



Proceedings of the VLDB Endowment

Volume 14, No. 10 – June 2021

Editors in Chief:

Xin Luna Dong and Felix Naumann

Associate Editors:

**Alon Halevy, Anastasia Ailamaki, Angela Bonifati, Arun Kumar, Ashraf Aboulnaga,
Eugene Wu, Floris Geerts, Graham Cormode, Jeffrey Xu Yu, Jiannan Wang, Jingren Zhou,
Jorge Arnulfo Quiané Ruiz, Juliana Freire, Jun Yang, Martin Theobald, Nesime Tatbul,
Paolo Papotti, Rainer Gemulla, Stefan Manegold, Stratos Idreos, Surajit Chaudhuri,
Xuemin Lin, Yi Chen, Yufei Tao, Zachary Ives, Zhifeng Bao**

Publication Editors:

Thorsten Papenbrock and Hannes Mühleisen

PVLDB – Proceedings of the VLDB Endowment

Volume 14, No. 10, June 2021.

All papers published in this issue will be presented at the 47th International Conference on Very Large Data Bases, Copenhagen, Denmark, 2021.

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.

Volume 14, Number 10, June 2021

Pages i – vii and 1703 - 1921

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. 14	iv
Editorial	vii

Research Papers

How to Design Robust Algorithms using Noisy Comparison Oracle	1703
<i>Raghavendra Addanki, Sainyam Galhotra, Barna Saha</i>	
SAND: Streaming Subsequence Anomaly Detection	1717
<i>Paul Boniol, John Paparrizos, Themis Palpanas, Michael Franklin</i>	
Optimizing Fitness-For-Use of Differentially Private Linear Queries.....	1730
<i>Yingtai Xiao, Zeyu Ding, Yuxin Wang, Danfeng Zhang, Daniel Kifer</i>	
Cryptanalysis of An Encrypted Database in SIGMOD '14.....	1743
<i>Xinle Cao, Jian Liu, Hao Lu, Kui Ren</i>	
Unconstrained Submodular Maximization with Modular Costs: Tight Approximation and Application to Profit Maximization.....	1756
<i>Tianyuan Jin, Yu Yang, Renchi Yang, Jieming Shi, Keke Huang, Xiaokui Xiao</i>	
Distributed Deep Learning on Data Systems: A Comparative Analysis of Approaches.....	1769
<i>Yuhao Zhang, Frank Mcquillan, Nandish Jayaram, Nikhil Kak, Ekta Khanna, Orhan Kislal, Domino Valdano, Arun Kumar</i>	
PR-Sketch: Monitoring Per-key Aggregation of Streaming Data with Nearly Full Accuracy.....	1783
<i>Siyuan Sheng, Qun Huang, Sa Wang, Yungang Bao</i>	
Tensors: An abstraction for general data processing.....	1797
<i>Dimitrios Koutsoukos, Supun C Nakandala, Konstantinos Karanasos, Karla Saur, Gustavo Alonso, Matteo Interlandi</i>	
Budget Sharing for Multi-Analyst Differential Privacy.....	1805
<i>David A Pujol, Yikai Wu, Brandon T Fain, Ashwin Machanavajjhala</i>	
In the Land of Data Streams where Synopses are Missing, One Framework to Bring Them All.....	1818
<i>Rudi Poepfel-Lemaitre, Martin Kiefer, Joscha Von Hein, Jorge Arnulfo Quiane Ruiz, Volker Markl</i>	
Data Acquisition for Improving Machine Learning Models.....	1832
<i>Yifan Li, Xiaohui Yu, Nick Koudas</i>	
Efficiently Answering Reachability and Path Queries on Temporal Bipartite Graphs	1845
<i>Xiaoshuang Chen, Kai Wang, Xuemin Lin, Wenjie Zhang, Lu Qin, Ying Zhang</i>	

Preference Queries over Taxonomic Domains.....	1859
<i>Paolo Ciaccia, Davide Martinenghi, Riccardo Torlone</i>	
Revisiting the Design of LSM-tree Based OLTP Storage Engine with Persistent Memory.....	1872
<i>Baoyue Yan, Xuntao Cheng, Bo Jiang, Shibin Chen, Canfang Shang, Jianying Wang, Kenry Huang, Xinjun Yang, Wei Cao, Feifei Li</i>	
Kamino: Constraint-Aware Differentially Private Data Synthesis	1886
<i>Chang Ge, Shubhankar Mohapatra, Xi He, Ihab F Ilyas</i>	
Towards Cost-Effective and Elastic Cloud Database Deployment via Memory Disaggregation	1900
<i>Yingqiang Zhang, Chaoyi Ruan, Cheng Li, Jimmy Yang, Wei Cao, Feifei Li, Bo Wang, Jing Fang, Yuhui Wang, Jingze Huo, Chao Bi</i>	
Dual-Objective Fine-Tuning of BERT for Entity Matching	1913
<i>Ralph Peeters, Christian Bizer</i>	

