



Proceedings of the VLDB Endowment

Volume 14, No. 13 – September 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. 13, September 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.

Volume 14, Number 13, July 2021

Pages i – ix and 3253 - 3420

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	ix

Research Papers

TSCache: An Efficient Flash-based Caching Scheme for Time-series Data Workloads	3253
<i>Jian Liu, Kefei Wang, Feng Chen</i>	
MP-RW-LSH: An Efficient Multi-Probe LSH Solution to ANNS-L_1	3267
<i>Huayi Wang, Jingfan Meng, Long Gong, Jun Xu, Mitsunori Ogihara</i>	
View Selection over Knowledge Graphs in Triple Stores	3281
<i>Theofilos Mailis, Yannis Kotidis, Stamatis Christoforidis, Evgeny Kharlamov, Yannis Ioannidis</i>	
Frequency-Hiding Order-Preserving Encryption with Small Client Storage.....	3295
<i>Dongjie Li, Siyi Lv, Yanyu Huang, Yijing Liu, Tong Li, Zheli Liu, Liang Guo</i>	
Modularis: Modular Relational Analytics over Heterogeneous Distributed Platforms.....	3308
<i>Dimitrios Koutsoukos, Ingo Müller, Renato Marroquin, Ana Klimovic, Gustavo Alonso</i>	
Time-Topology Analysis	3322
<i>Yunkai Lou, Chaokun Wang, Tiankai Gu, Hao Feng, Jun Chen, Jeffrey Xu Yu</i>	
Quantifying identifiability to choose and audit epsilon in differentially private deep learning	3335
<i>Daniel Bernau, Günther Eibl, Philip-william Grassal, Hannah Keller, Florian Kerschbaum</i>	
Data Management in Microservices: State of the Practice, Challenges, and Research Directions ..	3348
<i>Rodrigo N Laigner, Yongluan Zhou, Marcos Antonio Vaz Salles, Yijian Liu, Marcos Kalinowski</i>	
PerfGuard: Deploying ML-for-Systems without Performance Regressions, Almost!	3362
<i>H M Sajjad Hossain, Marc T Friedman, Hiren Patel, Shi Qiao, Soundar Srinivasan, Markus Weimer, Rammelt Ammerlaan, Lucas Rosenblatt, Gilbert Antonius, Peter Orenberg, Vijay Ramani, Abhishek Roy, Irene Shaffer, Alekh Jindal</i>	
DSB: A Decision Support Benchmark for Workload-Driven and Traditional Database Systems.....	3376
<i>Bailu Ding, Surajit Chaudhuri, Johannes Gehrke, Vivek Narasayya</i>	
Computing How-Provenance for SPARQL Queries via Query Rewriting	3389
<i>Daniel Hernández, Luis Galárraga, Katja Hose</i>	
UDO: Universal Database Optimization using Reinforcement Learning	3402
<i>Junxiong Wang, Immanuel Trummer, Debabrota Basu</i>	

Scalable Data Science Invited Talks

Internet Traffic Analysis at Scale.....	3415
<i>Anja Feldmann</i>	
The Power of Summarization in Graph Mining and Learning: Smaller Data, Faster Methods, More Interpretability	3416
<i>Danai Koutra</i>	
Summarizing Patients Like Mine via an On-demand Consultation Service	3417
<i>Nigam Shah</i>	
Towards Scalable Online Machine Learning Collaborations with OpenML.....	3418
<i>Joaquin Vanschoren</i>	
From ML Models to Intelligent Applications: The Rise of MLOps	3419
<i>Manasi Vartak</i>	
Designing Production-Friendly Machine Learning.....	3420
<i>Matei Zaharia</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 (University of Marburg)
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)

Distinguished Associate Editors of PVLDB

Volume 14

Zhifeng Bao
Arun Kumar
Yufei Tao
Graham Cormode

Distinguished Review Board Members of PVLDB

Volume 14

Ashwin Machanavajjhala
Boris Glavic
Chris Jermaine
George Fletcher
Ingo Müller
Jennie Rogers
Ju Fan
Khuzaima Daudjee
Laurel Orr
Lei Chen
Marco Serafini
Marcos Antonio Vaz Salles
Mohamed Mokbel
Mourad Ouzzani
Nan Tang
Norman May
Odysseas Papapetrou
Oliver A. Kennedy
Raul Castro Fernandez
Saravanan Thirumuruganathan
K. Selçuk Candan
Srikanth Kandula
Sudeepa Roy
Tamer Özsu
Thomas Neumann
Wendy Hui Wang
Xi He
Xiaolan Wang
Xu Chu
Yeye He
Yuchen Li

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)

Additional Reviewers

Aarati Kakaraparthi
Ahmed Mahmood
Aida Sheshbolouki
Akil Sevim
Aleksey Charapko
Ali Sadeghian
Alkis Simitsis
Amir Galad
Amol Deshpande
André Luiz da Costa Carvalho
Anh Dinh
Anil Pacaci
Anthony Colas
Arkaprava Saha
Asked Matteo Lissandrini
Baotong Lu
Benjamin Hilprecht
Boli Fang
Bonaventura Del Monte
Brad Glasbergen
Brian Hentschel
Brian Tsan
Bruce Lai
Caihua Shan
Chang Ge
Chao Tao
Chenghong Wang
Chengyuan Zhang
Chuan Xiao
Damian Jimenez
Darshana Balakrishnan
David Garcia Soriano
David Pujol
Dimitrel Loghin
Dimitrios Papadopoulos
Dimitris Tsaras
Dinh Tien Tuan Anh
Dominik Durner
Dong Wen
Donghe Kang
Edleno Silva de Moura
Emanuel Sallinger
Farahnaz Akrami
Fatma Arslan
Felix Neutatz
Fereshteh Ramzi

