



# Proceedings of the VLDB Endowment

Volume 15, No. 7 – March 2022

Editors in Chief:

**Fatma Özcan, Juliana Freire and Xuemin Lin**

Associate Editors:

**Arun Kumar, Azza Abouzied, Beng Chin Ooi, Boris Glavic, Dan Suciu,  
Divyakant Agrawal, Eugene Wu, Georgia Koutrika, Ioana Manolescu,  
Jeffrey Xu Yu, Julia Stoyanovich, Jun Yang, K. Selçuk Candan,  
Khuzaima Daudjee, Laure Berti-Equille, Lei Chen, Mohamed Mokbel,  
Neoklis Polyzotis, Paolo Papotti, Peter Boncz, Sebastian Schelter,  
Sourav S Bhowmick, Surajit Chaudhuri, Themis Palpanas, Vanessa Braganholo,  
Viktor Leis, Wang-Chiew Tan, Wenjie Zhang, Wook-Shin Han, Xiaofang Zhou**

Publication Editors:

**Lijun Chang and Xin Cao**

PVLDB – Proceedings of the VLDB Endowment

Volume 15, No. 7, March 2022.

All papers published in this issue will be presented at the 48th International Conference on Very Large Data Bases, Sydney, Australia, 2022.

## **Copyright 2022 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 15, Number 7, March 2022

Pages i – vii and 1311 - 1518

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. 15 .....	iv

### Research Papers

Prefix Filter: Practically and Theoretically Better Than Bloom.....	1311
<i>Tomer Even, Guy Even, Adam Morrison</i>	
Scalar DL: Scalable and Practical Byzantine Fault Detection for Transactional Database Systems ...	1324
<i>Hiroyuki Yamada, Jun Nemoto</i>	
In-Network Leaderless Replication for Distributed Data Stores.....	1337
<i>Gyuyeong Kim, Wonjun Lee</i>	
Fast Algorithms for Core Maximization on Large Graphs.....	1350
<i>Xin Sun, Xin Huang, Di Jin</i>	
NLC: Search Correlated Window Pairs on Long Time Series.....	1363
<i>Shuye Pan, Peng Wang, Chen Wang, Wei Wang, Jianmin Wang</i>	
Edge-based Local Push for Personalized PageRank .....	1376
<i>Hanzhi Wang, Zhewei Wei, Junhao Gan, Ye Yuan, Xiaoyong Du, Ji-rong Wen</i>	
Continuous Social Distance Monitoring in Indoor Space.....	1390
<i>Harry Kai-ho Chan, Huan Li, Xiao Li, Hua Lu</i>	
An In-Depth Study of Continuous Subgraph Matching.....	1403
<i>Xibo Sun, Shixuan Sun, Qiong Luo, Bingsheng He</i>	
OnlineSTL: Scaling Time Series Decomposition by 100x .....	1417
<i>Abhinav Mishra, Ram Sriharsha, Sichen Zhong</i>	
Stingy Sketch: A Sketch Framework for Accurate and Fast Frequency Estimation .....	1426
<i>Haoyu Li, Qizhi Chen, Yixin Zhang, Tong Yang, Bin Cui</i>	
A Study of Database Performance Sensitivity to Experiment Settings.....	1439
<i>Yang Wang, Miao Yu, Yujie Hui, Fang Zhou, Yuyang Huang, Rui Zhu, Xueyuan Ren, Tianxi Li, Xiaoyi Lu</i>	
The Inherent Time Complexity and An Efficient Algorithm for Subsequence Matching Problem .....	1453
<i>Zemin Chao, Hong Gao, Yinan An, Jianzhong Li</i>	
Selective Data Acquisition in the Wild for Model Charging.....	1466
<i>Chengliang Chai, Jiabin Liu, Nan Tang, Guoliang Li, Yuyu Luo</i>	
Discovering Association Rules from Big Graphs.....	1479
<i>Wenfei Fan, Wenzhi Fu, Ruochun Jin, Ping Lu, Chao Tian</i>	