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 14

Editors in Chief of PVLDB

Xin Luna Dong (Amazon)
Felix Naumann (HPI, University of Potsdam)

Associate Editors of PVLDB

Ashraf Aboulnaga (Qatar Computing Research Institute,
Hamad Bin Khalifa University)
Anastasia Ailamaki (EPFL)
Zhifeng Bao (RMIT University)
Angela Bonifati (Lyon 1 University)
Surajit Chaudhuri (Microsoft Research)
Yi Chen (New Jersey Institute of Technology)
Graham Cormode (University of Warwick)
Juliana Freire (New York University)
Floris Geerts (University of Antwerp)
Rainer Gemulla (University of Mannheim)
Alon Halevy (Facebook)
Stratos Idreos (Harvard University)
Zachary Ives (University of Pennsylvania)
Arun Kumar (UC San Diego)
Xuemin Lin (University of New South Wales)
Stefan Manegold (CWI, Leiden University)
Paolo Papotti (Eurecom)
Jorge Arnulfo Quiané Ruiz (Technical University of Berlin)
Yufei Tao (Chinese University of Hong Kong)
Nesime Tatbul (Intel Labs and MIT)
Martin Theobald (Université du Luxembourg)

Jiannan Wang (Simon Fraser University)
Eugene Wu (Columbia University)
Jun Yang (Duke University)
Jeffrey Xu Yu (The Chinese University of Hong Kong)
Jingren Zhou (Alibaba)

Publication Editors

Thorsten Papenbrock (HPI, University of Potsdam)
Hannes Mühleisen (CWI)

PVLDB Managing Editor

Wolfgang Lehner (Dresden University of Technology)

PVLDB Advisory Committee

Divesh Srivastava (AT&T Labs-Research)
M. Tamer Özsu (University of Waterloo)
Juliana Freire (New York University)
Xin Luna Dong (Amazon)
Peter Boncz (CWI)
Lei Chen (Hong Kong University of Science and
Technology)
Graham Cormode (University of Warwick)
Xiaofang Zhou (University of Queensland)
Magdalena Balazinska (University of Washington)
Fatma Ozcan (IBM Almaden)
Felix Naumann (HPI, University of Potsdam)
Peter Triantafillou (University of Warwick)

Review Board

Abolfazl Asudeh (University of Illinois)
Ahmed Eldawy (University of California, Riverside)
Alan Fekete (University of Sydney)
Alekh Jindal (Microsoft)
Alex Ratner (University of Washington)
Altigran da Silva (Universidade Federal do Amazonas)
Anthony Tung (National University of Singapore)
Antonios Deligiannakis (Technical University of Crete)
Arijit Khan Nanyang (Technological University, Singapore)
Arnau Prat (Sparsity Technologies)
Ashwin Machanavajjhala (Duke University)
Asterios Katsifodimos (Technical University of Delft)
Avrilia Floratou (Microsoft)
Babak Salimi (University of Washington)
Badrish Chandramouli (Microsoft Research)
Beng Chin Ooi (National University of Singapore)
Bin Yang (Aalborg University)
Boris Glavic (Illinois Institute of Technology)
Byron Choi (Hong Kong Baptist University)
Carlos Scheidegger (University of Arizona)
Carsten Binnig (Technical University of Darmstadt)
Ce Zhang (ETH Zurich)
Chengfei Liu (Swinburne University of Technology)
Chengkai Li (University of Texas at Arlington)
Chris Jermaine (Rice University)
Christian Bizer (University of Mannheim)
Cong Yu (Google)
Daisy Zhe Wang (University of Florida)
Danica Porobic (Oracle)
Davide Mottin (Aarhus University)
Dimitris Papadias (Hong Kong University of Science and Technology)
Dong Deng (Rutgers University)
Eric Lo (Chinese University of Hong Kong)
Essam Mansour (Concordia University)
Fatma Ozcan (IBM Research)
Flip Korn (Google)
Florin Rusu (University of California, Merced)
Fotis Psallidas (Microsoft)
Francesco Bonchi (ISI Foundation)
Gao Cong (Nanyang Technological University)
George Fletcher (Technical University of Eindhoven)
Georgia Koutrika (Athena Research Center)
Hao Wei (Amazon)
Heiko Mueller (New York University)
Hong Cheng (Chinese University of Hong Kong)
Hongzhi Wang (Harbin Institute of Technology)
Hung Ngo (RelationalAI)
Immanuel Trummer (Cornell University)
Ingo Müller (ETH Zürich)
Jana Giceva (Technical University of Munich)
Jennie Rogers (Northwestern University)
Jeong-Hyon Hwang (University at Albany, State University of New York)
Jiaheng Lu (University of Helsinki)
Jianliang Xu (Hong Kong Baptist University)

