Proceedings of the VLDB Endowment

Volume 14, No. 8 – April 2021

PVLDB – Proceedings of the VLDB Endowment

Volume 14, No. 8, April 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

Volume 14, Number 8, April 2021

Pages i – viii and 1254 - 1453

ISSN 2150-8097

Available at: http://www.pvldb.org and https://dl.acm.org/journal/pvldb

TABLE OF CONTENTS

Front Matter

Research Papers

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 Velegrakis (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

I am pleased to present the eighth issue of the Proceedings of the VLDB Endowment (PVLDB), Volume 14, in which a variety of interesting topics in the areas of core database systems, graph analytics, and algorithms is presented. From this volume, one can see that the database community continues to commit to the important problem of speeding up data and query processing, in all its shapes and forms. It does so, driven by new advances in technology and new challenges imposed by applications and changing practical needs.

In this volume, twelve papers are included in the regular research category. The variety of topics covered by these papers provides a nice kaleidoscopic view of the current interests in the database system research community.
Indeed, we find methods to improve query processing by fine-tuning the interaction with various levels in the memory hierarchy. For example, Suzuki et al. focus on read-intensive workloads, as common in key-value stores and graph analytics, propose a new hardware interface to flash memory that reduces CPU overhead, and showcase the obtained efficiency boost for various graph analytical tasks. Furthermore, Kulkarni et al. present Shadowfax, a distributed key-value store for events, which transparently spans various levels of the memory hierarchy and in which distribution reconfiguration is achieved without server-side key ownership checks or cross-core coordination.
In the good tradition of connecting databases and programming languages, Schiavio et al. bridge the gap between dynamic languages, for which the type of variables is checked at runtime, and Language-Integrated Query (LINQ) frameworks. A language-agnostic query engine, DynQ, is proposed, which can execute queries on dynamically typed collections.
As in previous volumes, we again see the creative use of machine learning methods for database problems. For example, Cereda et al. address the challenging problem of properly selecting database management configurations. To this aim, their proposed system, CGPTuner, uses contextual Gaussian process bandit optimization based on the DBMS's workload. In turn, Zou et al. propose the Lachesis system in which UDF-centric workflow performance is improved by exploring automatic partitioning for UDF-centric applications. The partition strategy leverages deep reinforcement learning. Finally, Wu et al. contribute to the exciting learned index paradigm by proposing a learned index with precise positions, which gracefully supports all standard index query and update operations.
Another connection to machine learning is proposed by Fan et al. They introduce entity enhancing rules, which embed machine learning predicates in a logical rule formalism, for the problem of discrepancy detection, and develop parallel (and incremental) scalable detection algorithms for this detection problem. Inspired by classical query language design and query optimization methods, Yuan et al. propose a declarative tensor programming language and develop a distributed query optimization method, based on rewriting techniques, for their tensor expressions.
Finally, a considerable number of papers in this volume deal with graphs and graph analytics in particular. For example, Yao et al. study the problem of size-bounded community search in large networks and develop efficient branch-reduce-and-bound algorithms for this purpose. Min et al. detect critical patterns in dynamic graphs by investigating the continuous subgraph matching problem. The proposed SymBi algorithm works for both edge insertion and deletions and is shown to outperform a state-of-the-art competitor. Liu et al. develop exact and approximate algorithms for finding the fastest paths through an indoor crowd and for finding paths encountering the least objects en route. Zhao et al. introduce a new minimum vertex augmentation graph problem, which is shown to encompass many interesting graph problems. They provide a theoretical and empirical analysis of the problem.

Alongside the regular research paper category, the volume contains two papers in the experiments, analysis & benchmarks category. More specifically, Orogat et al. introduce an extensible and informative benchmarking suite, CBench, for query answering on knowledge graphs. Moreover, Gupta et al. investigate the magnitude of use and complexity of procedural extensions of SQL and propose ProcBench as a benchmark. Another category, introduced this year, is the scalable Data Science category. Also, two papers in this category are included in this volume. Bharadwaj et al. address the challenging problem of finding data sources in data lakes, whereas Fang et al. develop a framework for traffic prediction based on graph convolutional en long short-term memory networks. Finally, in the vision category, Tang et al. outline their vision of how A.I. can help to automate human-easy but computer-hard data preparation tasks.

All papers have been carefully reviewed and most papers have undergone a revision. This resulted in high-quality papers that will be presented at the 47th International Conference on Very Large Data Bases, 2021, in Copenhagen. I sincerely thank all the authors for submitting their work and all the reviewers for their outstanding service reviewing the submissions. I hope that the reader will find this volume enjoyable.

Floris Geerts
PVLDB Associate Editor