DeepTEA: Effective and Efficient Online Time-dependent Trajectory Outlier Detection .....	1493
<i>Xiaolin Han, Reynold Cheng, Chenhao Ma, Tobias Grubenman</i>	
Entity Resolution On-Demand .....	1506
<i>Giovanni Simonini, Luca Zecchini, Sonia Bergamaschi, Felix Naumann</i>	

## **PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 15**

### **Editors in Chief of PVLDB**

Fatma Ozcan (Google)  
Juliana Freire (New York University)  
Xuemin Lin (University of New South Wales)

### **Associate Editors of PVLDB**

Arun Kumar (University of California, San Diego)  
Azza Abouzied (NYU Abu Dhabi)  
Beng Chin Ooi (NUS)  
Boris Glavic (Illinois Institute of Technology)  
Dan Suciu (University of Washington)  
Divyakant Agrawal (University of California, Santa Barbara)  
Eugene Wu (Columbia University)  
Georgia Koutrika (ATHENA)  
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)  
Jeffrey Xu Yu (Chinese University of Hong Kong)  
Julia Stoyanovich (New York University)  
Jun Yang (Duke University)  
K. Seçuk Candan (Arizona State University)  
Khuzaima Daudjee (University of Waterloo)  
Laks Lakshmanan (The University of British Columbia)  
Laure Berti-Equille (IRD)  
Lei Chen (Hong Kong University of Science and Technology)  
Mohamed Mokbel (University of Minnesota, Twin Cities)  
Neoklis Polyzotis (Google)  
Paolo Papotti  
Peter Boncz (CWI)  
Sebastian Schelter (University of Amsterdam)  
Sharad Mehrotra (U.C. Irvine)  
Sourav S Bhowmick (Nanyang Technological University)

Surajit Chaudhuri (Microsoft Research)  
Themis Palpanas (University of Paris)  
Vanessa Braganholo (Fluminense Federal University)  
Viktor Leis (Friedrich Schiller University Jena)  
Wang-Chiew Tan (Megagon Labs)  
Wenjie Zhang (University of New South Wales)  
Wook-Shin Han (POSTECH)  
Xiaofang Zhou (Hong Kong University of Science and Technology)

### **Publication Editors**

Lijun Chang (University of Sydney)  
Xin Cao (University of New South Wales)

### **PVLDB Managing Editor**

Wolfgang Lehner (Dresden University of Technology)

### **PVLDB Advisory Committee**

Felix Naumann (HPI)  
Juliana Freire (New York University)  
Xuemin Lin (U of New South Wales)  
Georgia Koutrika (Athena Research Center)  
Jun Yang (Duke University)  
Vanessa Braganholo (Universidade Federal Fluminense)  
Sourav S Bhowmick (Nanyang Technological University)  
Chris Jermaine (Rice University)  
Peter Triantafillou (University of Warwick)  
Xin Luna Dong (Facebook)  
Fatma Ozcan (Google)  
Lei Chen (Hong Kong University of S&T)  
Graham Cormode (University of Warwick)  
Divesh Srivastava (AT&T Labs-Research)  
Wolfgang Lehner (TU Dresden)

