



# Proceedings of the VLDB Endowment

Volume 15, No. 2 – October 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. 2, October 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 2, October 2021

Pages i – vii and 141 - 387

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

(p,q)-biclique Counting and Enumeration for Large Sparse Bipartite Graphs.....	141
<i>Jianye Yang, Yun Peng, Wenjie Zhang</i>	
Evaluating Query Languages and Systems for High-Energy Physics Data .....	154
<i>Dan Graur, Ingo Müller, Mason Proffitt, Ghislain Fourny, Gordon T. Watts, Gustavo Alonso</i>	
Distributed Hop-Constrained s-t Simple Path Enumeration at Billion Scale.....	169
<i>Kongzhang Hao, Long Yuan, Wenjie Zhang</i>	
ETO: Accelerating Optimization of DNN Operators by High-Performance Tensor Program Reuse.....	183
<i>Jingzhi Fang, Yanyan Shen, Yue Wang, Lei Chen</i>	
Babelfish: Efficient Execution of Polyglot Queries.....	196
<i>Philipp Marian Grulich, Steffen Zeuch, Volker Markl</i>	
Butterfly Counting on Uncertain Bipartite Networks .....	211
<i>Alexander Zhou, Yue Wang, Lei Chen</i>	
METRO: A Generic Graph Neural Network Framework for Multivariate Time Series Forecasting.....	224
<i>Yue Cui, Kai Zheng, Dingshan Cui, Jiandong Xie, Liwei Deng, Feiteng Huang, Xiaofang Zhou</i>	
LargeEA: Aligning Entities for Large-scale Knowledge Graphs .....	237
<i>Congcong Ge, Xiaoze Liu, Lu Chen, Baihua Zheng, Yunjun Gao</i>	
HVS: Hierarchical Graph Structure Based on Voronoi Diagrams for Solving Approximate Nearest Neighbor Search.....	246
<i>Kejing Lu, Mineichi Kudo, Chuan Xiao, Yoshiharu Ishikawa</i>	
Origami: A High-Performance Mergesort Framework.....	259
<i>Arif Arman, Dmitri Loguinov</i>	
Learning to be a Statistician: Learned Estimator for Number of Distinct Values .....	272
<i>Renzhi Wu, Bolin Ding, Xu Chu, Zhewei Wei, Xiening Dai, Tao Guan, Jingren Zhou</i>	
ParChain: A Framework for Parallel Hierarchical Agglomerative Clustering using Nearest-Neighbor Chain .....	285
<i>Shangdi Yu, Yiqiu Wang, Yan Gu, Laxman Dhulipala, Julian Shun</i>	
Answering Regular Path Queries through Exemplars.....	299
<i>Komal Chauhan, Kartik Jain, Sayan Ranu, Srikanta Bedathur, Amitabha Bagchi</i>	
HET: Scaling out Huge Embedding Model Training via Cache-enabled Distributed Framework .....	312
<i>Xupeng Miao, Hailin Zhang, Yining Shi, Xiaonan Nie, Zhi Yang, Yangyu Tao, Bin Cui</i>	

FINEdex: A Fine-grained Learned Index Scheme for Scalable and Concurrent Memory Systems.....	321
<i>Pengfei Li, Yu Hua, Jingnan Jia, Pengfei Zuo</i>	
TaGSim: Type-aware Graph Similarity Learning and Computation .....	335
<i>Jiyang Bai, Peixiang Zhao</i>	
Analysis of Influence Contribution in Social Advertising .....	348
<i>Yuqing Zhu, Jing Tang, Xueyan Tang, Lei Chen</i>	
Scabbard: Single-Node Fault-Tolerant Stream Processing .....	361
<i>Georgios R Theodorakis, Fotios Kounelis, Peter Pietzuch, Holger Pirk</i>	
Enabling Personal Consent in Databases.....	375
<i>George Konstantinidis, Jet Holt, Adriane Chapman</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 Ibanez)  
Li Xiong (Emory University)  
Lianke Qin (University of California, Santa Barbara)  
Lijun Chang (The University of Sydney)  
Lin Ma (Carnegie Mellon University)  
Long Yuan (Nanjing University of Science and Technology)  
Lu Qin (UTS)  
Luciano Barbosa (Universidade Federal de Pernambuco)  
Marcelo Arenas (Universidad 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

We are pleased to present the second issue of PVLDB, Volume 15. This issue encompasses 16 regular research papers, 2 scalable data science papers, and 1 experiments analysis & benchmark paper. The papers cover a wide range of research topics including graph data processing, machine learning & applied AI for data management, database engines, data privacy, data mining, and query languages. From this issue, we can see the database community not only devotes great efforts to providing efficient solutions for solving the fundamental problems but also opens up novel research directions for addressing new challenges in the era of big data.

Reading this issue, we are delighted to see many papers on graph data processing and social networks. Particularly, two of the papers focus on bipartite graphs. Yang et al. study the problem of  $(p, q)$ -biclique counting and enumeration on bipartite graphs. Zhou et al. define the uncertain butterfly structure and propose efficient algorithms for solving the uncertain butterfly counting problem on uncertain bipartite graphs. We also have two papers addressing the path-related problems on graphs. Hao et al. devise a scalable distributed algorithm to tackle the challenges in enumerating the Hop-constrained  $s$ - $t$  simple paths over billion-scale graphs. Chauhan et al. present an effective framework for answering regular path queries through exemplars. For the fundamental graph similarity computation problem, Bai et al. propose TaGSim, a type-aware graph similarity learning and computation framework that can estimate graph edit distance in a fine-grained way. Zhu et al. analyze the influence contributions of influencers in social advertising. Ge et al. propose LargeEA to align entities between large-scale knowledge graphs. By applying AI techniques to database systems, Wu et al. explore a supervised learning framework to estimate the number of distinct values in a column. To improve the scalability of the embedding model training process in the field of machine learning, Miao et al. introduce a new distributed system framework HET. Fang et al. introduce Heracles, a framework for optimizing deep neural network operators based on reusing the information of performant tensor programs. Focusing on the technologies in database engines, Theodorakis et al. describe Scabbard, the first single-node multi-core stream processing engine that supports exactly once fault-tolerance semantics despite limited local I/O bandwidth. Grulich et al. present Babelfish, a new data processing engine designed for polyglot queries. Arman et al. introduce Origami, which is a high-performance mergesort framework with a four-phase pipeline. In addition, optimized solutions are developed for each phase. Li et al. propose FINEdex, a fine-grained learned index scheme with high scalability. Konstantinidis et al. study how to achieve fine-grained consent management in databases. In the field of data mining and analytics, Cui et al. investigate the multivariate time series forecasting problem and propose a generic framework METRO, which is integrated with multi-scale temporal graphs neural networks. Lu et al. introduce a novel hierarchical graph structure called HVS for approximate nearest neighbor search. Yu et al. study the hierarchical clustering problem and propose the ParChain framework for designing parallel hierarchical agglomerative clustering algorithms. Graur et al. evaluate query languages and systems for high-energy physics data.

All the papers will be presented at the 48th International Conference on Very Large Data Bases, 2022, in Sydney. We sincerely thank all the authors for submitting their work and all the reviewers for their outstanding service in reviewing the submissions. We hope that the reader will find this volume enjoyable.

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