



Proceedings of the VLDB Endowment

Volume 14, No. 6 – February 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. 6, February 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 6, February 2021

Pages i – vii and 863 - 1123

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

Errata for "Cerebro: A Data System for Optimized Deep Learning Model Selection"	863
<i>Supun Nakandala, Yuhao Zhang, Arun Kumar</i>	
ParaX: Boosting Deep Learning for Big Data Analytics on Many-Core CPUs.....	864
<i>Lujia Yin, Yiming Zhang, Zhaoning Zhang, Yuxing Peng, Peng Zhao</i>	
Optimization of Threshold Functions over Streams	878
<i>Walter Cai, Philip Bernstein, Wentao Wu, Badrish Chandramouli</i>	
Budget Constrained Interactive Search for Multiple Targets.....	890
<i>Xuliang Zhu, Xin Huang, Byron Choi, Jiaxin Jiang, Zhaonian Zou, Jianliang Xu</i>	
On the String Matching with k Differences in DNA Databases	903
<i>Yangjun Chen, Hoang Hai Nguyen</i>	
Fast Algorithm for Anchor Graph Hashing	916
<i>Yasuhiro Fujiwara, Sekitoshi Kanai, Yasutoshi Ida, Atsutoshi Kumagai, Naonori Ueda</i>	
Adaptive Code Generation for Data-Intensive Analytics	929
<i>Wangda Zhang, Junyoung Kim, Kenneth Ross, Eric Sedlar, Lukas Stadler</i>	
Materializing Knowledge Bases via Trigger Graphs.....	943
<i>Efthymia Tsamoura, David Carral, Enrico Malizia, Jacopo Urbani</i>	
Dealer: An End-to-End Model Marketplace with Differential Privacy.....	957
<i>Jinfei Liu, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, Jimeng Sun</i>	
NOAH: Interactive Spreadsheet Exploration with Dynamic Hierarchical Overviews	970
<i>Sajjadur Rahman, Mangesh Bendre, Yuyang Liu, Shichu Zhu, Zhaoyuan Su, Karrie Karahalios, Aditya Parameswaran</i>	
Efficient Bi-triangle Counting for Large Bipartite Networks.....	984
<i>Yixing Yang, Yixiang Fang, Maria Orlowska, Wenjie Zhang, Xuemin Lin</i>	
Glean: Structured Extractions from Templatic Documents	997
<i>Sandeep Tata, Navneet Potti, James Wendt, Lauro Beltrão Costa, Marc Najork, Beliz Gunel</i>	
ICS-GNN: Lightweight Interactive Community Search via Graph Neural Network	1006
<i>Jun Gao, Jiazun Chen, Zhao Li, Ji Zhang</i>	

Building Enclave-Native Storage Engines for Practical Encrypted Databases.....	1019
<i>Yuanyuan Sun, Sheng Wang, Huorong Li, Feifei Li</i>	
From Natural Language Processing to Neural Databases.....	1033
<i>James Thorne, Majid Yazdani, Marzieh Saeidi, Fabrizio Silvestri, Sebastian Riedel, Alon Halevy</i>	
Randomized Error Removal for Online Spread Estimation in Data Streaming.....	1040
<i>Haibo Wang, Chaoyi Ma, Olufemi O Odegbile, Shigang Chen, Jih-kwon Peir</i>	
Teseo and the Analysis of Structural Dynamic Graphs.....	1053
<i>Dean De Leo, Peter Boncz</i>	
Charting the Design Space of Query Execution using VOILA.....	1067
<i>Tim Gubner, Peter Boncz</i>	
Heracles: An Efficient Storage Model And Data Flushing For Performance Monitoring Timeseries	1080
<i>Zhiqi Wang, Jin Xue, Zili Shao</i>	
Fine-Grained Lineage for Safer Notebook Interactions	1093
<i>Stephen Macke, Aditya Parameswaran, Hongpu Gong, Doris Lee, Doris Xin, Andrew Head</i>	
FREDE: Anytime Graph Embeddings.....	1102
<i>Anton Tsitsulin, Marina Munkhoeva, Davide Mottin, Panagiotis Karras, Ivan Oseledets, Emmanuel Müller</i>	
On Analyzing Graphs with Motif-Paths	1111
<i>Xiaodong Li, Reynold Cheng, Kevin Chen-chuan Chang, Caihua Shan, Chenhao Ma, Hongtai Cao</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 Machanavajhala (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)

