



# Proceedings of the VLDB Endowment

Volume 14, No. 3 – November 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. 3, November 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 3, November 2020

Pages i – vii and 241 - 457

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

Breaking Down Memory Walls: Adaptive Memory Management in LSM-based Storage Systems ... <i>Chen Luo, Michael Carey</i>	241
Nearest Neighbor Classifiers over Incomplete Information: From Certain Answers to Certain Predictions .....	255
<i>Bojan Karlaš, Peng Li, Renzhi Wu, Nezihe Merve Gürel, Xu Chu, Wentao Wu, Ce Zhang</i>	
Elle: Inferring Isolation Anomalies from Experimental Observations .....	268
<i>Peter Alvaro, Kyle Kingsbury</i>	
Scotch: Generating FPGA-Accelerators for Sketching at Line Rate .....	281
<i>Martin Kiefer, Ilias Poulakis, Sebastian Bress, Volker Markl</i>	
ORBITS: Online Recovery of Missing Values in Multiple Time Series Streams .....	294
<i>Mourad Khayati, Ines Arous, Zakhar Tymchenko, Philippe Cudre-Mauroux</i>	
TURL: Table Understanding through Representation Learning .....	307
<i>Xiang Deng, Huan Sun, Alyssa Lees, You Wu, Cong Yu</i>	
EdgeDIPN: a Unified Deep Intent Prediction Network Deployed at the Edge .....	320
<i>Long Guo, Lifeng Hua, Rongfei Jia, Fei Fang, Binqiang Zhao, Bin Cui</i>	
LOCATER: Cleaning WiFi Connectivity Datasets for Semantic Localization ... ..	329
<i>Yiming Lin, Daokun Jiang, Roberto Yus, Georgios Bouloukakis, Andrew Chio, Sharad Mehrotra, Nalini Venkatasubramanian</i>	
Multi-Modal Transportation Recommendation with Unified Route Representation Learning .....	342
<i>Hao Liu, Jindong Han, Yanjie Fu, Jingbo Zhou, Xinjiang Lu, Hui Xiong</i>	
DISK: A Distributed Framework for Single-Source SimRank with Accuracy Guarantee .....	351
<i>Yue Wang, Ruiqi Xu, Zonghao Feng, Yulin Che, Lei Chen, Qiong Luo, Rui Mao</i>	
Toward a Better Understanding and Evaluation of Tree Structures on Flash SSDs .....	364
<i>Diego Didona, Nikolas Ioannou, Radu Stoica, Kornilios Kourtis</i>	
Answering Multi-Dimensional Range Queries under Local Differential Privacy .....	378
<i>Jianyu Yang, Tianhao Wang, Ninghui Li, Xiang Cheng, Sen Su</i>	
Ananke: A Streaming Framework for Live Forward Provenance .....	391
<i>Dimitris Palyvos-Giannas, Bastian Havers, Marina Papatriantafidou, Vincenzo Gulisano</i>	

RECEIPT: REfine CoarsE-grained IndePendent Tasks for Parallel Tip decomposition of Bipartite Graphs .....	404
<i>Kartik Lakhotia, Rajgopal Kannan, Viktor Prasanna, César De Rose</i>	
Comprehensive and Efficient Workload Compression .....	418
<i>Shaleen Deep, Anja Gruenheid, Paraschos Koutris, Jeff Naughton, Stratis Viglas</i>	
CoroBase: Coroutine-Oriented Main-Memory Database Engine .....	431
<i>Yongjun He, Jiacheng Lu, Tianzheng Wang</i>	
Scalable Querying of Nested Data .....	445
<i>Jaclyn Smith, Michael Benedikt, Milos Nikolic, Amir Shaikhha</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)  
 Sibow 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 third issue of PVLDB volume 13 exemplifies the amazing breadth of problems our community engages in. The papers in this volume address a bouquet of extremely relevant research challenges in streaming, storage, graph databases, experimental data, and machine learning algorithms, and propose systems and theory solutions. In a single volume one grasps the variety of moving targets researchers aim at, and at the same time the imagination runs wild as per the impact to applications and all the new challenges that emerge every day in our data-driven world.

Reading the volume, I am inspired by papers on predictability and privacy. Notwithstanding the efforts from academia and industry, it feels that these two problems grow as Lernaean Hydras. Predictability can be significantly boosted by intelligent machine learning models, but is undermined by the ever-growing complexity of software systems and the underlying microarchitectures (when predicting performance) and by human-related factors (when developing personalization strategies). Likewise, privacy is a self-conflicting goal in light of the “openness” in today’s world: the majority of the population discloses data to anyone who asks; ensuring privacy in a data-driven, data-managed, data-dependent world is a truly daunting task.

The role of machine learning is catalytic in much work that appears in this issue. The data management community uses machine learning in creative ways to enrich input with much-needed knowledge, and our contributions help evolve used learning models and inspire research in other disciplines, as is evident by the many cross-disciplinary citations. This issue also charges ahead with exciting systems research on flash storage and coroutine-sided main-memory databases, as well as interesting ideas on workload compression and on using experimental observations to find holes in isolation — out-of-the-box thinking to solve really hard problems that experimental computer science has been facing for decades.

Last but not least, I cannot help but notice that almost all papers in this issue have undergone a revision. This PVLDB initiative is a true guard for high-quality; through the tireless work of the shepherding reviewers, cutting-edge research and novel results are brought closer to the reader through polishing processes that stem from a scientific dialog with the authors. The revision process also benefits the authors, as making the work more well-rounded and accessible to a broad audience instigates followup research and improves potential for collaborations.

I hope that you are healthy and safe and that you will enjoy reading this issue.

Anastasia Ailamaki  
PVLDB Associate Editor