



# Proceedings of the VLDB Endowment

Volume 14, No. 2 – October 2020

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. 2, October 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**

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](mailto:info@vldb.org).

Volume 14, Number 2, October 2020

Pages i – vii and 74 - 240

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 .....	v
Editorial .....	vii

### Research Papers

Tsunami: A Learned Multi-dimensional Index for Correlated Data and Skewed Workloads .....	74
<i>Jialin Ding, Vikram Nathan, Mohammad Alizadeh, Tim Kraska</i>	
Jointly Optimizing Preprocessing and Inference for DNN-based Visual Analytics .....	87
<i>Daniel Kang, Ankit Mathur, Teja Veeramacheni, Peter Bailis, Matei Zahari</i>	
Permutable Compiled Queries: Dynamically Adapting Compiled Queries without Recompiling .....	101
<i>Prashanth Menon, Amadou Ngom, Todd Mowry, Andrew Pavlo, Lin Ma</i>	
EMOGI: Efficient Memory-access for Out-of-memory Graph-traversal In GPUs ... ..	114
<i>Seung Won Min, Vikram Sharma Mailthody, Zaid Qureshi, Jinjun Xiong, Eiman Ebrahimi, Wen-Mei Hwu</i>	
On-Off Sketch: A Fast and Accurate Sketch on Persistence .....	128
<i>Yinda Zhang, Jinyang Li, Yutian Lei, Tong Yang, Zhetao Li, Gong Zhang, Bin Cui</i>	
Real-Time Distance-Based Outlier Detection in Data Streams .....	141
<i>Luan Tran, Min Mun, Cyrus Shahabi</i>	
Seagull: An Infrastructure for Load Prediction and Optimized Resource Allocation .....	154
<i>Olga Poppe, Tayo Amunke, Dalitso Banda, Aritra De, Ari Green, Manon Knoertzer, Ehi Nosakhare, Karthik Rajendran, Deepak Shankargouda, Meina Wang, Alan Au, Carlo Curino, Qun Guo, Alekh Jindal, Ajay Kalhan, Morgan Oslake, Sonia Parchani, Vijay Ramani, Raj Sellappan, Saikat Sen, Sheetal Shrotri, Soundararajan Srinivasan, Ping Xia, Shize Xu, Alicia Yang, Yiwen Zhu</i>	
On the Efficiency of K-Means Clustering: Evaluation, Optimization, and Algorithm Selection .....	163
<i>Sheng Wang, Yuan Sun, Zhifeng Bao</i>	
RapidMatch: A Holistic Approach to Subgraph Query Processing .....	176
<i>Shixuan Sun, Xibo Sun, Yulin Che, Qiong Luo, Bingsheng He</i>	
Taurus: Lightweight Parallel Logging for In-Memory Database Management Systems.....	189
<i>Yu Xia, Xiangyao Yu, Andrew Pavlo, Srinivas Devadas</i>	
Improving Execution Efficiency of Just-in-time Compilation based Query Processing on GPUs.....	202
<i>Johns Paul, Bingsheng He, Shengliang Lu, Chiew Tong Lau</i>	
PPQ-Trajectory: Spatio-temporal Quantization for Querying in Large Trajectory Repositories .....	215
<i>Shuang Wang, Hakan Ferhatosmanoglu</i>	

Aggregated Deletion Propagation for Counting Conjunctive Query Answers ..... 228  
*Xiao Hu, Shouzhuo Sun, Shweta Patwa, Debmalya Panigrahi, Sudeepa Roy*

## **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 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

I am pleased to present the second issue of the Proceedings of the VLDB Endowment (PVLDB), Volume 14. PVLDB publishes premium-quality articles that address intellectually challenging topics in database systems. Submissions are permitted at the beginning of each month and reviewed through a journal-style process.

This volume collects 11 papers under the "regular search" category. Ding et al. propose a multidimensional index whose construction considers both the data and the query distribution. They show that the index is more amenable to skewed queries than the previous solutions. Kang et al. revisit the computational cost of deep neural networks and discuss the possibility of achieving higher efficiency by reducing data resolution and increasing the neural network's complexity. Aiming to integrate adaptive query processing (AQP) with just-in-time (JIT) query compilation, Menon et al. design the permutable compiled query (PCQ) technique to lower the re-compilation overhead in query plan modification. Min et al. tackle the cache-fault issue in performing graph traversals using GPUs. They alleviate the issue by allowing GPU threads to access data in the main memory directly. Zhang et al. consider the persistence estimation problem on data streams and describe the "on-off sketch" to achieve accurate estimation under a tight space budget. Tran et al. propose an algorithm to find distance-based outliers on data streams with a small memory footprint. Sun et al. discuss the benefits of combining two existing paradigms for answering subgraph queries: the exploration approach and the worst-case optimal join approach. Xia et al. investigate transaction logging for in-memory multi-core systems and describe how to accomplish the purpose by tracking the dependency among transactions. Paul et al. describe a JIT-based approach that improves the performance of GPU-based query algorithms by increasing their hardware utilization. Wang et al. study access methods to support spatiotemporal queries on trajectory data. They introduce the partition-wise predictive quantizer (PPQ) technique to enable effective filtering with attractive precision guarantees. Hu et al. propose the aggregated deletion propagation problem, where the objective is to remove the least number of tuples to shrink a query result by a designated amount. They establish several results on the problem's computational hardness.

In addition, the volume features a paper in the "scalable data science" category where Pope et al. explore how to use data science techniques to predict the server loads in Microsoft Azure for automatic resource allocation. The volume also includes a paper by Wang et al. in the "experiments, analyses, and benchmarks" category. They present an experimental evaluation of the k-means algorithm's existing implementations and analyze their advantages/disadvantages from different perspectives.

All the papers 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.

Yufei Tao  
PVLDB Associate Editor