Jianxin Li (Deakin University)
Jignesh Patel (University of Wisconsin)
Johann Gamper (Free University of Bozen-Bolzano)
Johannes Gehrke (Microsoft)
Jonas Traub (Technical University of Berlin)
Joy Arulraj (Georgia Tech)
Ju Fan (Renmin University of China)
K. Selçuk Candan (Arizona State University)
Kai Zeng (Alibaba)
Katja Hose (Aalborg University)
Ken Salem (University of Waterloo)
Kenneth A. Ross (Columbia University)
Khuzaima Daudjee (University of Waterloo)
Konstantinos Karanasos (Microsoft)
Laurel Orr (Stanford University)
Lei Chen (Hong Kong University of Science and Technology)
Lei Zou (Peking University)
Li Xiong (Emory University)
Lu Chen (Aalborg University)
Lu Qin (University of Technology Sydney)
Manasi Vartak (Verta)
Manos Athanassoulis (Boston University)
Manos Karpathiotakis (Facebook)
Marco Serafini (University of Massachusetts Amherst)
Marcos Antonio Vaz Salles (University of Copenhagen)
Mark Callaghan (MongoDB)
Markus Weimer (Microsoft)
Matei Zaharia (Stanford University, Databricks)
Matteo Interlandi (Microsoft)
Matthaios Olma (Microsoft Research)
Meihui Zhang Beijing (Institute of Technology)
Miao Qiao (University of Auckland)
Michael H. Böhlen (University of Zurich)
Michael Cafarella (University of Michigan)
Mirek Riedewald (Northeastern University)
Mohamed Mokbel (Qatar Computing Research Institute)
Mohamed Sarwat (Arizona State University)
Mohammad Sadoghi (University of California, Davis)
Mourad Ouzzani (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Muhammad Aamir Cheema (Monash University)
Murat Demirbas (University at Buffalo, SUNY)
Nan Tang (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Nick Koudas (University of Toronto)
Nikolaus Augsten (University of Salzburg)
Norman May (SAP)
Norman Paton (University of Manchester)
Odysseas Papapetrou (Technical University of Eindhoven)
Oliver A. Kennedy (University at Buffalo, SUNY)
Paolo Merialdo (Roma Tre University)
Paraschos Koutris (University of Wisconsin – Madison)
Peter Boncz (Centrum Wiskunde & Informatica)
Qin Zhang Indiana (University Bloomington)
Raja Appuswamy (Eurecom)
Ralf Schenkel (University of Trier)

Raul Castro Fernandez (University of Chicago)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
Reynold Cheng (The University of Hong Kong)
Reza Akbarinia (INRIA)
Ruoming Jin (Kent State University)
Ryan Johnson (Amazon Web Services)
S. Sudarshan (IIT Bombay)
Sanjay Krishnan (University of Chicago)
Saravanan Thirumuruganathan (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Sebastian Schelter (University of Amsterdam)
Semih Salihoglu (University of Waterloo)
Senjuti Basu Roy (New Jersey Institute of Technology)
Shaoxu Song (Tsinghua University)
Shimin Chen (Chinese Academy of Sciences)
Sibo Wang (The Chinese University of Hong Kong)
Silu Huang (Microsoft Research)
Spyros Blanas (Ohio State University)
Srikanth Kandula (Microsoft Research)
Steffen Zeuch (German Research Centre for Artificial Intelligence - DFKI)
Stijn Vansummeren (Université libre de Bruxelles)
Sudeepa Roy (Duke University)
Sudip Roy (Google)
Tamer Özsu (University of Waterloo)
Themis Palpanas (University of Paris, French University Institute - IUF)
Tianzheng Wang (Simon Fraser University)
Tingjian Ge (University of Massachusetts, Lowell)
Thomas Heinis (Imperial College)
Thomas Neumann (Technical University of Munich)
Toon Calders (Universiteit Antwerpen)