Francesco Fabbri
Gabor Szarnyas
Gerome Miklau
Giulia Preti
Haiquan Chen
Hanchen Wang
Haodi Ma
Haralampos Gavriilidis
Harry Sivasubramaniam
Ioana Hulpus
Ioannis Demertzis
Jacob Devasier
Jian Lou
Jiaxin Jiang
Jiecao Chen
Jinfei Liu
Jing Ma
Johes Bater
Jon Kleinberg
Joseph Near
Josue Caraballo
Junhao Gan
Jyoti Leeka
Kai Wang
Kaiyu Feng
Kajetan Maliszewski
Kevin Gaffney
Kirill Voloshin
LAI Ziliang
Lei Chen
Li Xiucheng
Lijun Chang
Lu Shengliang
Marc Makkes
Meihui Zhang
Mel Chekol
Michael Abebe
Michael Mitzenmacher
Mo Sha
Mohammed Samiul Saeef
Nasim Shirvani Mahdavi
Neoklis Polyzotis
Nikolai Karpov
Nur Al Hasan Halidar
Paul Boniol
Pengfei Tang
Peter Alvaro
Philipp Grulich
Philipp M. Grulich

Poonam Kumari
Prashant Shiralkar
Qian Tao
Qiong Luo
Qitong Wang
Qiuchen Zhang
Razvan-Gabriel Cirstea
Rogers Jeffrey Leo John
Romila Pradhan
Runsheng Guo
Rüdiger Kapitzka
Saba Eskandarian
Saheli Ghosh
Sainyam Galhotra
Samriddhi Singla
Shubhankar Mohapatra
Shufan Zhang
Shuyue Wei
Sun Shixuan
Tarikul Islam Papon
Tian Xia
Tin Vu
Tsz Nam Chan
Weiguo Zheng
Weiran Huang
Weizhen Ding
Wenyu Du
William Spoth
Xiang Lian
Xiangyu Ke
Xiaolan Gu
Xiaolin Han
Xiaoshuang Chen
Xun Jian
Yan Zhao
Yang Cao
Yannis Chronis
Ye Liu
Yifan Chen
Yifan Wang
Yilun Huang
Yinan Li
Ying Yan
Yuanhang Yu
Yubao Liu
Yuchao Tao
Yue Wang
Yuxiang Zeng
Zhengyuan Zhu

EDITORIAL

On behalf of the editors-in-chief and all associate editors of the Proceedings of the VLDB Endowment (PVLDB) Volume 14, it is my pleasure to present to you the final issue of this volume. PVLDB publishes papers accepted through a journal-style reviewing process from papers submitted on a year-round monthly submission schedule.

This issue of PVLDB includes 12 excellent papers, all of which went through a thorough and rigorous revision process. This issue has a broad spectrum of topics, ranging from data analytics, efficiency, optimization, privacy and security. For analytics, advances are made for time-topology analysis on temporal graphs, computing How-provenance for SPARQL queries on knowledge graphs, and modular relational analytics over heterogeneous distributed platforms. For efficiency, effective techniques are proposed for a caching system for time-series data workload, view materialization and view selection methodologies for querying knowledge graphs, and scalable and efficient approximate nearest neighbor search on high-dimension data. Proposals are presented on how to innovatively use machine learning (ML) techniques for query optimization. For security and privacy, novel techniques are proposed for a frequency-hiding order-preserving encryption with small client storage, and for determining privacy parameters in differentially private deep learning. As we can see, diverse data types are studied, including relational databases, distributed data, time series data, knowledge graphs, and temporal graphs.

Besides the research papers, this issue includes a systematic literature review of data management in microservices, and the design of a new benchmark for evaluating both classical and workload-driven database systems on modern decision support workloads (in the Experiments, Analysis & Benchmark category).

This issue also includes six abstracts for the invited Scalable Data Science Talks at VLDB. Vartak discusses the path from ML models to intelligent applications. To promote massively collaborative ML development, Vanschoren presents OpenML, an open-source initiative to create a platform that allows anyone to share datasets, machine learning pipelines, and full experiments, with rich metadata. Zaharia proposes to design production-friendly ML. Koutra discusses the benefits of using summarization techniques in graph mining and learning for faster speed and better interpretability. Finally, Feldmann showcases the Internet traffic analysis project.

We thank all authors and all reviewers for their outstanding work and dedication of striving for the highest quality for PVLDB. We hope the selected papers will provide valuable insights and inspire novel system contributions and follow-up research. All PVLDB papers published in this issue will be presented at the VLDB 2022 Conference, to be held in Sydney, Australia. We hope you will enjoy the conference, and hope that you will continue to submit your best work to PVLDB's future issues!

Yi Chen