



Proceedings of the VLDB Endowment

Volume 14, No. 7 – March 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. 7, March 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 7, March 2021

Pages i – vi and 1124 - 1253

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

Collective Influence Maximization for Multiple Competing Products with an Awareness-to-Influence Model	1124
<i>Dimitris Tsaras, George Trimponias, Lefteris Ntaflos, Dimitris Papadias</i>	
Finding Group Steiner Trees in Graphs with both Vertex and Edge Weights.....	1137
<i>Yahui Sun, Xiaokui Xiao, Bin Cui, Saman Halgamuge, Theodoros Lappas, Jun Luo</i>	
Optimizing Bipartite Matching in Real-World Applications by Incremental Cost Computation	1150
<i>Tenindra Abeywickrama, Victor Liang, Kian-lee Tan</i>	
The Case for NLP-Enhanced Database Tuning: Towards Tuning Tools that “Read the Manual” ...	1159
<i>Immanuel Trummer</i>	
Errata for “Unifying Consensus and Atomic Commitment for Effective Cloud Data Management”.	1166
<i>Sujaya Maiyya, Faisal Nawab, Divy Agrawal, Amr El Abbadi</i>	
Software-Defined Data Protection: Low Overhead Policy Compliance at the Storage Layer is Within Reach!.....	1167
<i>Zsolt István, Soujanya Ponnappalli, Vijay Chidambaram</i>	
TRACE: Real-time Compression of Streaming Trajectories in Road Networks	1175
<i>Tianyi Li, Lu Chen, Christian Jensen, Torben Bach Pedersen</i>	
Shortest Paths and Centrality in Uncertain Networks.....	1188
<i>Arkaprava Saha, Ruben Brokkelkamp, Yllka Velaj, Arijit Khan, Francesco Bonchi</i>	
Adaptive Data Augmentation for Supervised Learning over Missing Data.....	1202
<i>Tongyu Liu, Yingqing Luo, Ju Fan, Nan Tang, Guoliang Li, Xiaoyong Du</i>	
KLL±: Approximate Quantile Sketches over Dynamic Datasets	1215
<i>Fuheng Zhao, Sujaya Maiyya, Ryan Weiner, Divy Agrawal, Amr El Abbadi</i>	
Distributed Numerical and Machine Learning Computations via Two-Phase Execution of Aggregated Join Trees	1228
<i>Dimitrije Jankov, Binhang Yuan, Shangyu Luo, Chris Jermaine</i>	
An Inquiry into Machine Learning-based Automatic Configuration Tuning Services on Real-World Database Management Systems.....	1241
<i>Dana Van Aken, Dongsheng Yang, Sebastien Brillard, Ari Fiorino, Bohan Zhang, Christian Billian, Andrew Pavlo</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

As Associate Editors of PVLDB Vol 14 Research Track, specifically for the inaugural year of the Scalable Data Science (SDS) category, we are delighted to present this seventh issue of the volume. Going over the papers in this issue holistically, even just this one issue exemplifies the incredible breadth and depth, as well as intellectual diversity of the VLDB community on all aspects of data management research. We observe at least three kinds of diversity:

First, the topics of the papers in this issue span a diverse set of timely and interesting problems, including both traditional topics, such as stream processing, storage systems, transactions, data mining, graph mining, and data cleaning, as well as the emerging strong bidirectional bridges between the worlds of DB and ML/AI. This includes an SDS paper that leverages incremental computation to enable more scalability for a well-known matching algorithm, thus devising a more practical approach. This paper also evaluates its ideas on interesting new real-world datasets from an industry that has become a key part of our modern app-driven lives, namely ride-sharing apps. While a more comprehensive overview of the inaugural year of SDS will be released in due course, we observe that many of this paper's attributes – important practical problem, interesting technical contributions (even if not entirely novel), use of real-world datasets, evaluation of scalability, and high potential for (or already demonstrated) practical impact – were a common theme for all accepted SDS papers in Vol 14 so far. We believe all this reflects the early success of the aspirational goals and rationales of this new category under the Research Track.

Second, there are no less than 5 kinds of publications in this one issue! This includes all 4 kinds of Research Track papers: 7 Regular papers, 2 Vision papers, 1 SDS paper, and 1 Experiments, Analysis, and Benchmarks (EAB) paper. Finally, there is also 1 Errata for a prior PVLDB paper, demonstrating the continued appeal of this healthy publishing practice. This shows that the carefully designed multi-track, multi-category approach of PVLDB continues to inspire these various forms of research, each of which brings something different and valuable to the table.

Finally, the amazing geographic diversity of the VLDB community is also well-reflected in just this single issue, with the authors spanning 11 nations across 4 continents. Just over half the papers have at least one author with an affiliation in the USA; exactly half have at least one author from Asia-Pacific; a third from Europe; and one from the Middle East. As different computing research communities aim to expand their geographic footprint, the geographic diversity of the VLDB community remains a source of strength for its research output.

Almost all the papers in this issue were published after a revision, demonstrating the continued utility and benefit of this mechanism for both the quality and quantity of VLDB publications to remain healthy. As more computing research communities start adopting this mechanism after seeing PVLDB's decade of success with this approach, we hope more computing venues' peer review processes become more conversational and less adversarial.

We wish you good health as the pandemic thankfully comes to an end. Our best wishes for a successful hybrid VLDB 2021 conference. We look forward to it and to the return of normalcy with regular in-person events that enable true interaction and networking among all researchers and practitioners in the VLDB community.

Arun Kumar, Alon Halevy, and Nesime Tatbul
Associate Editors for Scalable Data Science Research Category