



Proceedings of the VLDB Endowment

Volume 14, No. 9 – May 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. 9, May 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 9, May 2021

Pages i – vii and 1454 - 1702

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

Minimum Vertex Augmentation	1454
<i>Jianwen Zhao, Yufei Tao</i>	
Database Isolation by Scheduling	1467
<i>Kevin P Gaffney, Robert K Claus, Jignesh Patel</i>	
SaS: SSD as SQL Database System.....	1481
<i>Jong-hyeok Park, Soyeon Choi, Gihwan Oh, Sang Won Lee</i>	
FLAT: Fast, Lightweight and Accurate Method for Cardinality Estimation	1489
<i>Rong Zhu, Ziniu Wu, Yuxing Han, Kai Zeng, Andreas Pfadler, Zhengping Qian, Jingren Zhou, Bin Cui</i>	
Fast Augmentation Algorithms for Network Kernel Density Visualization	1503
<i>Tsz Nam Chan, Zhe Li, Leong Hou U, Jianliang Xu, Reynold Cheng</i>	
AutoGR: Automated Geo-Replication with Fast System Performance and Preserved Application Semantics.....	1517
<i>Jiawei Wang, Cheng Li, Kai Ma, Jingze Huo, Feng Yan, Xinyu Feng, Yinlong Xu</i>	
Local Algorithms for Distance-generalized Core Decomposition over Large Dynamic Graphs.....	1531
<i>Qing Liu, Xuliang Zhu, Xin Huang, Jianliang Xu</i>	
Viper: An Efficient Hybrid PMem-DRAM Key-Value Store.....	1544
<i>Lawrence Benson, Hendrik Makait, Tilmann Rabl</i>	
Estimating Spread of Contact-Based Contagions in a Population Through Sub-Sampling	1557
<i>Sepanta Zeighami, Cyrus Shahabi, John Krumm</i>	
Trident: Task Scheduling over Tiered Storage Systems in Big Data Platforms.....	1570
<i>Herodotos Herodotou, Elena Kakoulli</i>	
Comprehensible Counterfactual Explanation on Kolmogorov-Smirnov Test.....	1583
<i>Zicun Cong, Lingyang Chu, Yu Yang, Jian Pei</i>	
Accelerating Large Scale Real-Time GNN Inference using Channel Pruning	1597
<i>Hongkuan Zhou, Ajitesh Srivastava, Hanqing Zeng, Rajgopal Kannan, Viktor K Prasanna</i>	
Towards Cost-Optimal Query Processing in the Cloud	1606
<i>Viktor Leis, Maximilian Kuschewski</i>	

Automating Incremental Graph Processing with Flexible Memoization	1613
<i>Shufeng Gong, Chao Tian, Qiang Yin, Wenyuan Yu, Yanfeng Zhang, Liang Geng, Song Yu, Ge Yu, Jingren Zhou</i>	
In-Network Support for Transaction Triaging	1626
<i>Theo Jepsen, Alberto Lerner, Fernando Pedone, Robert Soule, Philippe Cudre-Mauroux</i>	
Are We Ready for Learned Cardinality Estimation?	1640
<i>Xiaoying Wang, Changbo Qu, Weiyuan Wu, Jiannan Wang, Qingqing Zhou</i>	
On the algebra of data sketches	1655
<i>Jakub Lemiesz</i>	
Massively Parallel Algorithms for Personalized PageRank	1668
<i>Guanhao Hou, Xingguang Chen, Sibor Wang, Zhewei Wei</i>	
GeCo: Quality Counterfactual Explanations in Real Time.....	1681
<i>Maximilian Schleich, Zixuan Geng, Yihong Zhang, Dan Suciu</i>	
Automated Feature Engineering for Algorithmic Fairness	1694
<i>Ricardo Salazar, Felix Neutatz, Ziawasch Abedjan</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

I am delighted to introduce the ninth issue of the Proceedings of the VLDB Endowment (PVLDB), Volume 14. This issue comprises 16 regular research papers and two papers in each the Scalable Data Science (SDS) track, the Vision track, and the Experiment, Analysis and Benchmark (EA&B) track. The contributions collected in this issue cover a wide range of areas and topics, including database system design in today's and tomorrow's execution environments, graph processing and algorithms, data mining and data analysis, approximation algorithms, explanations, fairness, and much more.

Data management. Benson et al. explore persistent memory technologies in the context of key-value stores and propose a hybrid DRAM-PMem store called Viper. In their vision paper, Choi et al. discuss a new architecture called SSD as SQL database engine, in which the database engine is tightly integrated with the SSD architecture. Jepsen et al. introduce transaction triaging, which leverages programmable network hardware to manipulate transaction requests and responses to reduce networking overhead. Herodotou and Kakoulli investigate how to exploit storage tier information for task scheduling and propose an approach called Trident, which is implemented in Spark and Hadoop. Leis and Kuschewski envision cost-optimal query processing in the cloud by modelling workload, hardware, and cost. Gaffney et al. propose a method to disaggregate transaction isolation from the transactional storage manager using a design they call database isolation by scheduling. Wang et al. propose a framework termed AutoGR, which adds geo-replication features to non-replicated applications in an automated way.

Cardinality estimation. Zhu et al. propose a novel approach to cardinality estimation based on an adaptive tree-structured graphical model. In their EA&B paper, Wang et al. explore whether learned cardinality estimation techniques are ready to be deployed in production. They experimentally compared traditional and learned methods in both static and dynamic environments and outline directions for further research. Jakub Lemiesz designs a data sketch to estimate the weighted cardinality of weighted data streams as well as set theory operations involving multiple such streams.

Graphs. Zhao and Tao introduce, analyze and study experimentally a general class of graph problems called minimum vertex augmentation, which aim to find a minimum-cost recoloring of the vertices of a graph such that a user-defined predicate is satisfied. Liu et al. revisit distance-generalized core decompositions over dynamic graphs and propose efficient local algorithms for computing and maintaining such decompositions. Gong et al. propose the Ingress system (yes, indeed!) for incremental graph processing, which automatically incrementalizes vertex-centric computations. Hou et al. present Delta-Push, a distributed framework for single-source and top-k personalized page rank queries.

Data mining. Zeighami et al. investigate methods to estimate the spread of a phenomenon based on partial knowledge about contacts of individuals. Chan et al. present and analyze efficient methods for network kernel density visualization, which scale to datasets with millions of points. Addanki et al. study robust methods for finding maximums, nearest and farthest neighbors, and clustering based on relative distance queries under adversarial or probabilistic noise. Boniol et al. propose a method called SAND to detect anomalies in data streams in a domain-agnostic fashion. Cong et al. develop a method to produce counterfactual explanations for why a Kolmogorov-Smirnov test failed. In their contribution to the EA&B track, Schleich et al. present GeCo, a system to compute plausible and feasible counterfactual explanations of a prediction in real time. Zhou et al. propose to accelerate inference in graph neural networks by dimension pruning in their contribution to the SDS track. Also in the SDS track, Diaz et al. propose an automated feature selection and construction strategy that aims to provide algorithmic fairness and high accuracy.

All papers have undergone the PVLDB review process and will be presented at the 47th International Conference on Very Large Data Bases, 2021, in Copenhagen. I'd like to thank the authors, reviewers, fellow associate editors, program chairs and everybody else involved in compiling this issue. I hope that you find the contributions in this issue interesting and insightful.

Rainer Gemulla
PVLDB Associate Editor