Proceedings of the VLDB Endowment

Volume 15, No. 6 – February 2022

PVLDB – Proceedings of the VLDB Endowment

Volume 15, No. 6, February 2022.

All papers published in this issue will be presented at the 48th International Conference on Very Large Data Bases, Sydney, Australia, 2022.

Copyright 2022 VLDB Endowment

Volume 15, Number 6, February 2022

Pages i – vii and 1132 - 1310

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. 15

Review Board

Abolfazl Asudeh (University of Michifan)
Aécio Santos (New York University)
Ahmed Eldawy (University of California, Riverside)
Alexander Hall (RelationalAI)
Alexander J Ratner (University of Washington)
Aline Bessa (New York University)
Alkis Simitsis (Athena Research Center)
Altigran da Silva (Universidade Federal do Amazonas)
AnHai Doan (University of Wisconsin-Madison)
Anna Fariha (Microsoft)
Anton Dignös (Free University of Bozen-Bolzano)
Antonio Cavalcante Araujo Neto (University of Alberta)
Arijit Khan (Nanyang Technological University)
Arvind Arasu (Microsoft)
Babak Salimi (University of California, San Diego)
Bailu Ding (Microsoft Research)
Bertram Ludaescher (University of Illinois)
Bolong Zheng (Huazhong University of Science and Technology)
Brandon Haynes (Gray Systems Lab, Microsoft)
Byron Choi (Hong Kong Baptist University)
Carlo Curino (Microsoft -- GSL)
Carlos Scheidegger (The University of Arizona)
Carsten Binnig (TU Darmstadt)
Ce Zhang (ETH)
Cheng Long (Nanyang Technological University)
Chengfei Liu (Swinburne University of Technology)
Chuan Lei (Instacart)
Chunbin Lin (Amazon AWS)
Curtis Dyreson (Utah State University)
Dan Kifer (Pennsylva State University)
Dana M Van Aken (Carnegie Mellon University)
Daniel Deutch (Tel Aviv University)
Daniel Oliveira (UFF, Brazil)
David Koop (Northern Illinois University)
Davide Mottin (Aarhus University)
Dong Xie (Penn State University)
Eduardo Ogasawara (CEFET-RJ)
Eleni Tzirita Zacharatou (TU Berlin)
Fabio Porto (LNCC)
Faisal Nawab (University of California at Irvine)
Fan Zhang (Guangzhou University)
Fatemeh Nargesian (University of Rochester)
Fei Chiang (McMaster University)
Florin Rusu (UC Merced)
Floris Geerts (University of Antwerp)
Fotis Psallidas (Microsoft)
George Fletcher (Eindhoven University of Technology)
George Papadakis (University of Athens)
Gerhard Weikum (Max-Planck-Institut fur Informatik)
Germain Forestier (University of Haute Alsace)
Guoliang Li (Tsinghua University)
Haipeng Dai (Nanjing University)
Harish Doraiswamy (Microsoft Research India)
Heiko Mueller (DeepReason.ai)
Herodotos Herodotou (Cyprus University of Technology)
Holger Pirk (Imperial College)
Hongzhi Yin (The University of Queensland)
Huiping Cao (New Mexico State University)
Immanuel Trummer (Cornell)
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)
Ippokratis Pandis (Amazon)
Ishtiyaque Ahmad (University of California, Santa Barbara)
Jae-Gil Lee (KAIST)
Jana Giceva (TU Munich)
Jeffrey Xu Yu (Chinese University of Hong Kong)
Jens Teubner (TU Dortmund University)
Jia Zou (Arizona State University)
Jian Pei (Simon Fraser University)
Jianguo Wang (Purdue University)
Jiannan Wang (Simon Fraser University)
Jianxin Li (Deakin University)
Jianye Yang (Central South University)
Jiwon Seo (Hanyang University)
Johannes Gehrke (Microsoft)
Jorge Arnulfo Quiane Ruiz (TU Berlin)
Joseph Near (University of Vermont)
Junhu Wang (Griffith University)
Kaiping Zheng (National University of Singapore)
Kangfei Zhao (The Chinese University of Hong Kong)
Karima Echihabi (Mohammed VI Polytechnic University)
Katja Hose (Aalborg University)
Kenneth A Ross (Columbia University)
Kostas Zoumpatianos (Snowflake Computing)
Lei Zou (Peking University)
Leopoldo Bertossi (Universidad Adolfo Ibanez)
Li Xiong (Emory University)
Lianke Qin (University of California, Santa Barbara)
Lijun Chang (The University of Sydney)
Lin Ma (Carnegie Mellon University)
Long Yuan (Nanjing University of Science and Technology)
Lu Qin (UTS)
Luciano Barbosa (Universidade Federal de Pernambuco)
Marcelo Arenas (Universidad Catolica & IMFD)
Maria Luisa Sapino (U. Torino)
Matteo Lissandrini (Aalborg University)
Matthias Boehm (Graz University of Technology)
Matthias Renz (University of Kiel)
Max Heimel (Snowflake)
Maximilian Schleich (University of Washington)
Meihui Zhang (Beijing Institute of Technology)
Melanie Herschel (Universität Stuttgart)
Michael Abebe (University of Waterloo)
Min Xie (Instacart)
Mirella M Moro (Universidade Federal de Minas Gerais)
Mohamed Sarwat (Arizona State University)
Mohammad Dashti (MongoDB)
Mohammad Javad Amiri (University of Pennsylvania)
Mohammad Sadoghi (University of California, Davis)
Muhammad Aamir Cheema (Monash University)