## Review Board

Abolfazl Asudeh (University of Michigan)  
Aécio Santos (New York University)  
Ahmed Eldawy (University of California, Riverside)  
Alexander Hall (RelationalAI)  
Alexander J Ratner (University of Washington)  
Aline Bessa (New York University)  
Alkis Simitsis (Athena Research Center)  
Altigran da Silva (Universidade Federal do Amazonas)  
AnHai Doan (University of Wisconsin-Madison)  
Anna Fariha (Microsoft)  
Anton Dignös (Free University of Bozen-Bolzano)  
Antonio Cavalcante Araujo Neto (University of Alberta)  
Arijit Khan (Nanyang Technological University)  
Arvind Arasu (Microsoft)  
Babak Salimi (University of California, San Diego)  
Bailu Ding (Microsoft Research)  
Bertram Ludascher (University of Illinois)  
Bolong Zheng (Huazhong University of Science and Technology)  
Brandon Haynes (Gray Systems Lab, Microsoft)  
Byron Choi (Hong Kong Baptist University)  
Carlo Curino (Microsoft -- GSL)  
Carlos Scheidegger (The University of Arizona)  
Carsten Binnig (TU Darmstadt)  
Ce Zhang (ETH)  
Cheng Long (Nanyang Technological University)  
Chengfei Liu (Swinburne University of Technology)  
Chuan Lei (Instacart)  
Chunbin Lin (Amazon AWS)  
Curtis Dyreson (Utah State University)  
Dan Kifer (Pennsylvania State University)  
Dana M Van Aken (Carnegie Mellon University)  
Daniel Deutch (Tel Aviv University)  
Daniel Oliveira (UFF, Brazil)  
David Koop (Northern Illinois University)  
Davide Mottin (Aarhus University)  
Dong Xie (Penn State University)  
Eduardo Ogasawara (CEFET-RJ)  
Eleni Tzirita Zacharitou (TU Berlin)  
Fabio Porto (LNCC)  
Faisal Nawab (University of California at Irvine)  
Fan Zhang (Guangzhou University)  
Fatemeh Nargesian (University of Rochester)  
Fei Chiang (McMaster University)  
Florin Rusu (UC Merced)  
Floris Geerts (University of Antwerp)  
Fotis Psallidas (Microsoft)  
George Fletcher (Eindhoven University of Technology)  
George Papadakis (University of Athens)  
Gerhard Weikum (Max-Planck-Institut für Informatik)  
Germain Forestier (University of Haute Alsace)  
Guoliang Li (Tsinghua University)  
Haipeng Dai (Nanjing University)  
Harish Doraiswamy (Microsoft Research India)  
Heiko Mueller (DeepReason.ai)  
Herodotos Herodotou (Cyprus University of Technology)

Holger Pirk (Imperial College)  
Hongzhi Yin (The University of Queensland)  
Huiping Cao (New Mexico State University)  
Immanuel Trummer (Cornell)  
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)  
Ippokratis Pandis (Amazon)  
Ishtiyaque Ahmad (University of California, Santa Barbara)  
Jae-Gil Lee (KAIST)  
Jana Giceva (TU Munich)  
Jeffrey Xu Yu (Chinese University of Hong Kong)  
Jens Teubner (TU Dortmund University)  
Jia Zou (Arizona State University)  
Jian Pei (Simon Fraser University)  
Jianguo Wang (Purdue University)  
Jiannan Wang (Simon Fraser University)  
Jianxin Li (Deakin University)  
Jianye Yang (Central South University)  
Jiwon Seo (Hanyang University)  
Johannes Gehrke (Microsoft)  
Jorge Arnulfo Quiane Ruiz (TU Berlin)  
Joseph Near (University of Vermont)  
Junhu Wang (Griffith University)  
Kaiping Zheng (National University of Singapore)  
Kangfei Zhao (The Chinese University of Hong Kong)  
Karima Echiabi (Mohammed VI Polytechnic University)  
Katja Hose (Aalborg University)  
Kenneth A Ross (Columbia University)  
Kostas Zoumpatianos (Snowflake Computing)  
Lei Zou (Peking University)  
Leopoldo Bertossi (Universidad Adolfo Ibanez)  
Li Xiong (Emory University)  
Lianke Qin (University of California, Santa Barbara)  
Lijun Chang (The University of Sydney)  
Lin Ma (Carnegie Mellon University)  
Long Yuan (Nanjing University of Science and Technology)  
Lu Qin (UTS)  
Luciano Barbosa (Universidade Federal de Pernambuco)  
Marcelo Arenas (Universidad Católica & IMFD)  
Maria Luisa Sapino (U. Torino)  
Matteo Lissandrini (Aalborg University)  
Matthias Boehm (Graz University of Technology)  
Matthias Renz (University of Kiel)  
Max Heimerl (Snowflake)  
Maximilian Schleich (University of Washington)  
Meihui Zhang (Beijing Institute of Technology)  
Melanie Herschel (Universität Stuttgart)  
Michael Abebe (University of Waterloo)  
Min Xie (Instacart)  
Mirella M Moro (Universidade Federal de Minas Gerais)  
Mohamed Sarwat (Arizona State University)  
Mohammad Dashti (MongoDB)  
Mohammad Javad Amiri (University of Pennsylvania)  
Mohammad Sadoghi (University of California, Davis)  
Muhammad Aamir Cheema (Monash University)