Umar Farooq Minhas (Microsoft Research)
Viktor Leis (Friedrich Schiller University Jena)
Walid Aref (Purdue University)
Wei-Shinn Ku (Auburn University)
Weiren Yu (University of Warwick)
Wendy Hui Wang (Stevens Institute of Technology)
Wenjie Zhang (University of New South Wales)
Wolfgang Gatterbauer (Northeastern University)
Xi He (University of Waterloo)
Xiang Zhao (National University of Defence Technology)
Xiangyao Yu (University of Wisconsin – Madison)
Xiaokui Xiao (National University of Singapore)
Xiaolan Wang (Megagon Labs)
Xin Cao (University of New South Wales)
Xu Chu (Georgia Tech)
Yannis Velegarakis (Utrecht University)
Ye Yuan (Beijing Institute of Technology)
Yeye He (Microsoft Research)
Ying Zhang (University of Technology Sydney)
Yinghui Wu (Case Western Reserve University)
Yongjoo Park (University of Illinois at Urbana-Champaign)
Yongxin Tong (Beihang University)
Yu Yang (City University of Hong Kong)
Yuchen Li (Singapore Management University)
Yudian Zheng (Twitter)
Yunjun Gao (Zhejiang University)
Zechao Shang (University of Chicago)
Zhenjie Zhang (Singapore R&D, Yitu Technology Ltd.)
Zhewei Wei (Renmin University of China)
Ziawasch Abedjan (Technical University of Berlin)
Zoi Kaoudi (Technical University of Berlin)

EDITORIAL

On behalf of the editors in chief and all associate editors of Volume 14 of the Proceedings of the VLDB Endowment (PVLDB), it is my very pleasure to present to you the tenth issue of this volume.

PVLDB publishes research papers on a broad range of data systems related topics accepted through a journal-style reviewing process from papers submitted on a year-round monthly submission schedule.

Comprising a total of 17 excellent papers, this issue slightly exceeds the average of exactly 15 papers per issue for the first nine issues of this volume. This shows that despite the world-wide pandemic situation that we all have been facing and battling for the last 15+ months, the excellent research work in the data systems community continues unabated in both quantity and (foremost) quality. The latter is ensured by the rigorous journal-style reviewing process of PVLDB. The fact that all papers in this issue went through the revision process demonstrates that both authors and reviewers are strongly committed to collaborate on achieving the highest possible quality. The result benefits not only the readers and conference audience, but also the authors themselves.

As with previous issues, the 14 regular research papers continue to cover a broad spectrum of topics, with two "hot-spots" in the areas of data streams and differential privacy (three papers each). The other three papers cover the remaining three paper categories (one each): Experiment, Analysis & Benchmark (EA&B), Scalable Data Science (SDS), and Vision. Interestingly, all of them are on machine learning related topics. In their EA&B paper, Zhang et al. focus on Deep Learning (DL) on data systems and present a comparative analysis, both analytical and empirically, of four canonical approaches to bring Model Hopper Parallelism (MOP), a parallel DL model selection technique, to DB-resident data. In their SDS paper, Peeters and Bizer present JointBERT, a dual-objective training method for BERT, combining binary entity matching with multi-class classification, and experimentally compare JointBERT to various other matching methods. In their vision paper, Koutsoukos et al. study to what extent Tensor Computation Runtimes --- designed to support Deep Learning workloads --- could be used to support non-ML data processing workloads, in particular graph processing and relational operators.

Many thanks to all the authors for their creative and inspiring ideas, and for the hard work of writing them up in excellent papers to share them with the community (and beyond)!

Many thanks also to the reviewers for their hard work of reviewing and discussing the papers, as well as carefully guiding them through revision, that allowed us to assemble the selection of high-quality papers presented in this issue.

Last, but sure not least, I would like to express my big "Thank you" to the editors in chief, Xin Luna Dong and Felix Naumann, for their excellent and tireless work advising, guiding and helping us associate editors in the process of compiling an extraordinary PVLDB volume 14.

Stefan Manegold
PVLDB Associate Editor