Nikita Bhutani (Megagon Labs)
Oliver A Kennedy (University at Buffalo, SUNY)
Panos K. Chrysanthis (University of Pittsburgh)
Paolo Missier (Newcastle University)
Parth Nagarkar (NMSU)
Paul Groth (University of Amsterdam)
Peng CHENG (East China Normal University)
Peter Pietzuch (Imperial College London)
Pierangela Samarati (Universita delgi Studi di Milano)
Pinar Karagoz (METU, Turkey)
Pinar Tozun (IT University of Copenhagen)
Prithu Banerjee (UBC)
Raoni Lourenço (New York University)
Raul Castro Fernandez (UChicago)
Ravi Ramamurthy (Microsoft)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
Renata Borovica-Gajic (University of Melbourne)
Reynold Cheng (The University of Hong Kong)
Rui Mao (Shenzhen University)
Ruoming Jin (Kent State University)
Sai Wu (Zhejiang University)
Sainyam Galhotra (University of Chicago)
Sanjay Krishnan (University of Chicago)
Sanjib Kumar Das (Google)
Sayan Ranu (IIT Delhi)
Sebastian Link (University of Auckland)
Semih Salihoglu (University of Waterloo)
Senjuti Basu Roy (New Jersey Institute of Technology)
Sergey Melnik (Google)
Shantanu Sharma (New Jersey Institute of Technology)
Shaoxu Song (Tsinghua University)
Sheng Wang (New York University)
Shimin Chen (Chinese Academy of Sciences)
Shumo Chu (University of California, Santa Barbara)
Shweta Jain (University of Illinois, Urbana-Champaign)
Sibo Wang (The Chinese University of Hong Kong)
Srinivasan Keshav (University of Cambridge)
Steffen Zeuch (DFKI GmbH)
Steven E Whang (KAIST)
Subarna Chatterjee (Harvard University)
Sudip Roy (Google)
Supun C Nakandala (University of California, San Diego)
Tamer Özsu (University of Waterloo)
Tarique A Siddiqui (Microsoft Research)
Thomas Heinis (Imperial College)
Thomas Neumann (TUM)
Tianzheng Wang (Simon Fraser University)
Tien Tuan Anh Dinh (Singapore University of Technology and Design)
Tilmann Rabl (HPI, University of Potsdam)
Ting Yu (Qatar Computing Research Institute)
Torben Bach Pedersen (Aalborg University)
Torsten Grust (Universität Tübingen)
Umar Farooq Minhas (Microsoft Research)
Vasiliki Kalavri (Boston University)
Verena Kantere (National Technical University of Athens)
Victor Zakhary (Oracle)
Vivek Narasayya (Microsoft Research)
Vraj Shah (University of California, San Diego)
Walid G Aref (Purdue)
Wasay Abdul (Harvard)
Wei Wang (Hong Kong University of Science and Technology (Guangzhou))
Wei Lu (Renmin university of china)
Weiren Yu (University of Warwick)
Wen Hua (The University of Queensland)
Wolfgang Lehner (TU Dresden)
Xi He (University of Waterloo)
Xiang Lian (Kent State University)
Xiao Qin (IBM Research)
Xiaofei Zhang (University of Memphis)
Xiaokui Xiao (National University of Singapore)
Xiaolan Wang (Megagon Labs)
Xiaoyang Wang (Zhejiang Gongshang University)
Xin Huang (Hong Kong Baptist University)
Yael Amsterdamer (Bar-Ilan university)
Yanyan Shen (Shanghai Jiao Tong University)
Ye Yuan (Northeastern University)
Yeye He (Microsoft Research)
Yi Chen (NJIT)
Yi Lu (MIT)
Yikai Zhang (Chinese University of Hong Kong)
Yinan Li (Microsoft Research)
Ying Zhang (University of Technology Sydney)
Yongxin Tong (Beihang University)
Yuanyuan Zhu (Wuhan University)
Yue Wang (Shenzhen Institute of Computing Sciences, Shenzhen University)
Yufei Tao (Chinese University of Hong Kong)
Yuliang Li (Megagon Labs)
Yuncheng Wu (National University of Singapore)
Yunjun Gao (Zhejiang University)
Yuval Moskovitch (University of Michigan)
Zhifeng Bao (RMIT University)
Zhongle Xie (Zhejiang University)
Zi Huang (University of Queensland)
Ziawasch Abedjan (Leibniz Universität Hannover)
Zohar Karnin (Amazon)
Zsolt István (IT University of Copenhagen)

