Proceedings of the VLDB Endowment

Volume 14, No. 1 – September 2020

PVLDB – Proceedings of the VLDB Endowment

Volume 14, No. 1, September 2020.

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

Copyright 2020 VLDB Endowment

Volume 14, Number 1, September 2020

Pages i – vi and 1 - 85

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

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)

LETTER FROM THE EDITORS IN CHIEF

We are very pleased to present the first issue of PVLDB's Volume 14. The Proceedings of the VLDB present the latest research in the area of database and information system technology. Together with expert boards of associate editors and reviewers, submissions are carefully peer-reviewed, often entering a revision phase, then published in the journal and ultimately presented at the following VLDB conference. We are very grateful for all the great colleagues who contribute to the ongoing success of PVLDB. Before introducing the individual papers in this first issue, we would like to report on some changes for PVLDB that were implemented for this volume and of which we hope they improve the scope of the journal, the quality and rapid dissemination of the research, the quality of the reviewing process, and finally the availability and reproducibility of database research.

A new paper category, Scalable Data Science (SDS), seeks to accept papers that describe the design, implementation, experience, and evaluation of solutions and systems for practical data science and data engineering tasks on large-scale data. These papers are limited to 8 pages of content and do not necessarily propose new breakthrough algorithms or models, but emphasize solutions that either solve or advance the understanding of issues related to data science technologies in the real world. In the first months, about 15% of all submissions were in this SDS category. Another category change is the extension of the traditional Experiments and Analysis (E&A) category to include Benchmarks (EA&B). We recognize that benchmarks are drivers of research, providing experimental settings and a common evaluation for relevant problems. Designing, testing and publishing such benchmarks are important research contributions. Among the six papers of this first edition are one SDS and one EA&B contribution.

PVLDB strives to give high quality and constructive feedback in the form of reviews and meta-reviews. The latter summarize the reviews and the results of the three-week discussion phase in which reviewers exchange their view of the paper and converge to a joint decision. In the past, positive decisions were restricted to either a direct "accept" or a "major revision". To account for submissions that are already close to an accept decision and for which reviewers had only smaller requests, we have introduced a "minor revision". Among the papers of this issue, two had originally received a minor revision decision and were ultimately accepted.

Finally, our community has recognized that reproducibility, repeatability, and replicability are very important factors to advance database research. The ability to re-use code, data and experimental settings and thus compare one's own result to previous work allows independent validation of the previous results and accelerates the advancement of state of the art. For PVLDB's volume 14 we have introduced several measures to encourage authors to make available their research artifacts. First, we allow authors to submit supplemental material, such as data and code, and explain those files in a submission form question. In turn, reviewers are asked to answer a review question about the availability and potential reproducibility of the presented research. Finally, we have instituted an optional new metadatum on the front page of accepted papers, namely an "Artifacts available" tag along with an author-supplied stable URL. PVLDB's reproducibility chairs verify the availability of research artifacts, yet without performing reproducibility checks.

For this issue, the review board has selected six contributions, addressing both core database technology and work about non-relational data management. Interestingly, five of these contributions make direct use of machine learning technologies. Marcus et al. present an indexing benchmark specifically designed to evaluate and compare the novel so-called learned indexes, which train a model of the underlying data instead of relying on a pre-determined data structure. Wang et al. also address a traditional database problem, namely that of query plan cost estimation, but design a cost model for incremental data processing tasks as they appear for instance in stream processing systems. Yang et al. also apply machine learning to improve core database performance, in this case learning join cardinalities - an important ingredient in cost models. Heo et al. report on their solution to employ weak supervision in the form of labeling functions to annotate industrial images to identify and locate defects, such as scratches. Yang et al. address the important problem of scalability for the complex task of computing attributed network embeddings without compromising the quality of the embedding. Finally, Li et al. address the well-known entity matching task by applying language models, for instance to the sometimes-complex names of products.

All papers will be presented at the 2021 Conference on Very Large Databases (VLDB'21) in Copenhagen. We hope you enjoy reading and look forward to seeing you there.

Xin Luna Dong and Felix Naumann
Editors-in-Chief of PVLDB Volume 14
Program Chairs for VLDB 2021