EDITORIAL

The sixth issue of PVLDB's Volume 14, which we have the pleasure to announce with this letter, covers a variety of exciting topics in the areas of core database systems, machine learning, graph data, data streams, and data accessibility. We were delighted to see that the 21 papers of this issue, which represent the research effort of our community, show a strong commitment to real-world problems in highly relevant areas, such as medicine, green-IT and digitalization: Never has it been more important to produce accurate results, be careful with our limited resources and build reliable and safe systems.

This month, 17 regular research papers, three scalable data science papers, and one vision paper have been published. With eight research projects in the area of graph data, graphs clearly dominate the topic list of the current issue: Leo and Boncz propose a novel analytical database for structural dynamic graphs, Tsitsulin et al. found a novel graph embedding, Yang et al. accelerate the counting of bi-triangles in bipartite graphs, Zhu et al. search for multiple targets in graphs, Fujiwara et al. improve anchor graph hashing, Tsamoura et al. accelerate the materialization of knowledge bases with trigger graphs, Gao et al. facilitate graph neural networks for community discovery, and Li et al. carefully study motif-paths. Similar to our past five issues, research on and with machine learning techniques also plays a major role in Issue 6: Yin et al. make CPUs competitive for deep learning tasks, Liu et al. study model marketplaces, Tata et al. use machine learning to extract structured information from form-like documents, and Gao et al. facilitate graph neural networks for the discovery of communities that contain a certain query vertex. Issue 6 also features a solid set of core database research papers: Thorne et al. envision neural databases, Gubner and Boncz explore the query engine design space with a novel framework, Sun et al. demonstrate how to improve the throughput of encrypted databases, Chen and Nguyen support DNA databases in finding k-similar substring DNAs to any given query pattern, and Zhang et al. introduce database-style query optimization techniques into a programming language compiler. In the area of data streaming, Wang et al. improve the accuracy of estimating flow spread in high-rate data streams, Walter Cai et al. optimize threshold functions over streams, and Wang et al. improve the performance of monitoring time series. Last but not least, Rahman et al. feature dynamic hierarchical overviews to make large spreadsheets accessible and Macke et al. improve computational notebooks so that the notebooks can prevent erroneous interactions.

Volume 14 as a whole also marks a special milestone in the history of PVLDB, because we introduced several changes to the publication format in the background. All published papers are now PDF/A-compliant and, hence, archivable. In this way, PVLDB makes an effort towards long-term preservation of its content, which directly ties in with its promise to authors and readers that the research results are kept accessible. In our role as proceedings chairs, we also retired the old PVLDB style template in which many errors had accumulated over the years. Submissions to PVLDB are now based on the ACM style template, which is well maintained and updated regularly. With the new template, authors can switch their papers more easily between ACM conferences and benefit from a modern design. To ease the process of starting a new publication, we now provide a PVLDB template project on Overleaf - a collaborative cloud-based LaTeX editor. Another feature of the new style template is a well-defined artifact availability tag that authors can activate by providing a link to their project artifacts, which primarily include code and data, but may also be scripts, measurements and other assets. The linked artifacts are checked in the publication process and contribute to PVLDB's reproducibility efforts. The community feedback to all these changes was largely positive and sparked many suggestions for further improvements, such as the use of certain LaTeX features in the review process and a stricter template enforcement - features that we might see in PVLDB Volume 15.

As proceedings chairs, we want to take the opportunity of writing this editorial to thank all authors for their help in producing high-quality camera-ready copies of their papers. Fitting all content on 12 pages and still adhering to the format regulations is often not an easy task. In particular cases, last-minute actions and fast responses were needed to publish an issue in time, for which we are very grateful.

We hope you all enjoy Issue 6 and that it may inspire new ideas.

Hannes Mühleisen and Thorsten Papenbrock
Publication Editors of PVLDB Volume 14 & Proceedings Chairs for VLDB 2021