LETTER FROM THE EDITORS IN CHIEF

We are pleased to present the sixth issue of PVLDB, Volume 15. This issue contains 14 papers in total including 11 regular research papers and 3 scalable data science (SDS) papers. A broad range of topics are covered in this issue including distributed database systems, machine learning & applied AI for data management, spatial data management, graph data management, database engines, provenance and workflows, data quality, data mining, and information retrieval.

For the first paper in this issue, Wang et al. propose an efficient distributed algorithm for agglomerative hierarchical clustering. Next, Chang et al. present a near-optimal approach for solving the edge connectivity-based hierarchical graph decomposition problem. Tong et al. study the problem of secure spatial queries over data federation and present efficient solutions to solve the problem. Fuchs et al. introduce Sortledton, a universal graph data structure that is optimized for the most relevant data access patterns used by graph computation kernels. Zhang et al. propose NBTree, a lock-free persistent-memory-friendly B+-Tree for eADR-enabled persistent memory systems. Tuli et al. propose TranAD, a deep transformer network-based model for anomaly detection in multivariate time series data. Zhao et al. present the deterministic algorithms to solve the frequency estimation and frequent item problems in the bounded-deletion model. Zhao et al. propose a distributed graph neural network system ByteGNN to support efficient GNN training. Jiang et al. introduce graph neural network models for community search and attributed community search problems. Li et al. present Hyper-Tune, an efficient and robust distributed hyper-parameter tuning framework. Minartz et al. propose efficient algorithms for detecting multivariate correlations in static and streaming data. Poppe et al study the problem of proactive auto-scaling in Microsoft Azure SQL Database Serverless. Cheng et al. propose PGE that leverages both text information and graph structure in product graphs to learn embeddings for error detection. Manne et al. present effective solutions for the multiversion replay problem.

All the papers in this issue will be presented at the 48th International Conference on Very Large Data Bases, 2022, in Sydney. We sincerely thank all the authors for submitting their work and all the reviewers for their outstanding service in reviewing the submissions. We hope that the reader will find this volume enjoyable.

Fatma Özcan, Juliana Freire and Xuemin Lin
Editors-in-Chief of PVLDB Volume 15
Program Chairs for VLDB 2022