



Proceedings of the VLDB Endowment

Volume 14, No. 5 – January 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. 5, January 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 5, January 2021

Pages i – vi and 721 - 862

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

| | |
|--|-----|
| FlashP: An Analytical Pipeline for Real-time Forecasting of Time-Series Relational Data | 721 |
| <i>Yan Shuyuan, Bolin Ding, Wei Guo, Jingren Zhou, Zhewei Wei, Xiaowei Jiang, Sheng Xu</i> | |
| Efficient Streaming Subgraph Isomorphism with Graph Neural Networks..... | 730 |
| <i>Chi Thang Duong, Dung Hoang, Hongzhi Yin, Matthias Weidlich, Quoc Viet Hung Nguyen, Karl Aberer</i> | |
| Epoch-based Commit and Replication in Distributed OLTP Databases | 743 |
| <i>Yi Lu, Xiangyao Yu, Lei Cao, Samuel Madden</i> | |
| Hierarchical Core Maintenance on Large Dynamic Graphs..... | 757 |
| <i>Zhe Lin, Fan Zhang, Xuemin Lin, Wenjie Zhang, Zhihong Tian</i> | |
| Analyzing and Mitigating Data Stalls in DNN Training | 771 |
| <i>Jayashree Mohan, Amar Phanishayee, Ashish Raniwala, Vijay Chidambaram</i> | |
| Persistent Memory Hash Indexes: An Experimental Evaluation..... | 785 |
| <i>Daokun Hu, Zhiwen Chen, Jianbing Wu, Jianhua Sun, Hao Chen</i> | |
| Optimizing An In-memory Database System For AI-powered On-line Decision Augmentation Using Persistent Memory..... | 799 |
| <i>Cheng Chen, Jun Yang, Mian Lu, Taize Wang, Zhao Zheng, Yuqiang Chen, Wenyuan Dai, Bingsheng He, Weng-Fai Wong, Guoan Wu, Yuping Zhao, Andy Rudoff</i> | |
| DBTagger: Multi-Task Learning for Keyword Mapping in NLIDBs Using Bi-Directional Recurrent Neural Networks..... | 813 |
| <i>Arif Usta, Akifhan Karakayali, Özgür Ulusoy</i> | |
| Improving Information Extraction from Visually Rich Documents using Visual Span Representations | 822 |
| <i>Ritesh Sarkhel, Arnab Nandi</i> | |
| Zen: a High-Throughput Log-Free OLTP Engine for Non-Volatile Main Memory | 835 |
| <i>Gang Liu, Leying Chen, Shimin Chen</i> | |
| Differentially Private Binary- and Matrix-Valued Data Query: An XOR Mechanism..... | 849 |
| <i>Tianxi Ji, Pan Li, Emre Yilmaz, Erman Ayday, Yanfang Ye, Jinyuan Sun</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 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

Welcome to Issue 5 of PVLDB Volume 14. The issue contains eleven papers that address a diverse set of topics, from new hardware to graph data management to privacy. The papers come from academia and industry, and they fall in three of the four PVLDB paper categories: Research, Experiments Analysis & Benchmarks, and Scalable Data Science. The diversity of paper topics and categories is a testament to the openness of our community and the value of PVLDB's innovation in creating the new Scalable Data Science category.

In the first paper of the issue, Yan et al. present a system for scalable, real-time forecasting of high-dimensional time series data. Next, Duong et al. present a technique for speeding up the NP-hard problem of subgraph isomorphism by caching and reusing results. Lu et al. describe COCO, a distributed OLTP database that improves performance by using the interesting idea of grouping transactions into epochs and treating each epoch as a commit unit. Lin et al. propose a solution for incrementally maintaining the k-core hierarchy of a graph as the graph evolves. Mohan et al. present the first Experiments, Analysis, and Benchmarking (EA&B) paper in the issue. They study the effect of fetching data from storage and preprocessing it in memory on the training time of deep neural networks and present several insights based on their analysis. Hu et al. also present an EA&B paper in which they study the performance of hash indexes that reside in persistent memory. Hash indexes are a foundational access method in database systems, so understanding their performance on new memory hardware is very useful. Chen et al. also consider persistent memory. They present a distributed in-memory database system that supports on-line feature extraction and uses a data structure optimized for persistent memory. Usta et al. is a Scalable Data Science paper that deals with an essential step in converting natural language queries to SQL: mapping keywords to database elements. Sarkhel and Nandi propose a method for information extraction from visually rich documents that substantially outperforms the state of the art. Liu et al. describe Zen, a high-throughput OLTP engine designed for non-volatile memory. Finally, Ji et al. propose a technique for guaranteeing differential privacy for binary- and matrix-valued data by performing an XOR operation between the query results and noise generated according to a particular distribution.

This issue is the result of the efforts of the PVLDB editors, the reviewers, and most importantly the authors. I thank them all.

Ashraf Aboulnaga
PVLDB Associate Editor