Nikita Bhutani (Megagon Labs)  
 Oliver A Kennedy (University at Buffalo, SUNY)  
 Panos K. Chrysanthis (University of Pittsburgh)  
 Paolo Missier (Newcastle University)  
 Parth Nagarkar (NMSU)  
 Paul Groth (University of Amsterdam)  
 Peng CHENG (East China Normal University)  
 Peter Pietzuch (Imperial College London)  
 Pierangela Samarati (Universita delgi Studi di Milano)  
 Pinar Karagoz (METU, Turkey)  
 Pinar Tozun (IT University of Copenhagen)  
 Prithu Banerjee (UBC)  
 Raoni Lourenço (New York University)  
 Raul Castro Fernandez (UChicago)  
 Ravi Ramamurthy (Microsoft)  
 Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)  
 Renata Borovica-Gajic (University of Melbourne)  
 Reynold Cheng (The University of Hong Kong)  
 Rui Mao (Shenzhen University)  
 Ruoming Jin (Kent State University)  
 Sai Wu (Zhejiang University)  
 Sainyam Galhotra (University of Chicago)  
 Sanjay Krishnan (University of Chicago)  
 Sanjib Kumar Das (Google)  
 Sayan Ranu (IIT Delhi)  
 Sebastian Link (University of Auckland)  
 Semih Salihoglu (University of Waterloo)  
 Senjuti Basu Roy (New Jersey Institute of Technology)  
 Sergey Melnik (Google)  
 Shantanu Sharma (New Jersey Institute of Technology)  
 Shaoxu Song (Tsinghua University)  
 Sheng Wang (New York University)  
 Shimin Chen (Chinese Academy of Sciences)  
 Shumo Chu (University of California, Santa Barbara)  
 Shweta Jain (University of Illinois, Urbana-Champaign)  
 Sibow Wang (The Chinese University of Hong Kong)  
 Srinivasan Keshav (University of Cambridge)  
 Steffen Zeuch (DFKI GmbH)  
 Steven E Whang (KAIST)  
 Subarna Chatterjee (Harvard University)  
 Sudip Roy (Google)  
 Supun C Nakandala (University of California, San Diego)  
 Tamer Özsu (University of Waterloo)  
 Tarique A Siddiqui (Microsoft Research)  
 Thomas Heinis (Imperial College)  
 Thomas Neumann (TUM)  
 Tianzheng Wang (Simon Fraser University)  
 Tien Tuan Anh Dinh (Singapore University of Technology and Design)  
 Tilmann Rabl (HPI, University of Potsdam)  
 Ting Yu (Qatar Computing Research Institute)  
 Torben Bach Pedersen (Aalborg University)  
 Torsten Grust (Universität Tübingen)  
 Umar Farooq Minhas (Microsoft Research)  
 Vasiliki Kalavri (Boston University)  
 Verena Kantere (National Technical University of Athens)  
 Victor Zakhary (Oracle)  
 Vivek Narasayya (Microsoft Research)  
 Vraj Shah (University of California, San Diego)  
 Walid G Aref (Purdue)  
 Wasay Abdul (Harvard)  
 Wei Wang (Hong Kong University of Science and Technology (Guangzhou))  
 Wei Lu (Renmin university of china)  
 Weiren Yu (University of Warwick)  
 Wen Hua (The University of Queensland)  
 Wolfgang Lehner (TU Dresden)  
 Xi He (University of Waterloo)  
 Xiang Lian (Kent State University)  
 Xiao Qin (IBM Research)  
 Xiaofei Zhang (University of Memphis)  
 Xiaokui Xiao (National University of Singapore)  
 Xiaolan Wang (Megagon Labs)  
 Xiaoyang Wang (Zhejiang Gongshang University)  
 Xin Huang (Hong Kong Baptist University)  
 Yael Amsterdamer (Bar-Ilan university)  
 Yanyan Shen (Shanghai Jiao Tong University)  
 Ye Yuan (Northeastern University)  
 Yeye He (Microsoft Research)  
 Yi Chen (NJIT)  
 Yi Lu (MIT)  
 Yikai Zhang (Chinese University of Hong Kong)  
 Yinan Li (Microsoft Research)  
 Ying Zhang (University of Technology Sydney)  
 Yongxin Tong (Beihang University)  
 Yuanyuan Zhu (Wuhan University)  
 Yue Wang (Shenzhen Institute of Computing Sciences, Shenzhen University)  
 Yufei Tao (Chinese University of Hong Kong)  
 Yuliang Li (Megagon Labs)  
 Yuncheng Wu (National University of Singapore)  
 Yunjun Gao (Zhejiang University)  
 Yuval Moskovitch (University of Michigan)  
 Zhifeng Bao (RMIT University)  
 Zhongle Xie (Zhejiang University)  
 Zi Huang (University of Queensland)  
 Ziawasch Abedjan (Leibniz Universität Hannover)  
 Zohar Karnin (Amazon)  
 Zsolt István (IT University of Copenhagen)

