>Doris Lee, Dixin Tang, Kunal Agarwal, Thyne Boonmark, Caitlyn Chen, Jake Kang, Ujjaini Mukhopadhyay, Jerry Song, Micah Yong, Marti A. Hearst, Aditya Parameswaran</i>	
Flexible Rule-Based Decomposition and Metadata Independence in Modin: A Parallel Dataframe System .....	739
<i>Devin Petersohn, Dixin Tang, Rehan S Durrani, Areg Melik-adamyam, Joseph Gonzalez, Anthony Joseph, Aditya Parameswaran</i>	

## **PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 15**

### **Editors in Chief of PVLDB**

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 Suci (University of Washington)  
Divyakant Agrawal (University of California, Santa Barbara)  
Eugene Wu (Columbia University)  
Fatma Ozcan (Google)  
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 Michigan)  
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 (Pennsylvania 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 Zacharitou (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 für 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 Ibáñez)  
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 Católica & IMFD)  
Maria Luisa Sapino (U. Torino)  
Matteo Lissandrini (Aalborg University)  
Matthias Boehm (Graz University of Technology)  
Matthias Renz (University of Kiel)  
Max Heimerl (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)  
 Sibow 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

The third issue of PVLDB Vol. 15 includes 28 articles covering different categories in the research track: 22 regular research papers, 3 Scalable Data Science papers, and 3 Experiments, Analyses & Benchmarks papers. These papers went through a thorough review process, and 27 of them went through a revision phase. We are very grateful to all AEs, reviewers, reproducibility chairs, publication editors, and authors who contributed to this issue.

This issue represents the breadth of problem areas our community is working on from core data database systems topics such as query optimization and indexing, to spatio-temporal analytics, entity linkage, and provenance management. We can also find many creative approaches that combine methods from different disciplines to address important problems within data management and beyond. We have seen an increasing number of submissions that propose the use of machine learning methods to address data management challenges, as well as data management methods to address machine learning challenges. Similarly, our community has been exploring the synergies between data management and visualization.

As part of our effort to ensure the transparency of papers published in PVLDB, authors are strongly encouraged to submit code and data used in their papers as supplementary materials. The PVLDB Reproducibility chairs check, for every accepted paper, whether the supplementary materials are available; these papers are awarded the ACM Artifacts Available badge (<https://www.acm.org/publications/policies/artifact-review-and-badging-current>). For issue 3, 71.4% of papers have received this badge.

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!

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