## LETTER FROM THE EDITORS IN CHIEF

Welcome to the seventh issue of Proceedings of the VLDB Endowment (PVLDB), Volume 15. PVLDB provides a high-quality publication service to the data management research community. Each volume offers twelve monthly submission deadlines on the first day of each month and a quick, six weeks, reviewing cycle. This publication model was pioneered by PVLDB and combines a journal-style reviewing process, which includes a three-month revision cycle, with the agility and visibility provided by rapid on-line publication, and presentation at the annual VLDB conference.

This issue contains 16 papers: 13 regular research papers, 2 experiments analysis & benchmark (EA&B) papers, and 1 scalable data science (SDS) paper. Particularly, it covers papers in graph and time series analytics, distributed database systems, scalable ML, spatial and temporal data management, and entity resolution. All the papers in this issue are accepted after two rounds of review to achieve a high quality.

Yamada and Nemoto present Scalar DL, a middleware designed to extend existing database systems to provide an abstraction of a transactional database system with two way byzantine fault detection. Kim and Lee describe NetLR, a new replicated data store architecture that leverages a programmable top-of-rack network switch as a replication coordinator for performing leaderless data replication. Even et.al. propose a new sketch structure to test for set membership, which is space-efficient and supports both fast insertions and queries. Li et.al. introduces a sketch framework for highly skewed streaming data, which optimizes both speed and accuracy.

Sun, Huang and Jin study the problem of k-core maximization in large graphs by adding new edges under a budget constraint. Wang et.al. propose an edge-based local push method for computing approximate personalized page rank on weighted graphs. Fan et.al. revisit the graph association rule discovery, addressing the challenges with an excessive number of rules and scaling it to large graphs.

Chao et.al. propose an extended piecewise aggregate approximation method for time series subsequence matching problem. Pan et.al. study the problem of searching non-linear correlated window pairs on two long time series under the constraints of time delay and window length.

Chai et.al. present a mechanism to augment a labelled training dataset with additional data points from other similar labelled datasets for supervised ML. Chan et.al. study the problem of continuous social distance monitoring to predict the distance between two moving objects in indoor spaces. Han et.al. address the online time-dependent trajectory outlier detection problem, which aims to extract abnormal movements of vehicles on the roads, using a deep-learning framework.

Simonini et.al. integrate Select-Project-Order By queries with entity-resolution algorithms in an on-demand entity resolution framework by answering a SQL SP query on dirty data by utilizing traditional blocking and the order of seeds while progressively returning results as if they were issued on cleaned data.

The EAB paper by Sun et.al. evaluates five continuous subgraph matching algorithms. Wang et.al. investigates the sensitivity of the throughput performance of 11 different systems under a wide range of configuration settings for TPC-C and YCSB benchmarks.

Finally, the SDS paper by Mishra, Sriharsha and Zhong proposes a scalable online algorithm for time series decomposition into trend, seasonality and remainder components.

All the papers in this issue will be presented at the 48th International Conference on Very Large Data Bases, 2022, in Sydney. We sincerely thank all the authors for submitting their work and all the reviewers and associate editors for their outstanding service in reviewing the submissions. We hope that the readers will find the selected papers engaging, and thought provoking, providing valuable insights and inspiring follow-up research.

Fatma Özcan, Juliana Freire and Xuemin Lin  
Editors-in-Chief of PVLDB Volume 15  
Program Chairs for VLDB 2022